

Web Hosting

Definition

The World Wide Web (WWW), a web of worldwide servers connected to the Internet, provides an easily used and understood method of accessing electronic content. Accessing information requires data communication between a Webbrowser client and a Web-server application. Web hosting, then, is a means of hosting the Web-server application on a computer system through which electronic content on the Internet is readily available to any Web-browser client.

Overview

This tutorial will provide a basic overview of the main components that enable the Web, present two basic methods of Web hosting known as dedicated and shared, and discuss the challenges of resource management.

Topics

1. Overview of the Web
 2. Overview of Web Hosting
 3. Web-Hosting Implementation on a Dedicated Platform
 4. Web-Hosting Implementation on a Shared Platform
 5. Web Hosting–Resource Management Challenges
 6. Advanced Web-Hosting Methods
- Self-Test
Correct Answers
Glossary

1. Overview of the Web

In late 1990 while working at CERN, the European Laboratory for Particle Physics Research in Geneva, Switzerland, Tim Berners-Lee invented the Web, including the definitions of universal resource locator (URL), hypertext transfer protocol (HTTP), and hypertext markup language (HTML). The Web provides a

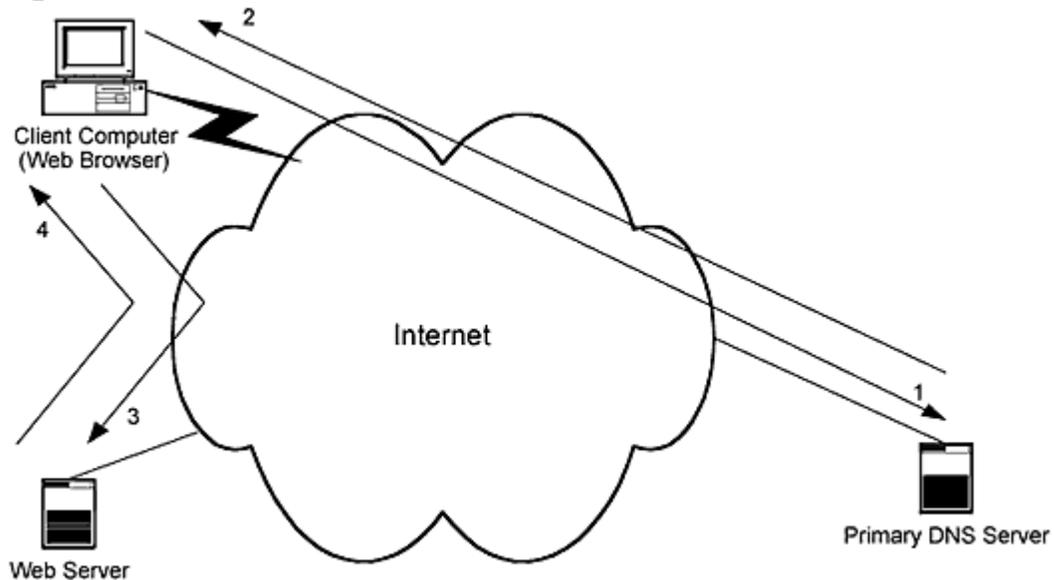
method for easily linking content contained on computer systems distributed throughout the world and connected to the Internet. Utilizing the Web, content on servers from many locations can be seamlessly linked and presented as a comprehensive resource collection. The Web further strengthens the power of the Internet's foundation of distributed computing.

The Web and the Internet remained the world's best-hidden resource until 1993 when Marc Andreessen, an undergraduate at the University of Illinois in Champaign, and a team at the National Center for Supercomputing Applications (NCSA) created the NCSA Mosaic browser. The NCSA Mosaic browser was the first Web-browser client that provided a friendly, point-and-click method for navigating the Internet using the Web.

The invention of the NCSA Mosaic browser was the start of the unprecedented growth of Internet users, Internet service providers (ISPs), and Internet business opportunities. By means of a user-friendly approach to searching and viewing the vast amount of information on the Internet, the Web-browser client enabled nontechnical individuals to benefit from the power and resources of the Internet.

Accessing content through the Web consists of communication between a Web-browser client and a Web server utilizing HTTP (see *Figure 1*).

