# 網頁程式設計 Web Programming

Ch1: Introduction (WWW and HTTP)

#### **Outline**

- Why Study Web Programming?
- World Wide Web
- Hypertext Transfer Protocol

REF: 新觀念PHP7+MySQL+AJAX網頁設計範例教本(第五版),旗標。

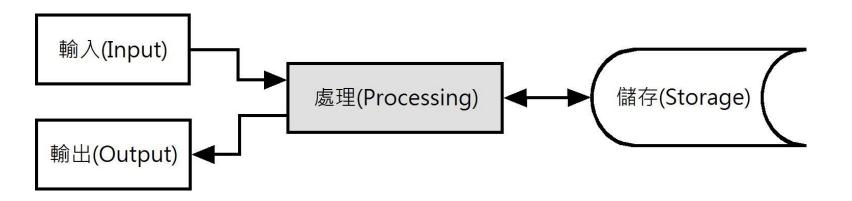
REF: 輕鬆玩HTML5+CSS3+JavaScript網頁程式設計,全華。

## Why Study Web Programming?

- A web application basically includes user-side (presenting text, figure, video, gaming...) and server-side (processing actions) these two parts.
- In the past, web application can only be running at the server-side, presenting its exportation results on the client-side browser (now it can be mutual).
- Most web applications aim to response client requests and make interaction. For instance, Facebook, Plurk, and Twitter are providing such scenario.
- You may use C/C++, PHP, Perl, Python, Ruby, and other web programming languages to develop the web application.

## Why Study Web Programming? (cont.)

- Web application is one of Information Processing Systems, with input, output, and processing actions.
- The invention of web has improved the information dissemination in the world.



#### Why Study Web Programming? (cont.)

- Basic elements for enabling the web service:
- Webpage content (text, figure, video...)
- Web application (collaborate with database, accounting system...)
  - Programming languages (like Ruby, JavaScript...)
  - Web frameworks (like Ruby on Rails, NodeJS...)
- Webserver software (Apache, IIS, Tomcat, Nginx...)
- Webhosting (PC, Server, Virtual Machine, VPC w/ cPanel...)
- Presenting to the client → front-end
- Others → back-end

(前後端簡略的區分方式)

#### **World Wide Web**

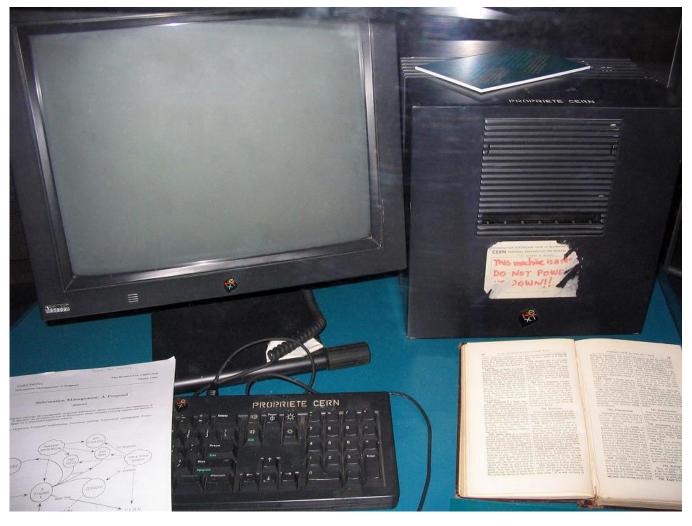
- World Wide Web (WWW, 全球資訊網), called **web** in short, is an information system that invented by research team of CERN (歐洲核子研究組織), leading by Sir Timothy Berners-Lee in 1989.
- In WWW, it uses Uniform Resource Locator (**URL**, 統一資源定位符) and Uniform Resource Identifier (**URI**, 統一資源識別碼) to identified available resources, and these resources can be inter-linked by **hyperlink** (超鏈結).





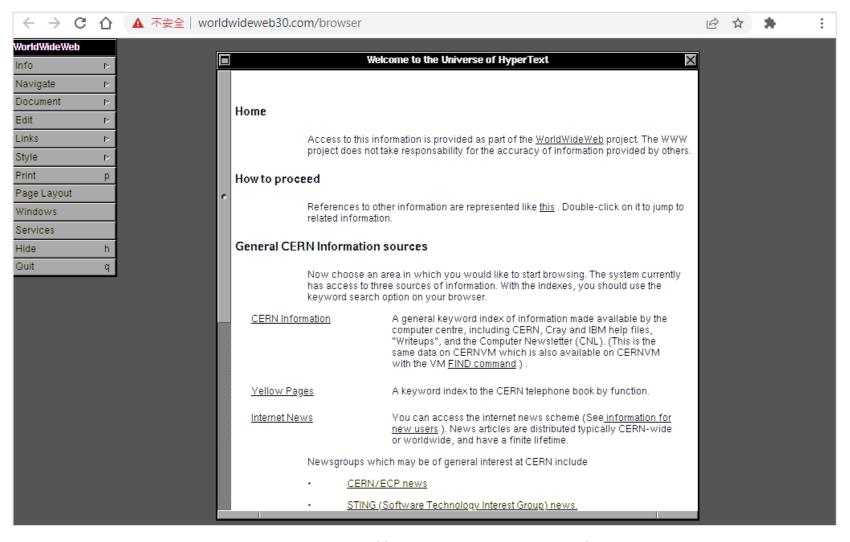
REF: https://zh.wikipedia.org/wiki/File:Where\_the\_WEB\_was\_born.jpg

