Chapter 8
Web Server Hardware and Software

At a Glance

Instructor’s Manual Table of Contents
- Chapter Overview
- Chapter Objectives
- Instructor Notes
- Quick Quizzes
- Discussion Questions
- Additional Resources
- Key Terms
**Chapter Overview**

This chapter provides background information on the basic technologies used to build Web sites that can support online business operations. It includes a discussion of server software and hardware. It also includes an introduction to software that these sites use to perform utility functions such as site maintenance, diagnostics, and e-mail management.

**Chapter Objectives**

In this chapter, you will learn about:

- Web server basics
- Software for Web servers
- E-mail management and spam control issues
- Internet and Web site utility programs
- Web server hardware

**Instructor Notes**

**Web Server Basics**

The main job of a Web server computer is to respond to requests from Web client computers. The three main elements of a Web server are the hardware (computers and related components), operating system software, and Web server software. All three of these elements must work together to provide sufficient capacity in a given situation.

**Types of Web Sites**

Companies create Web sites for a wide variety of reasons and in a wide variety of forms. Each has a different purpose, requires different computer hardware and software, and requires different monetary and personnel resources.

**Types of Web Sites:**

- Development sites
- Intranets
- Extranets
- Transaction-processing sites
- Content-delivery sites
Web Clients and Web Servers

When people use their Internet connections to become part of the Web, their computers become Web client computers on a worldwide client/server network. Client/server architectures are used in LANs, WANs, and the Web. In a client/server architecture, the client computers typically request services, such as printing, information retrieval, and database access, from the server, which processes the clients’ requests. The computers that perform the server function usually have more memory and larger, faster disk drives than the client computers they serve.

Dynamic Content

A dynamic page is a Web page whose content is shaped by a program in response to user requests, whereas a static page is an unchanging page retrieved from disk. A server delivering mostly static Web pages performs better than the same server delivering dynamic Web pages because static page delivery requires less computing power than dynamic page delivery. The largest performance differences between competing Web server products appear when servers deliver dynamic pages.

Dynamic content is nonstatic information constructed in response to a Web client’s request. For example, if a Web client inquires about the status of an existing order by entering a unique customer number or order number into a form, the Web server searches the customer information and generates a dynamic Web page based on the customer information it found, thus fulfilling the client’s request. Assembled from back-end databases and internal data on the Web site, a dynamic page is a specific response to the requester’s query.

Various Meanings of “Server”

A server is any computer used to provide (or “serve”) files or make programs available to other computers connected to it through a network (such as a LAN or a WAN). The software that the server computer uses to make these files and programs available to the other computers is sometimes called server software. Sometimes this server software is included as part of the operating system that is running on the server computer. Thus, some information systems professionals informally refer to the operating system software on a server computer as server software, a practice that adds considerable confusion to the use of the term “server.”

Web Server Basics:

- **Web Client/Server Communication:** When a person uses a Web browser to visit a Web site, the Web browser (also known as a Web client) requests files from the Web server at the company or organization that operates the Web site. Using the Internet as the transportation medium, the request is formatted by the browser using HTTP and sent to the server computer. A moment later, when the server receives the request, it retrieves the file containing the Web page or other information that the client requested, formats it using HTTP, and sends it back to the client over the Internet.
Two-tier client/server architecture: Typical request message from a client to a server consists of three major parts: request line, optional request headers, and an optional entity body.

Three-tier and N-tier client/server architectures: A three-tier architecture extends the two-tier architecture to allow additional processing to occur before the Web server responds to the Web client’s request. Higher-order architectures - that is, those that have more than three tiers - are usually called n-tier architectures.

Software for Web Servers

In this section, you will learn about the operating system software used on most Web servers, the Web server software itself, and other programs, such as Internet utilities and e-mail software that companies often run on Web servers or other computers as part of electronic commerce operations.

Operating Systems for Web Servers

Most Web servers run on computers that use one of the following operating systems: Microsoft Windows NT Server, Microsoft Windows 2000 or 2003 Server products, Linux, or one of several UNIX-based operating systems, such as Solaris or FreeBSD. Many companies believe that Microsoft server products are simpler for their information systems staff to learn and use than UNIX-based systems. Other companies worry about the security weaknesses caused by the tight integration between application software and the operating system in Microsoft products. UNIX-based Web servers are more popular, and many users believe that UNIX is a more secure operating system on which to run a Web server.