Figure 1. Web Overview



The following is a step-by-step description of the communication path, as shown in *Figure 1*. It assumes that the Web server, the primary domain naming system (DNS) server, and the client computer are connected to the Internet and that all communication is conducted through the Internet.

steps 1 and 2—The end-user types a URL into the Web browser. The client computer finds the Internet protocol (IP) number associated with the domain name in the URL from the primary DNS server.

steps 3 and 4—The client computer uses the IP number obtained from the primary DNS server to request, through HTTP, the default HTML file from the Web server associated with the URL. The Web server sends the default HTML file to the client computer. The default HTML file provides information to the client computer for requesting all associated files—such as graphics—for the Web site's complete home page.

When the client computer and Web browser request and receive files from the same URL, the client computer is not required to perform a DNS lookup as described in steps 1 and 2. When the client computer attempts to retrieve a Web site from a different URL, the client computer must then perform steps 1 and 2 again.

2. Overview of Web Hosting

The complex web of servers consists of computer systems installed with Web-server software and connected to the Internet. These servers can be found in any facility with Internet connectivity. The process of maintaining and operating one of these servers is called Web hosting. Web hosting can be conducted in-house by the owner of the Web site, or it can be outsourced to a Web presence provider (WPP).

WPPs are typically companies with one or more data-center facilities that are connected to the Internet. Web hosting provided by WPPs can vary widely with respect to service quality and cost. Some providers consist simply of a room in the basement of a house and a tier-1 (T1) line connected to the local ISP. Others, however, are corporations with state-of-the-art hosting centers consisting of redundant fiber paths for high-speed Internet connections, redundant electrical power sources, a dry pipe–fire suppression system, and an experienced operations group, available 24 hours a day, seven days a week.

Web hosting can be provided on a shared computer environment or on a dedicated computer system. When a Web site consists only of standard HTML code and receives a small number of visitors, shared hosting service is the best solution. When a Web site consists of complex common gateway interface (CGI) scripts and proprietary programs and receives a large number of visitors, dedicated hosting service is the best solution.

3. Web-Hosting Implementation on a Dedicated Platform

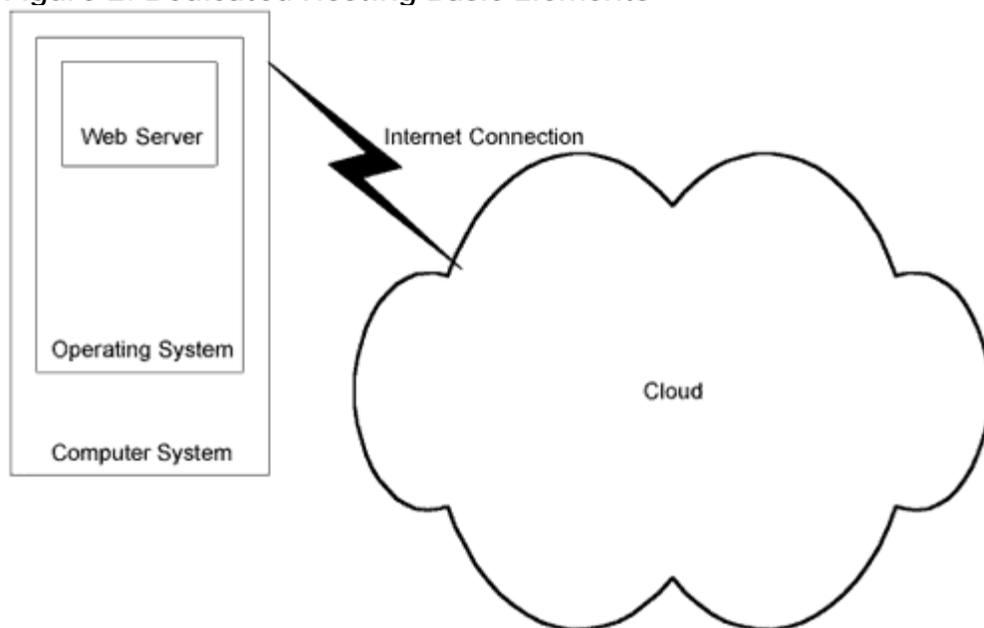
The basic concept of Web hosting on a dedicated computer system consists of hosting one Web site on one computer system. The dedicated environment offers complete flexibility and security to both the WPP and the customer.

Web hosting on a dedicated computer system is the simplest and most straightforward method of operating a Web site. Because the computer system contains only one Web site, the configuration of software is standardized, as outlined in the software-installation documentation. Furthermore, system resources are dedicated to only one Web site and, therefore, are not constrained by any other process not associated with the operations of that site.