# World's First Web Server (at CERN)



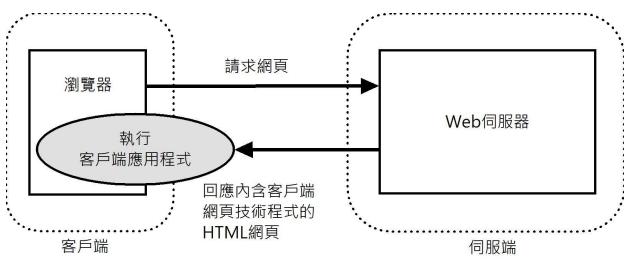
REF: https://en.wikipedia.org/wiki/CERN\_httpd#/media/File:First\_Web\_Server.jpg

#### CERN 2019 WorldWideWeb 30th Rebuild

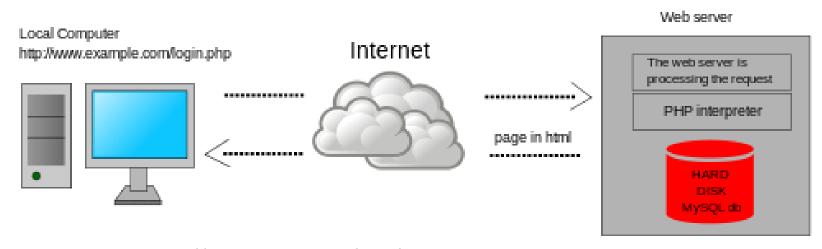


REF: https://worldwideweb30.com/

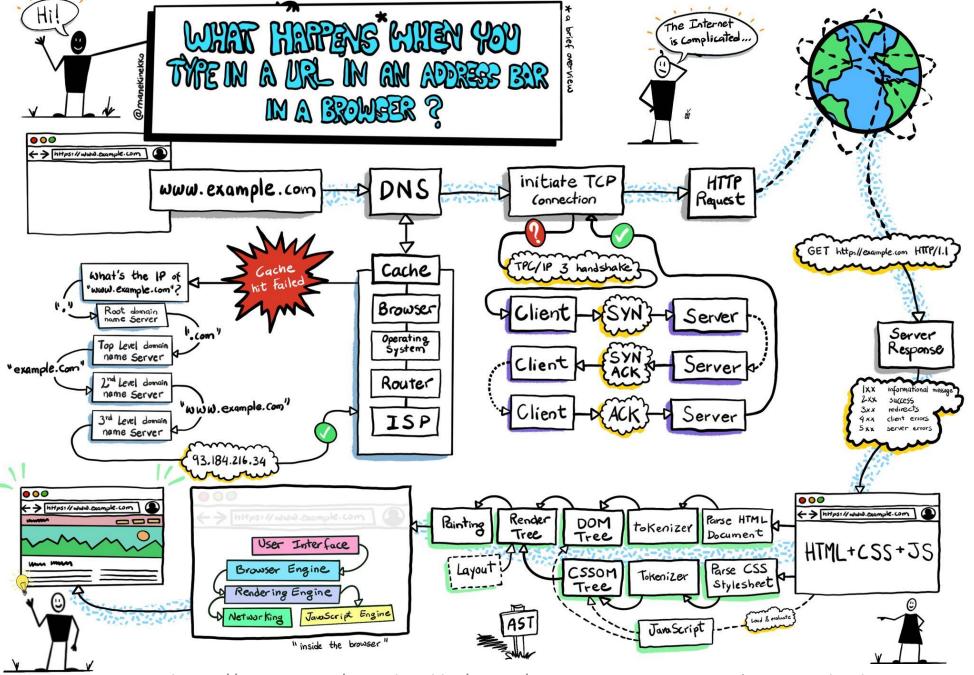
- Hypertext Markup Language (HTML, 超文本標記語言) is a standard language used to create webpage. A webpage may involve html and additional information (e.g., Cascading Style Sheets/CSS and JavaScript/JS) to present its content.
- The resources on web are transferred via the **Hypertext Transfer Protocol** (HTTP, 超文本傳輸協定). By using **web browser** (網頁瀏覽器), user are able to access web resources through the Internet.

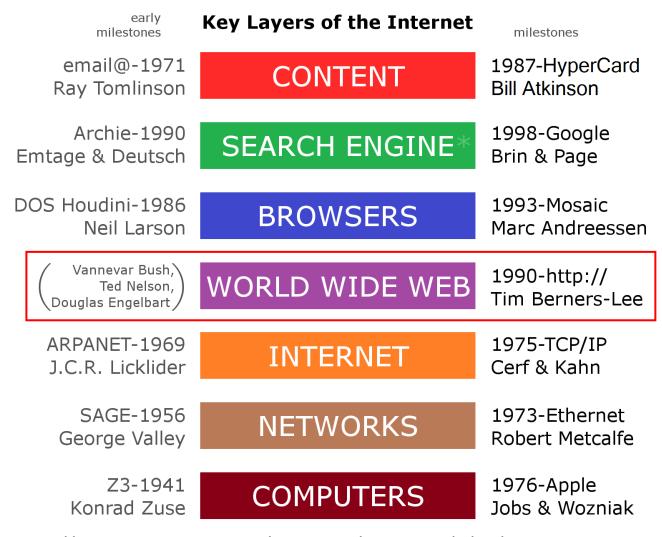


- Website aggregates resources in web server, enabling web service—a daemon that responds to requests made over the Internet from web client(s).
- Web server and client leverages HTTP to make communication, which are application-level protocols in the layer-7 of Open Systems Interconnection (OSI) reference model.