Web Server Software

This section describes the most popular Web server programs in use today: Apache HTTP Server, Microsoft Internet Information Server (IIS), and Sun Java System Web Server (JSWS) (often called by its former names, Sun ONE, iPlanet Enterprise Server, and Netscape Enterprise Server). These popularity rankings were accumulated through surveys done by Netcraft, a networking consulting company in Bath, England, known throughout the world for its Web server survey. Netcraft conducts continual surveys to tally the number of Web sites in existence and measure the relative popularity of Internet Web server software.

| Web Server Software: | Apache HTTP Server: Has dominated the Web since 1996 because it is free and performs very efficiently. It is powerful enough that IBM includes it in its WebSphere application server package. |
Microsoft Internet Information Server: Comes bundled with current versions of Microsoft Windows Server operating systems. IIS is used on many corporate intranets because many companies have adopted Microsoft products as their standard products. Small sites running personal Web pages use IIS, as do some of the largest electronic commerce sites on the Web.

Sun Java System Web Server: Supports dynamic application development for server-side applications. Sun JSWS provides connectivity to a number of database products as well.

Quick Quiz

1. A(n) _____ is a Web page whose content is shaped by a program in response to user requests.
   Answer: dynamic page

2. _____ is nonstatic information constructed in response to a Web client’s request.
   Answer: Dynamic content

3. A(n) _____ is any computer used to provide (or “serve”) files or make programs available to other computers connected to it through a network.
   Answer: server

4. _____ is developed by a community of programmers who make the software available for download at no cost.
   Answer: Open-source software

Electronic Mail (E-Mail)

Electronic mail originated in the 1970s on the ARPANET. Although the goals of the ARPANET were to control weapons systems and transfer research files, general communications uses emerged on the network. In 1972, Ray Tomlinson, an ARPANET researcher, wrote a program that could send and receive messages over the network.

E-Mail Benefits
One useful feature of e-mail is that documents, pictures, movies, worksheets, or other information can be sent along with the message itself. These attachments are frequently the most important part of the message.

**E-mail Drawbacks**

Despite its many benefits, e-mail does have some drawbacks. One annoyance associated with e-mail is the amount of time that businesspeople spend answering their e-mail today. A second major irritation brought by e-mail is the computer virus, more simply known as a virus, which is a program that attaches itself to another program and can cause damage when the host program is activated.

**Spam**

The sheer magnitude of the spam problem is hard to believe. During one 24-hour period in 2005, researchers estimated that 106 billion spam e-mail messages were sent. Many researchers who track the growth in spam believe that current trends will continue and that more than 90 percent of all e-mail messages (including messages transmitted to both business and personal users) will be spam before any effective technical solutions can be implemented. Other researchers believe that the growth of spam is showing signs of leveling out.

**Solutions to the Spam Problem**

One way individuals can limit spam is to reduce the likelihood that a spammer can automatically generate their e-mail addresses. Many organizations create e-mail addresses for their employees by combining elements of each employee’s first and last names. For example, small companies often combine the first letter of an employee's first name with the entire last name to generate e-mail addresses for all employees at small companies (larger companies often must use more complex algorithms as they are likely to have both a Jane Smith and a Judy Smith working for them). Any spam sender able to obtain an employee list can generate long lists of potential e-mail addresses using the names on the list. If no employee list is available, the spam sender can simply generate logical combinations of first initials and common names. The cost of sending e-mail is so low that a spammer can afford to send thousands of emails to randomly generated addresses in the hope that a few of them are valid. By using an e-mail address that is more complex, such as xq7vy23@mycompany.com, individuals can reduce the chances that a spammer can randomly generate his or her address. Of course, such an address is hard to remember, which somewhat defeats the purpose of e-mail as a convenient way to communicate.

A second way to reduce spam is to control the exposure of an e-mail address. Spammers use software robots to search the Internet for character strings that include the “@” character (which appears in every e-mail address). These robots search Web pages, discussion boards, chat rooms, and any other online source that might contain e-mail addresses. Some individuals use multiple e-mail addresses to thwart spam. They use one address for display on a Web site, another to register for access to Web sites, another for shopping accounts, and so on. If a spammer starts using one of these addresses, the individual can stop using it and switch to another.
Web Site and Internet Utility Programs

In earlier chapters, you learned how companies are using e-mail as a key element in their electronic commerce strategies. In this section, you will learn some of the more significant technical details of how e-mail works.