The essential components of Web hosting on a dedicated computer system are as follows (see *Figure 2*):

- computer system hardware
- operating system (including transfer control protocol [TCP]/Internet protocol [IP] stack)
- Internet connection (IP number and domain name)
- Web server software (HTTP)

Figure 2. Dedicated Hosting Basic Elements



Additional software applications can be added to the computer system to enhance the Web site and to simplify the process of uploading content. One of these applications is a file transfer protocol (FTP) server for remote access to the computer system for transferring HTML content files.

4. Web-Hosting Implementation on a Shared Platform

The basic concept of Web hosting on a shared computer environment consists of hosting many different Web sites on one computer system. The shared environment offers economic benefits to both the WPP and the customer. Because the Web-hosting environment is the same for all customers, the provider gains economic benefits from allocating portions of the total cost of the hardware, software, maintenance and operation, and customer support amongst all customers. Therefore, the total fixed cost is less on a per-customer basis than with dedicated hosting. The customer gains economic benefit by the reduced price of the Web-hosting service.

The essential components of Web hosting on a shared computer environment are the same as with dedicated hosting, except for the configuration of the software and the management of system resources. There are two basic ways to configure Web-server software for multiple Web sites. The first method is to configure the Web server with each Web site's specific configuration information. The second method is to operate multiple Web-server software on a single computer environment. The first method—a single configuration file with all of the Web site's information—has greater scalability but does not provide a means of limiting the resources consumed by each Web site. Therefore, a combination of both methods is ideal for creating a scalable shared-hosting service. A combination is achieved by using the single configuration file method for Web sites requiring small amounts of resources and using the multiple Web-server method to limit the resources consumed by Web sites that demand large amounts of resources.

When a Web site demands large amounts of system resources, the logical next step is to move the Web site to a dedicated computer system (i.e., dedicated hosting).

5. Web Hosting—Resource Management Challenges

Managing computer-system resources in the shared platform and the dedicated platform is challenging. As a Web site becomes more popular and is sought after by millions of Internet users, the Web site demands more and more system

resources. Being able to measure, monitor, and manage the amount of system resources is essential for Web-site availability and server performance.

Critical system resources to manage include the following:

- central processing unit (CPU) utilization
- memory utilization
- disk-swap space
- disk space
- disk input and output
- network input and output
- Internet bandwidth (not a computer-system resource but still requires monitoring and managing)

These critical system resources have a direct relationship with the performance of a specific Web site. A Web site can be created or modified to minimize the demand on these system resources. Some Web sites are developed without the consideration of system-resource utilization. When a Web site contains and executes a CGI script, CPU resources are demanded. If the Web site contains a large number of CGI scripts and requires these scripts to be executed by every Web-site visitor, then CPU resources become a major bottleneck and cause the Web site to appear slow. It is important for the Web-site designer and developer to balance system-resource demands with Web-site functionality and creativity.

To measure, monitor, and manage the computer-system resources, additional software must be installed on the computer system. Each type of computer system hardware requires specific software for resource management. The computer-system manufacturer and operating system–software developer should be able to identify the necessary software applications for measuring, monitoring, and managing the system resources for their specific computer systems.

