3. Web Fundamentals

What is the World Wide Web?

WWW or World Wide Web is a global information system where documents and other resources are accessible through this interconnected network of nodes (internet). These resources are generally interconnected through hypertext links identified by Uniform Resource Locators (URL). The web consists of several elements such as:

- Structuring language: HTML handles this. It's not a programming language but rather a markup language that creates the skeleton of a website, shaping each of its pages.
- Communication protocols: Primarily HTTP and HTTPS, which enable the loading and downloading of web pages or entire websites operating on a client-server request-response model.
- URL: They act as addresses that help locate every available resource on the network.
- Browsers: These are generally graphical applications through which users interact with resources on the web. Their main function is interpreting HTML, CSS, and client-side programming languages like JavaScript to present web content in a simple and visually appealing way.

It's important to mention that the internet and the web are not the same, but they are infrastructures that currently depend on each other and function under the same paradigm.

What is a domain and its types?

A domain is a unique identifier that cannot be repeated on the internet. It serves as an address that simplifies access to specific content, eliminating the need to memorize numeric sequences or IP addresses. Instead, access is facilitated through recognizable character strings.
Every server exposed to the internet has a unique IP address, and remote access requires entering this IP address. This is where DNS comes into play, translating these numerical addresses into human-readable domain names.
Domains fall into several categories:

- TLD or Top-Level Domain: The highest level of domains in the internet hierarchy, representing the last segment of a domain name located after the final dot.
- SLD or Second-Level Domain: These are one level below TLDs. SLDs are typically registered by individuals or entities and form the unique part of the domain name.
- Subdomain: These rank below TLDs and SLDs in the hierarchy and can be repeated across different domains.

Components of a URL

Example URL: https://www.example.com:8080/blog/article?id=123#comments
A URL (Uniform Resource Locator) is a unique address that points to a specific resource and is composed of the following parts:

- Scheme/Protocol: Specifies the method of access to obtain a resource.
- Delimiter segment: Represented by ://, it tells the browser how to interpret the following information.
- Domain/IP: Identifies the destination host where the resource resides.
- Port: An optional component, used only to force a connection through a specific port (:8080).
- Path: Indicates the specific location of the requested resource, often composed of multiple segments separated by slashes (/).
- Parameters: Provide additional information for the server request, preceded by a question mark (?).
- Fragment/Anchor: Represented by a hash symbol (#), typically used to link to specific sections of a document.

Differences between a website and a webpage

A webpage is a single document on the World Wide Web accessible through a browser, each identified by a unique URL. In contrast, a website is a collection of related webpages grouped under a single domain.

Hosting, VPS, and Dedicated Servers

- Hosting: The most common option for small websites. Multiple websites share the same physical server and hardware resources.
- Dedicated Server: Provides all its resources to a single client.
- VPS (Virtual Private Server): Operates independently with allocated resources within a single physical device.

Frontend and Backend and Their Vulnerabilities

Frontend and Its Vulnerabilities

The frontend is the part of the application users directly interact with. Key technologies include HTML, CSS, and JavaScript.
Common frontend vulnerabilities include:

- Cross-Site Scripting (XSS)

- Clickjacking
- Cross-Site Request Forgery (CSRF)

Backend and Its Vulnerabilities

The backend handles data processing, user interaction management, and application logic. It also manages communication between databases and the frontend. Common backend technologies include Java, Python, Ruby, PHP, and relational or non-relational databases, often with APIs.
Common backend vulnerabilities include:

- SQL Injection (SQLi)
- Remote Code Execution (RCE)
- API access and abuse
- Information leakage due to misconfigurations

Major Web Servers

Web servers are software that accept HTTP and HTTPS requests and respond by delivering content. Their main functions include processing requests, managing connection security, and maintaining activity logs. Major web servers include:

Apache

One of the most widely used web servers, written in C and based on a modular structure, allowing only necessary modules to be loaded. For example, mod_security acts as a firewall.

Nginx

Designed for high-traffic environments, load balancing, reverse proxying, and efficient delivery of static content.

IIS

Developed by Microsoft, primarily running on Windows Server. It's used for applications requiring Microsoft-specific technologies like ASP.net.

Apache Tomcat

Implements Java specifications, particularly Java Servlet and JavaServer Pages (JSP), and is mainly used to run Java applications.

Jetty

A lightweight server known for consuming minimal resources.

Major Content Management Systems (CMS)

CMS platforms have gained popularity for enabling people with limited technical knowledge to create websites. Major CMSs include:

WordPress

Originally created for blogging, WordPress evolved to power nearly 40% of the world's websites. It uses PHP and MySQL/MariaDB for database management.

Joomla

Similar to WordPress, using PHP and MySQL, but considered less user-friendly.

Drupal

Designed for scalable website structures.

HTTP Protocol Information Transmission Methods

These define the actions to perform on a resource identified by a URL:

- GET: Requests representation of a specific resource without changing its state.
- POST: Sends data to the server to create a new resource.
- PUT: Updates or creates a resource on the server.
- DELETE: Removes the specified resource.
- PATCH: Partially modifies a resource.
- HEAD: Requests only metadata without the response body.
- TRACE: Shows what an intermediate server receives in a request chain.
- CONNECT: Establishes a transparent communication channel, often for encrypted HTTPS connections via proxy.
- OPTIONS: Lists available communication methods for a target resource.

Major HTTP Protocol Headers

Essential components of HTTP requests and responses, transmitting additional information between client and server:

- Host: Identifies the domain and port the client wants to access.
- User-Agent: Identifies the client's browser and operating system.
- Accept: Specifies acceptable content types.

- Accept-Language: Indicates the user's preferred language.
- Accept-Encoding: Specifies supported data compression formats.
- Content-Type: Defines the media type of the request or response body.
- Content-Length: Specifies the body's size in bytes.
- Cookie: Transmits client-stored cookies for session management.
- Connection: Determines whether the connection stays open or closes after the request.
- Referer: Indicates the source URL of the current request.
- Authorization: Provides client authentication credentials.
- Cache-Control: Manages caching behavior for requests and responses.
- X-Forwarded-For: Reveals the client's original IP address through proxies.
- Set-Cookie: Instructs the client to store cookies with attributes like expiration, path, and domain.
- X-XSS-Protection: Configures the browser's XSS attack prevention filter.

HTTP Status Codes

Quickly indicate the result of a client's request:

- 1XX: Informational responses (e.g., 100 Continue)
- 2XX: Successful responses (e.g., 200 OK, 201 Created)
- 3XX: Redirections (e.g., 301 Moved Permanently, 302 Found)
- 4XX: Client errors (e.g., 400 Bad Request, 401 Unauthorized, 404 Not Found)
- 5XX: Server errors (e.g., 500 Internal Server Error, 502 Bad Gateway, 503 Service Unavailable)