Finger and Ping Utilities

Finger is a program that runs on UNIX operating systems and allows a user to obtain some information about other network users. A Finger command yields a list of users who are logged on to a network, or reports the last time a user logged on to the network. Many organizations have disabled the Finger command on their systems for privacy and security reasons.

A program called Ping, short for Packet Internet Groper, tests the connectivity between two computers connected to the Internet. Ping provides performance data about the connection between Internet computers, such as the number of computers (hops) between them. It sends two packets to the specified address and waits for a reply. Network technicians often use Ping to troubleshoot Internet connections.

Tracert and Other Route-Tracing Programs

Tracert (TRACE RouTe) sends data packets to every computer on the path (Internet) between one computer and another computer and clocks the packets’ roundtrip times. This provides an indication of the time it takes a message to travel from one computer to another and back, ensures that the remote computer is online, and pinpoints any data traffic congestion. Route-tracing programs also calculate and display the number of hops between computers and the time it takes to traverse the entire one-way path between machines.

Route-tracing programs such as Tracert work by sending a series of packets to a particular destination. Each router along the Internet path between the originating computer and the destination computer reports its IP address and the time it took to reach it. After the program completes its packet transmissions, it displays the number of hops and how much time it took to reach each node and travel the entire path.

Issues Box: CAMELOT Messaging Security

CAMELOT is a server side based spam filtering solution that behaves very much like a firewall that sits in front of the SMTP Mail Server. Messages are received through the CAMELOT gateway and are forwarded to a standard mail server after they are successfully processed. The SMTP handler verifies all the parameters of the SMTP transaction and denies
all further processing as soon as an invalid incident occurs. This prevents unauthorized access before the message is delivered to the mail server.

All messages are then subjected to an extensive content examination. The CAMELOT Content Processor allows for several methods of analysis and supports configurable properties and the order of execution can also be individually adjustable. Different Policy profiles for each network segment and/or domain are used to perform detailed content processing for all types of requirements.

The content examination process works with the combination of multiple analysis technologies. Unwanted words and phrases, as well as dangerous HTML references, can be detected automatically by special text pattern analysis.

Bayesian classification methods are used to rate incoming messages and process them according special rules. The help of artificial intelligence in self-learning filter systems can reduce the portion of unwanted messages to a minimum. Also pornographic contents in images can be detected in this way.


Questions

- Do you have any reservations about this system?
- What other solutions are available to fight spam?

Telnet and FTP Utilities

Telnet is a program that allows users to log on to a computer that is connected to the Internet. This remote login capability can be useful for running older software that does not have a Web interface. Several Telnet client programs are available as free downloads on the Internet, and Microsoft Windows systems include a Telnet client called Telnet.exe. Telnet lets a client computer give commands to programs running on a remote host. Telnet programs use a set of rules called the Telnet protocol.

The File Transfer Protocol (FTP) is the part of the TCP/IP rules that defines the formats used to transfer files between TCP/IP-connected computers. FTP can transfer files one at a time, or it can transfer many files at once. FTP also provides other useful services, such as displaying remote and local computers’ directories, changing the current client’s or server’s active directory, and creating and removing local and remote directories.

Indexing and Searching Utility Programs
Search engines and indexing programs are important elements of many Web servers. Search engines or search tools search either a specific site or the entire Web for requested documents. An indexing program can provide full-text indexing that generates an index for all documents stored on the server.

**Data Analysis Software**

Web servers can capture visitor information, including data about who is visiting a Web site (the visitor’s URL), how long the visitor’s Web browser viewed the site, the date and time of each visit, and which pages the visitor viewed. This data is placed into a Web log file. As you can imagine, the file grows very quickly - especially for popular sites with thousands of visitors each day.

**Link-Checking Utilities**

Dedicated site management tools include a standard set of features, starting with link checking. A link checker examines each page on the site and reports on any URLs that are broken, seem broken, or are in some way incorrect. It can also identify orphan files. An orphan file is a file on the Web site that is not linked to any page. Other important site management features include script checking and HTML validation. Some management tools can locate error-prone pages and code, list broken links, and e-mail maintenance results to site managers.

On the company Web site, it is important to regularly check links that point to pages both within and outside the corporate Web site. Some Web server software does contain link-checking features. A dead link, when clicked, displays an error message rather than a Web page. Maintaining a site that is free of dead links is vital because too many dead links on a site can cause visitors to jump to another site. Web-browsing customers are just a click away from going to a competitor’s site if they become annoyed with an errant Web link.