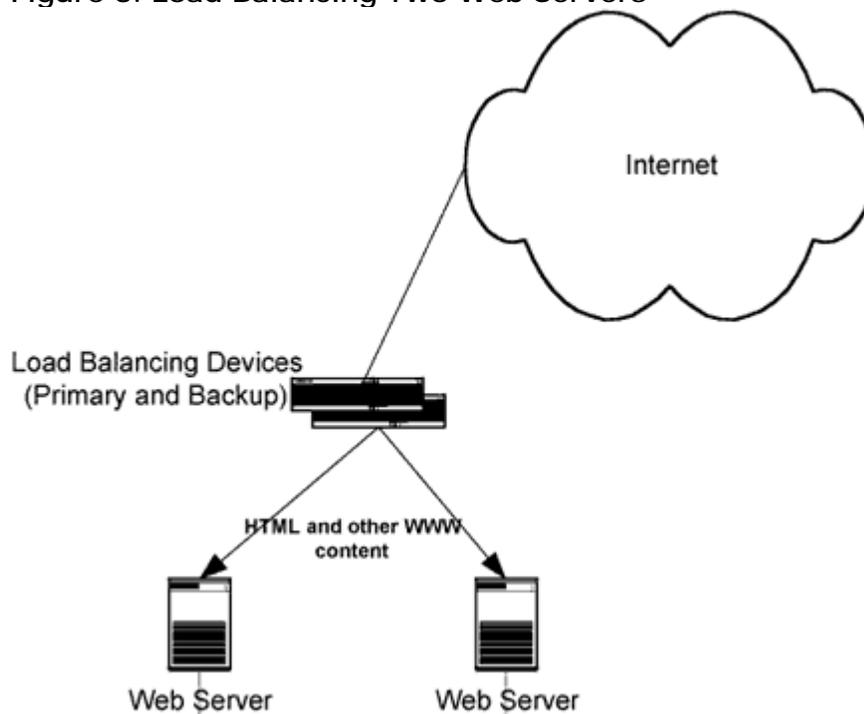
6. Advanced Web-Hosting Methods

During the last several years, Web hosting has evolved from simple one-computer system architectures to redundant, load-balanced server farms. A server farm is a network of computer systems. As a Web site demands more and more system resources, the traditional hosting environment is constrained by the limited amount of available resources. There are two basic means of providing more resources: a larger computer system or a distributed computer environment. To

provide redundancy and scalability, the distributed computer environment is the preferred method of expanding system resources.

The simplest distributed computer environment consists of two identical Web servers on the same local-area network (LAN) with a load-balancing device (see *Figure 3*). The load-balancing device is the gateway for all traffic entering and leaving the Web servers. The load balancer directs the incoming traffic to the best performing Web server, to alleviate all resource bottlenecks. With the load balancer as the gateway, the two Web servers appear as one large computing environment to all end-users on the Internet. This simple distributed computer environment can be expanded to accommodate more Web servers, providing greater scalability and consistently high performance levels.

Figure 3. Load Balancing Two Web Servers



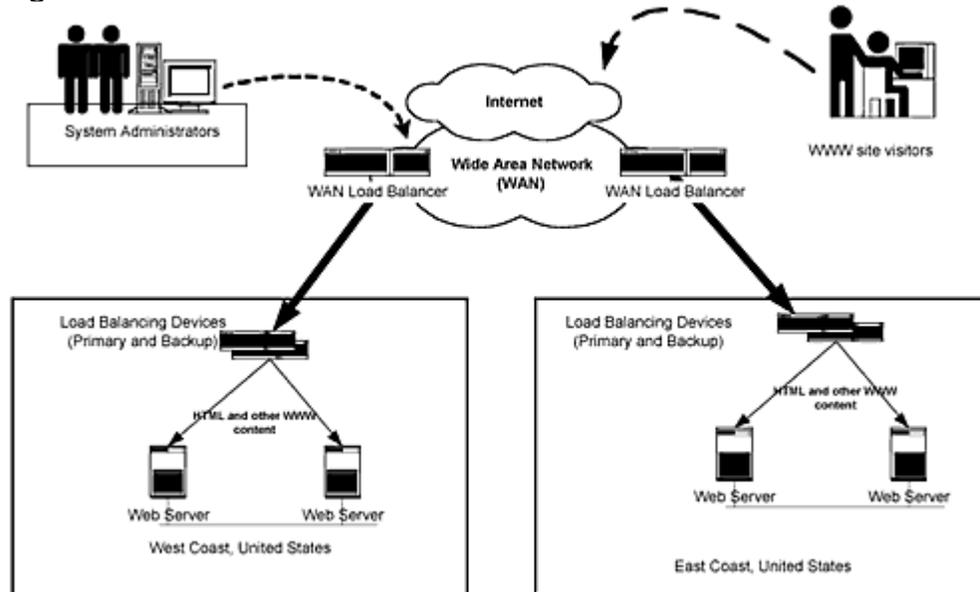
The simple distributed-computer environment provides a method for increasing the available computer-system resources, but it will not prevent performance problems associated with specific network issues within the LAN or with the Internet connection at that specific location. To overcome local network problems, Web hosting has continued to evolve into a geographically distributed computing–environment architecture.

By distributing the traffic of a Web site across multiple servers located in dispersed geographic locations, system resources can be added without interruptions in the Web-hosting service, and the Web site can always be available despite LAN or Internet-connection problems. Moreover, with

intelligent wide-area network (WAN) load balancing, Web-site performance will increase for all visitors, regardless of their geographic location.

Figure 4 illustrates Web hosting in a geographically distributed computing environment.