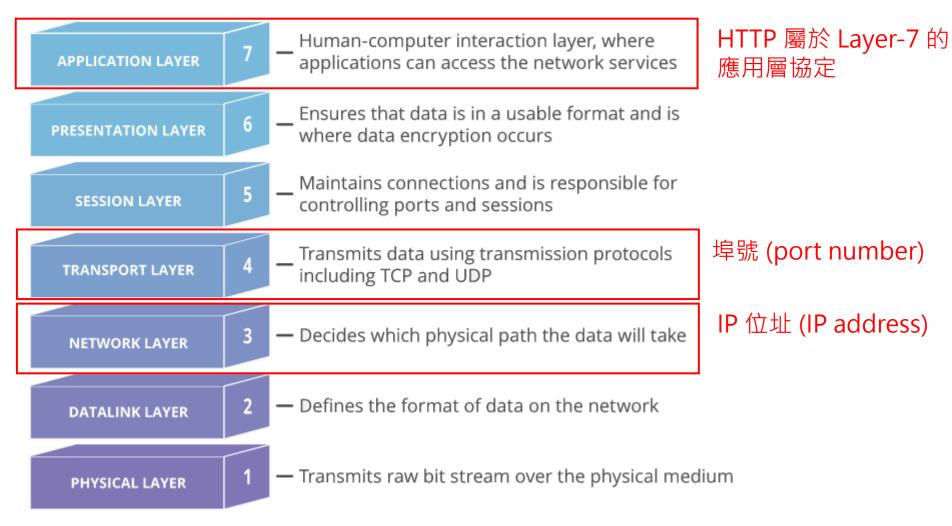
REF: https://en.wikipedia.org/wiki/File:Scheme\_dynamic\_page\_en.svg





REF: https://upload.wikimedia.org/wikipedia/commons/3/39/Internet\_Key\_Layers.png

## **Open Systems Interconnection (OSI) Reference Model**



REF: https://www.cloudflare.com/zh-tw/learning/ddos/what-is-layer-7/

## **Hypertext Transfer Protocol**

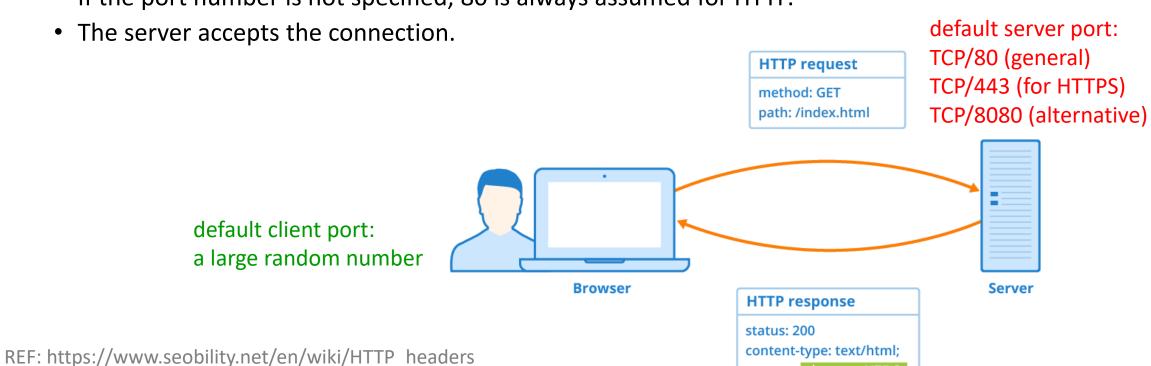
- HTTP is the general way of making data communication to transmit web resources through the network.
- A simple interaction between HTTP server and client is shown below:



- HTTP Requests for Comments (RFCs) has became a standard of Internet Engineering Task Force (IETF) and the World Wide Web Consortium (W3C). The first documented version of HTTP was HTTP/0.9 in 1991.
- Dave Raggett led the HTTP Working Group (HTTP WG) in 1995 and wanted to expand the protocol. The initial work was to document existing practice and short-term extensions. Subsequent work was to extend and revise the protocol.
- The WG directions include improving efficiency, extending operations & negotiation, getting richer meta-information, and adding security concepts.

REF: https://www.w3.org/Protocols/HTTP/AsImplemented.html

- Connection
  - The client makes a TCP-IP connection to the host using the domain name or IP address, and the port number given in the address.
  - If the port number is not specified, 80 is always assumed for HTTP.



charset= UTF-8

#### Request

- The client sends a document request consisting of a line of ASCII characters. A well-behaved server will not require the carriage return character.
- This request consists of the word "GET", a space, the document address, omitting the "http:, host and port parts when they are the coordinates just used to make the connection.
- The document address will consist of a single word.
- The search functionality of the protocol lies in the ability of the addressing syntax to describe a search on a named index .
- A search should only be requested by a client when the index document itself has been described as an index with tag .