**Remote Server Administration**

With remote server administration, a Web site administrator can control a Web site from any Internet-connected computer. Although all Web sites provide administrative controls - most through a workstation computer on the same network as the server computer or through a Web browser - it is convenient for an administrator to be able to fix the server from wherever he or she happens to be.

**Web Server Hardware**

Companies use a wide variety of computer brands, types, and sizes to host electronic commerce operations. Some small companies can run Web sites on desktop PCs. Most electronic commerce Web sites are operated on computers designed for site hosting, however.

**Server Computers**
Web server computers generally have more memory, larger (and faster) hard disk drives, and faster processors than the typical desktop or notebook PCs with which you are probably familiar. Many Web server computers use multiple processors; very few desktop PCs have more than one processor. Because Web server computers use more capable hardware elements and more of these elements, they are usually much more expensive than workstation PCs.

Today, a high-end desktop PC with 2 GB of RAM, a 3.6 GHz processor, a fast 400 GB SATA disk drive, a good monitor, and a complement of DVD/CD-RW drives costs between $2000 and $4000.

**Web Server Performance Evaluation**

Benchmarking Web server hardware and software combinations can help in making informed decisions for a system. Benchmarking, in this context, is testing that is used to compare the performance of hardware and software.

Elements affecting overall server performance include hardware, operating system software, server software, connection speed, user capacity, and type of Web pages being delivered. When evaluating Web server performance, a company should know exactly what factors are being measured and ensure that these are important factors relative to the expected use of the Web server. Another factor that can affect a Web server’s performance is the speed of its connection. A server on a T3 connection can deliver Web pages to clients much faster than on a T1 connection.

The number of users the server can handle is also important. This can be difficult to measure because results are affected by the server’s line speed, the clients’ line speeds, and the sizes of the Web pages delivered. Two factors to evaluate when measuring a server’s Web page delivery capability are throughput and response time. Throughput is the number of HTTP requests that a particular hardware and software combination can process in a unit of time. Response time is the amount of time a server requires to process one request. These values should be well within the anticipated loads a server can experience, even during peak load times.

**Web Server Hardware Architectures**

Administrators of large Web sites must plan carefully to configure their Web server computers, which can number in the hundreds or even thousands, to handle the daily Web traffic efficiently. These large collections of servers are called server farms because the servers are often lined up in large rooms, row after row, like crops in a field. One approach, sometimes called a centralized architecture, is to use a few very large and fast computers. A second approach is to use a large number of less powerful computers and divide the workload among them. This is sometimes called a distributed architecture or, more commonly, a decentralized architecture.

| Teaching Tip: | ♦ Research the different types of server hardware. What are the advantages and disadvantages of each type of server? Focus on the software. What type of businesses would be best served by each type of software? |
Quick Quiz

1. A program called _____ tests the connectivity between two computers connected to the Internet.
   Answer: Ping, Packet Internet Groper

2. _____ is electronic junk mail and can include solicitations, advertisements, or e-mail chain letters.
   Answer: Spam

3. A(n) _____ looks for From addresses in incoming messages that are known to be spammers.
   Answer: black list spam filter

4. _____ is a program that allows users to log on to a computer that is connected to the Internet.
   Answer: Telnet

5. A(n) _____ examines each page on the site and reports on any URLs that are broken, seem broken, or are in some way incorrect.
   Answer: link checker

Discussion Questions

- Discuss the financial implications of spam.
- How is benchmark testing conducted?

Additional Resources

- Spam: http://thechronicleherald.ca/Science/488086.html
- Telnet: http://en.wikipedia.org/wiki/Telnet
- How computer viruses work: http://computer.howstuffworks.com/virus1.htm

Key Terms

- **Anonymous FTP**: Allows users to log on as guests.
- **Benchmarking**: Testing that is used to compare the performance of hardware and software.
➢ **Computer virus:** A program that attaches itself to another program and can cause damage when the host program is activated.

➢ **Finger:** A program that runs on UNIX operating systems and allows a user to obtain some information about other network users.

➢ **Link checker:** Examines each page on the site and reports on any URLs that are broken, seem broken, or are in some way incorrect.

➢ **Open-source software:** Developed by a community of programmers who make the software available for download at no cost.

➢ **Orphan file:** A file on the Web site that is not linked to any page.

➢ **Ping:** Tests the connectivity between two computers connected to the Internet.