Figure 4. Two Site Architectures



Self-Test

1. Who invented the World Wide Web?
 - a. Albert Einstein
 - b. Marc Andreessen
 - c. Tim Berners-Lee
 - d. Bill Gates
2. With what server does the Web browser communicate to find the IP number of a Web site?
 - a. Web server
 - b. DNS server
 - c. FTP server

- d. HTTP server
3. What is a CPU?
 - a. central processing unit
 - b. the brain of the computer
 - c. a large hard drive
 - d. both a and b
 4. Maintaining and operating servers is performed through which of the following:
 - a. the operating system
 - b. the system board
 - c. RAM
 - d. Web hosting
 5. What computer-system resources are the most critical to measure, monitor, and manage for Web hosting?
 - a. CPU utilization
 - b. memory utilization
 - c. disk input and output
 - d. network input and output
 - e. all of the above
 6. Web hosting on a shared computer environment is normally accomplished by which method?
 - a. multiple disk drives, one for each Web site
 - b. multiple instances of Web-server software, one for each Web site
 - c. one configuration file with all Web site–configuration information
 - d. multiple CPUs, one for each Web site
 - e. both b and c

7. A Web server farm refers to which of the following:
 - a. land used for data-center facilities
 - b. a network of computer systems
 - c. one large computer system
 - d. a DNS and FTP server

8. Which of the following are essential components of Web hosting?
 - a. an FTP server
 - b. an Internet connection
 - c. electronic commerce
 - d. Web-server software
 - e. both b and d

9. What is the language used in the Web to create Web pages?
 - a. HTTP
 - b. HTML
 - c. FTP
 - d. URL

10. What protocol is used in the Web to transfer information from Web servers to Web browsers?
 - a. URL
 - b. HTTP
 - c. POP
 - d. OSI

Correct Answers

1. Who invented the World Wide Web?

- a. Albert Einstein
- b. Marc Andreessen
- c. Tim Berners-Lee**
- d. Bill Gates

See Topic 1.

2. With what server does the Web browser communicate to find the IP number of a Web site?

- a. Web server
- b. DNS server**
- c. FTP server
- d. HTTP server

See Topic 1.

3. What is a CPU?

- a. central processing unit**
- b. the brain of the computer
- c. a large hard drive
- d. both a and b

See Topic 5.

4. Maintaining and operating servers is performed through which of the following:

- a. the operating system
- b. the system board
- c. RAM

d. Web hosting

See Topic 2.

5. What computer-system resources are the most critical to measure, monitor, and manage for Web hosting?
- a. CPU utilization
 - b. memory utilization
 - c. disk input and output
 - d. network input and output
 - e. all of the above**

See Topic 5.

6. Web hosting on a shared computer environment is normally accomplished by which method?
- a. multiple disk drives, one for each Web site
 - b. multiple instances of Web-server software, one for each Web site
 - c. one configuration file with all Web site–configuration information
 - d. multiple CPUs, one for each Web site
 - e. both b and c**

See Topic 4.

7. A Web server farm refers to which of the following:
- a. land used for data-center facilities
 - b. a network of computer systems**
 - c. one large computer system
 - d. a DNS and FTP server

See Topic 6.

8. Which of the following are essential components of Web hosting?
- a. an FTP server

- b. an Internet connection
- c. electronic commerce
- d. Web-server software

e. both b and d

See Topic 3.

9. What is the language used in the Web to create Web pages?

a. HTTP

b. HTML

c. FTP

d. URL

See Topic 1.

10. What protocol is used in the Web to transfer information from Web servers to Web browsers?

a. URL

b. HTTP

c. POP

d. OSI

See Topic 1.

Glossary

CGI

common gateway interface; an application that takes input from the client, processes it on the server, and returns the results

DNS

domain naming system; a general-purpose distributed, replicated, data-query service; it is principally used to look up host IP addresses based on host names; some important domains are .com (commercial), .edu (educational), .net (network operations), .gov (U.S. government), and .mil (U.S. military); most countries also have a domain; the country domain names are based on the

International Standards Organization (ISO) 3166 (e.g., .us [United States], .uk [United Kingdom], .au [Australia])

FTP

file transfer protocol; a protocol that allows a user on one host to access and transfer files to and from another host over a network; FTP is commonly the name of the program the user invokes to execute the protocol

HTML

hypertext markup language; the language used in the Web to create Web pages with links to other documents, rich text enhancements (bold, italics, etc.), and so on; the source file for what is seen on a Web page is written in HTML

HTTP

hypertext transfer protocol; used in the Web to transfer information from Web servers to Web browsers

SSL

secure socket layer; an encryption technology for the Web used to provide secure transactions such as the transmission of credit-card numbers for e-commerce

URL

uniform resource locator; an identifier that describes the location of a particular piece of information (e.g., document) including the protocol used to retrieve that information

Web

World Wide Web; a hypertext-based, distributed information system created by researchers at CERN in Switzerland, enabling users to create, edit, or browse hypertext documents