 A request message from a client to a server, which includes the type of that message, the method used to be applied for the resource, the identifier of the resource, and the protocol version in use.

```
Request = Simple-Request | Full-Request | Method = "GET" | "HEAD" | "HEAD" | "POST" | extension-method | Request-Line | Request-Header | Entity-Header | Entity-Body | Entity-Body | Entity-Body |
```

- GET (用於取得資料)
  - The GET method means retrieve whatever information (in the form of an entity) is identified by the Request-URI.
  - GET /TheProject.html HTTP/1.0
     GET /pub/WWW/TheProject.html HTTP/1.0

- HEAD (類似 GET 但只需要使用標頭回應, 例如用於判斷資源是否存在)
  - The HEAD method is identical to GET except that the server must **not return any Entity-Body** in the response.
  - The meta information contained in the HTTP headers in response to a HEAD request should be identical to the information sent in response to a GET request.

- POST (上傳資料, 資料存放位置和如何接收處理由伺服器端決定)
  - The POST method is used to request that the destination server accept the entity enclosed in the request as a new subordinate of the resource identified by the Request-URI in the Request-Line.
  - POST is designed to allow a uniform method to cover the following functions:
    - Annotation of existing resources
    - Posting a message to a bulletin board, newsgroup, mailing list, or similar group of articles
    - Providing a block of data, such as the result of submitting a form, to a data-handling process
    - Extending a database through an append operation
- PUT, DELETE, TRACE, OPTIONS, CONNECT (also defined in RFC 7231)
- PATCH (defined in RFC 5789)

#### Response

- The response to a simple GET request is a message in hypertext mark-up language (HTML).
- It is a byte stream of ASCII characters.
- Lines shall be delimited by an optional carriage return followed by a mandatory line feed character. The client should not assume that the carriage return will be present.
- The format of the message is HTML. This format allows for menus and hit lists to be returned as hypertext. It also allows for plain ASCII text to be returned following the PLAINTEXT tag.
- The message is terminated by the closing of the connection by the server.
- Well-behaved web client will read the entire document as fast as possible. Web client should not wait for human-user's action before reading the whole of the document.
- The server may impose a timeout of the order on inactivity.
- Error responses are supplied in human readable text in HTML syntax.

- Status-Code in **response message** 
  - The Status-Code element is a 3-digit integer result code of the attempt to understand and satisfy the request. Web client is not required to examine or display the Reason-Phrase for human-user.
- 1xx: Informational Not used, but reserved for future use
- 2xx: Success The action was successfully received, understood, and accepted.
- 3xx: Redirection Further action must be taken in order to complete the request
- 4xx: Client Error The request contains bad syntax or cannot be fulfilled
- 5xx: Server Error The server failed to fulfill an apparently valid request

```
"200"
Status-Code
                          ; OK.
                  "201"
                          ; Created
                  "202"
                          ; Accepted
                  "204"
                          ; No Content
                  "301"
                          ; Moved Permanently
                  "302"
                          ; Moved Temporarily
                          ; Not Modified
                  "304"
                  "400"
                            Bad Request
                  "401"
                          ; Unauthorized
                          ; Forbidden
                  "403"
                  "404"
                          ; Not Found
                  "500"
                          ; Internal Server Error
                  "501"
                          ; Not Implemented
                          ; Bad Gateway
                  "502"
                          ; Service Unavailable
                  extension-code
```

Web Client (request)

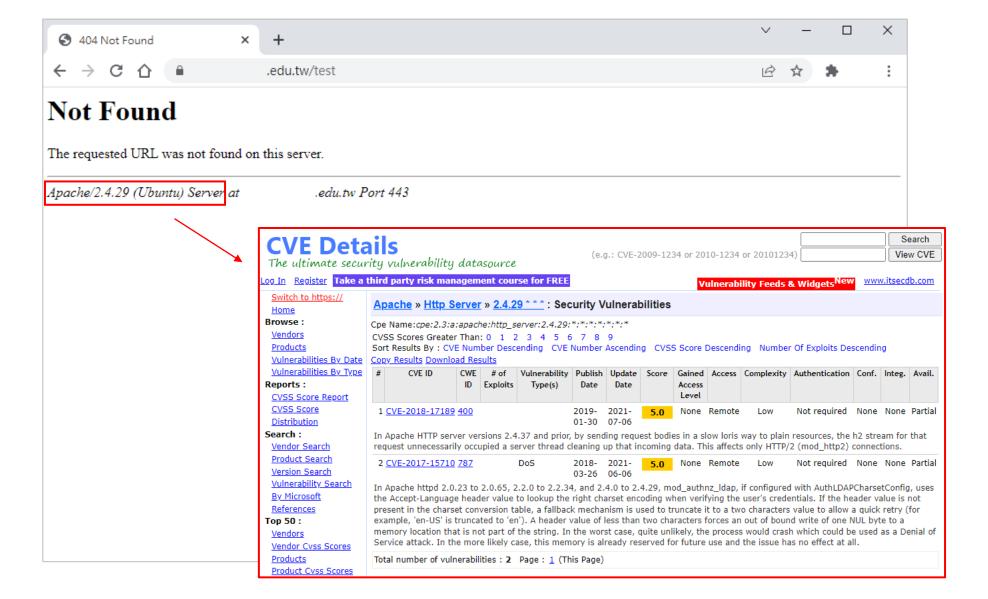
```
GET / HTTP/1.1
Host: www.example.com

網址會先透過 DNS 機制取得解析後的 IP 位址,若此值為IP 地址則不需要解析
```

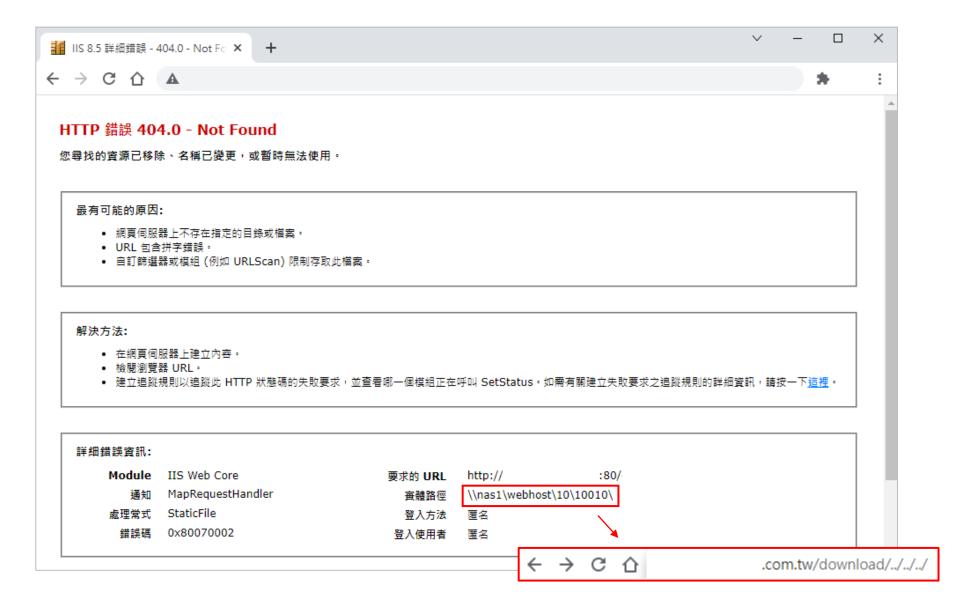
Web Server (response)

REF: https://en.wikipedia.org/wiki/Hypertext\_Transfer\_Protocol

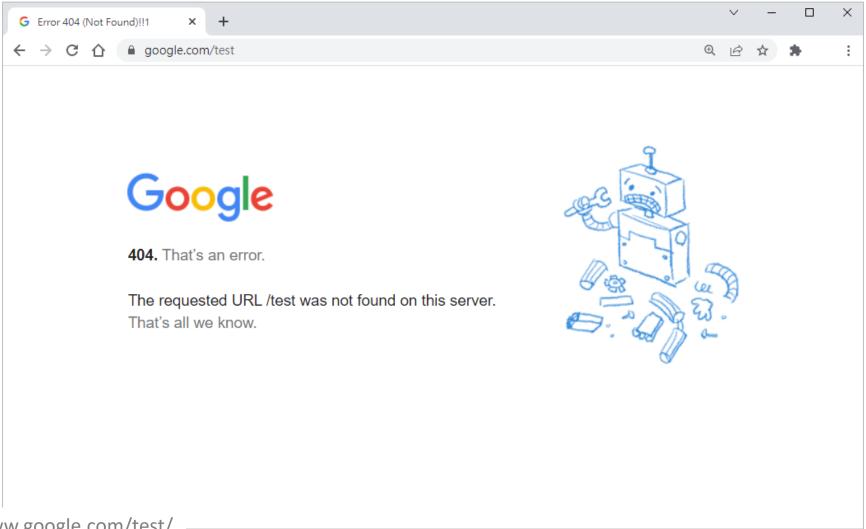
## **Status Code Leakage**



# **Status Code Leakage (cont.)**



## **Customized Response Message**



REF: https://www.google.com/test/

## **Customized Response Message (cont.)**



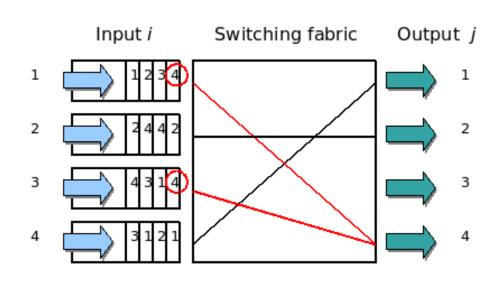
REF: https://www.nvidia.com/zh-tw/test/

#### Disconnection

- The TCP-IP connection is broken by the server when the whole document has been transferred.
- Web client may abort the transfer by breaking the connection before this, in which case the server shall not record any error condition.
- Web server need not store any information about the request after disconnection.

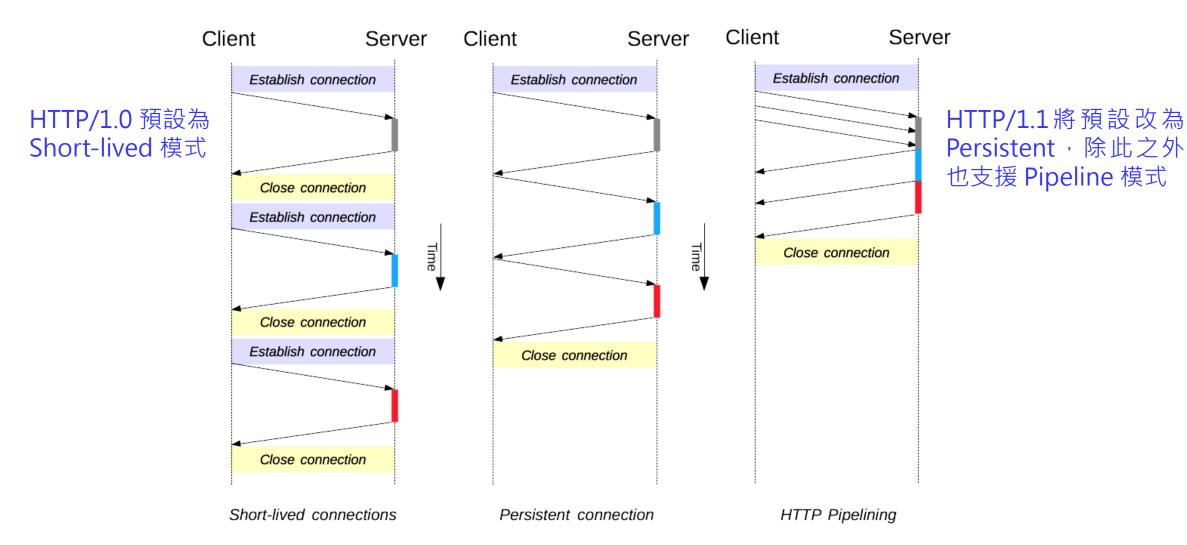
- HTTP/1.0
  - Documented in 1996 as RFC 1945, revised from HTTP/0.9
  - It separates connection for every resource request even accessing the same web server
  - With **head of line blocking** problem (HOL)

- Head of line blocking problem
- The 1st and 3rd input flows are competing to send packets to the same output interface
- If the switching fabric decides to transfer the packet from the 3rd input flow, the 1st input flow cannot be processed in the same time slot.
- The 1st input flow is blocking a packet for output interface 3, which is available for processing.



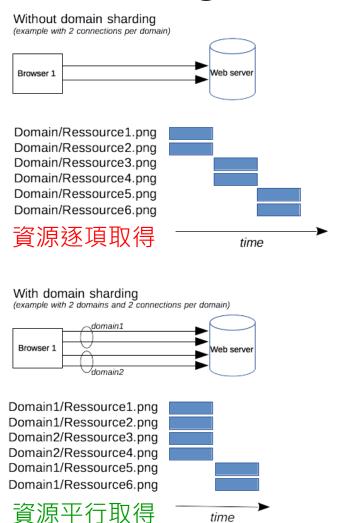
- HTTP/1.1
  - A revision of the HTTP/1.0
  - It can reuse the connection to access resource from the same web
  - "HTTP/1.1 200 OK" is the common example used to test HTTP server-client connection
  - When parallel requests in the browser is coming, the subsequent requests have to wait for the former ones to be completely served in HTTP/1.1, it still meets the head of line blocking problem

# **Connection management in HTTP/1.x**



REF: https://developer.mozilla.org/en-US/docs/Web/HTTP/Connection\_management\_in\_HTTP\_1.x

## Connection management in HTTP/1.x

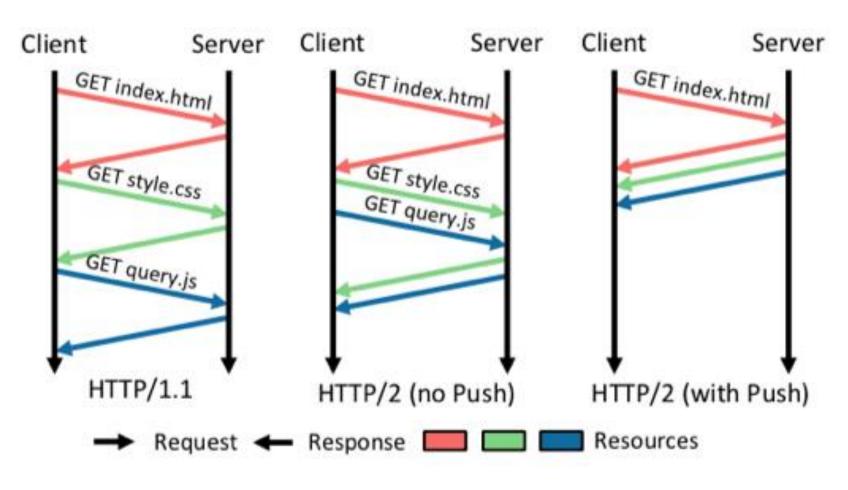


在 HTTP/1.x 使用 domain sharing 可以加快取得需要的資源 (建立很多 connection)

但在 HTTP/2 因為能更有效率地平行處理 (connection coalescing),故啟用這個功能反而可能會降低效率

- HTTP/2 (HTTP/2.0)
  - Published in 2015, and it is a major revision of the HTTP/1.1
  - Supported by most mainstream web browsers
  - Allow HTTP/2 Server Push (to respond with data for more queries than the client requested)
  - Pipelining of requests
  - Fix head of line blocking problem by multiplexing multiple requests over a single TCP connection
- SPDY (pronounced like "speedy") from HTTP/1.1 to HTTP/2
  - It was a previous HTTP-replacement protocol developed by Google (at session layer)
  - The HTTP WG considered proposed SPDY protocols to be used HTTP/1.1
  - HTTP/2 was built with SPDY concepts, each user action is given a "stream ID", providing is a single TCP channel connecting the user to the server

## **Server Push**



Server 主動將其他需要搭配的資源一起主動推送給Client,以節省 Client 送出 request 的耗時,進而達到提升效率之目的

## **Hypertext Transfer Protocol Secure (HTTPS)**

- Hypertext Transfer Protocol Secure (HTTPS) is a secured extension of the HTTP. Its communication encryption relies on SSL/TLS (HTTP over SSL/TLS).
- The SSL/TLS protocols are implemented as a transparent wrapper around the HTTP protocol in session layer.



## **Hypertext Transfer Protocol Secure (HTTPS) (cont.)**

• The motivations for HTTPS are security browsing and host authentication for the

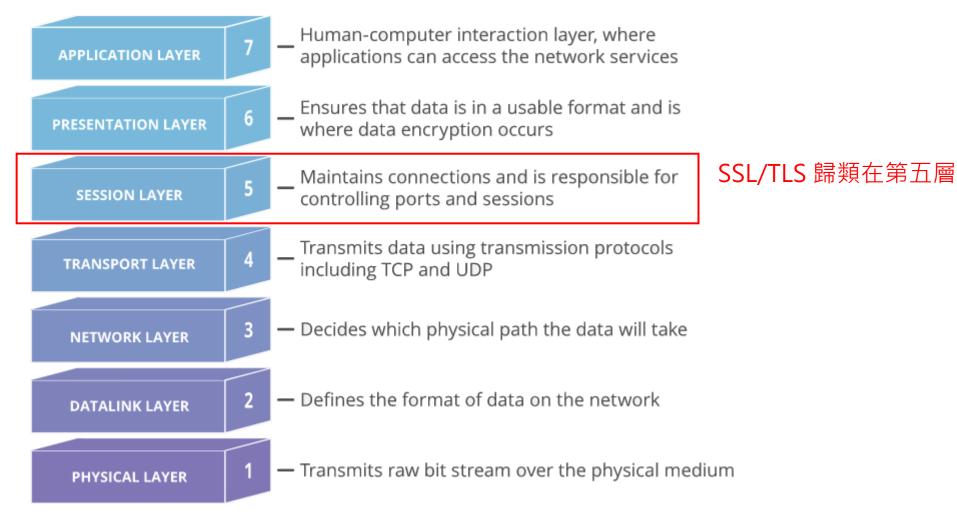
website to against MITM attack (sniffer and tamper)

 The aspect of trust websites is depending on certificate authorities (CA) build-in the browsers, such as Chrome, Firefox, and IE





## **Open Systems Interconnection (OSI) Reference Model**



REF: https://www.cloudflare.com/zh-tw/learning/ddos/what-is-layer-7/

## **Hypertext Transfer Protocol Secure (HTTPS) (cont.)**

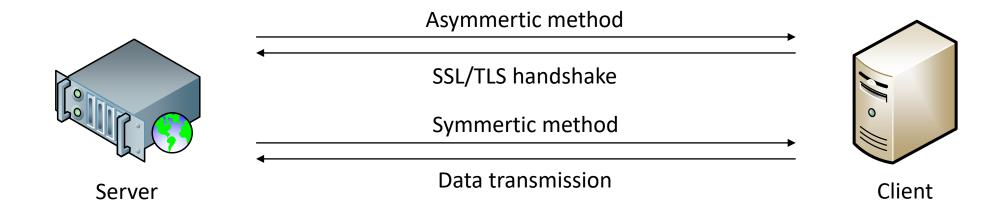
• HTTPS is an extension of HTTP, it is also referred to as HTTP over SSL/TLS.

(1-1) Client uses server's public key to encrypt message (send session key).

(1-2) Server uses its private key to decrypt message, getting client's public key, and uses this key to encrypt message (receive session key) and return an ack.

(1-3) handshake complete

(2-1) Using symmetric session key to transmission the data.



- HTTP/3
  - It is a new upcoming version of HTTP
  - The name was HTTP-over-QUIC originally
  - HTTP/3 uses QUIC technology, it aims to fix the head-of-line blocking issues completely
  - The change of migrating from legacy transmission way to QUIC speeding up the performance while it might sacrifice the reliability

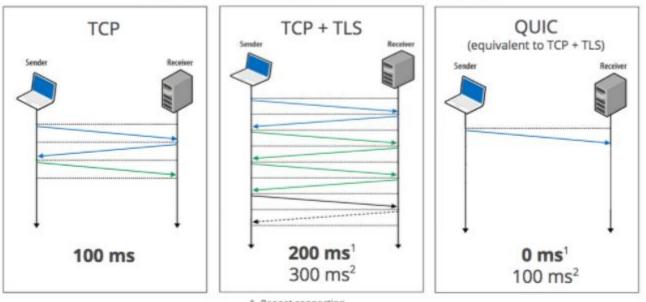


REF: https://en.wikipedia.org/wiki/HTTP/3

REF: https://solidgeargroup.com/en/quic-what-hides-behind-http-3/

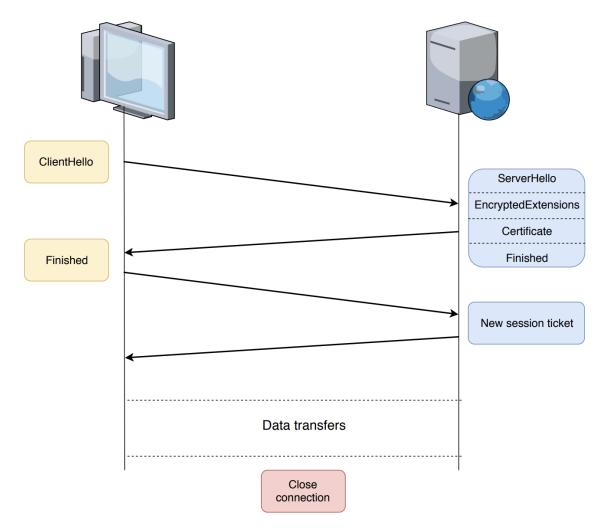
- Quick UDP Internet Connection (QUIC)
  - It is a transport layer network protocol developed initially by Google
  - Congestion control is used over User Datagram Protocol (UDP), not original Transmission Control Protocol (TCP) in HTTP/1.x and HTTP/2

#### **Zero RTT Connection Establishment**

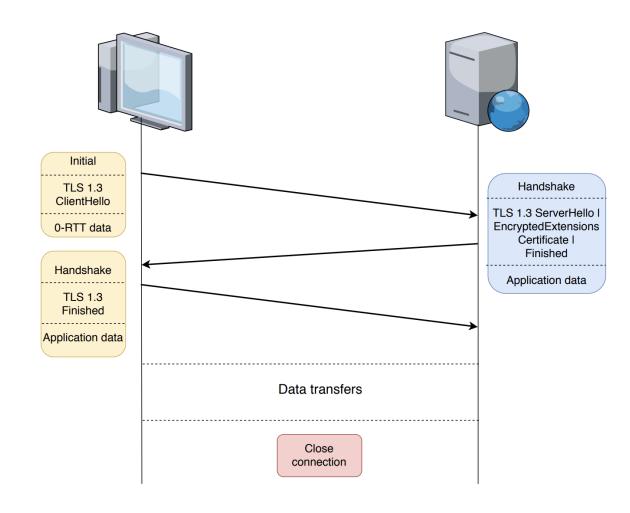


- 1. Repeat connection
- 2. Never talked to server before

- Quick UDP Internet Connection (QUIC)
  - QUIC supports a set of multiplexed connections over UDP/443 and has been designed to provide equivalent security to TLS/SSL by reducing latency in both connection and data transport
  - It makes web client-server communications to be more efficient
  - When each resource is assigned an individual conceptual flow, the transport layer may know that when a packet is lost, subsequent packets can still be used if they contain data from another resource that was not in the lost packet
  - It solves is the slow establishment of the TLS/SSL connection, which requires several information exchanges (round trip times, RTTs), while QUIC reduces the cost to practically zero RTTs



REF: https://solidgeargroup.com/en/quic-what-hides-behind-http-3/



REF: https://solidgeargroup.com/en/quic-what-hides-behind-http-3/

- However, some antivirus/firewall are still blocking QUIC by default
  - `Most firewalls have extensive functionality when dealing with HTTP and HTTPS traffic`
  - `In most architectures, when HTTP traffic is detected, it is passed on to a web protection module that performs web filtering, deep packet inspection`
  - `Enabling QUIC may cause network security appliances to be unavailable to restrict access, which also increases the risk of downloading malware when browsing the web`

Destination Address	all	<b>~ ©</b>
Schedule	always	•
Service	∠  ∠  ∠  ∠  ∠  ∠  ∠  ∠  ∠  ∠  ∠  ∠  ∠	▼ ΰ
Action	O DENY	•
Logging Options  ON Log Violation Traffic		

REF: https://kb.fortinet.com/kb/viewContent.do?externalId=FD36529

REF: https://www.fastvue.co/fastvue/blog/googles-quic-protocols-security-and-reporting-implications/

# 使用不同語言編寫的 Web Application

#### VBScript

• VBScript 是 Visual Basic 語言家族的成員,全名 Microsoft Visual Basic Scripting Edition,簡稱 VBScript。VBScript 是完全免費的直譯語言,早期的微軟 IE 瀏覽器都有支援 VBScript,能讓 網頁設計者開發互動多媒體的網頁內容,但現在已經較少看到。

### Java Applet

- 使用 Java 語言撰寫的一種 Java 應用程式,我們需要使用編譯器將原始程式碼編譯成位元組碼,即 Java Applet,瀏覽器需要使用 Java 直譯器 JVM (Java Virtual Machine) 來執行。
- 可以在網頁讀取的時候看到這樣的動畫標示



REF: https://java.com

### ActionScript and Flash

• ActionScript 是 Macromedia 公司開發的 Script 腳本語言 (後來被 Adobe 收購),可以讓 Flash 動畫電影檔產生互動效果,這是一種類似 JavaScript 語法的腳本語言。瀏覽器只需安裝 Flash 播放程式,就可以在網頁顯示 Flash 所建立的動畫效果,但 Flash 因為資安問題已經已經被主流瀏覽器停止支援,同時 Adobe 也不再提供維護。

#### Silverlight

• Silverlight 是一套用來開發豐富網際網路應用程式 (Rich Internet Application, RIAs) 的工具程式,其定位和 Macromedia 公司的 Flash 相同,屬微軟家族的 RIA 工具,提供網頁設計師另一種建立網頁內容的選擇,但由於發展因素已由微軟宣布終止維護。

#### JavaScript (JS)

• JavaScript 是 Netscape 開發的 Script 腳本語言,程式語法容易,初學者也可以快速在網頁上建立互動效果。Jscript 則為微軟推出相容 JavaScript 的 Script 語言,在過去的時代中 Netscape 或 Mozilla Firefox 支援 JavaScript 但 Internet Explorer 支援 Jscript。

#### AJAX

• AJAX 是 Asynchronous JavaScript And XML 的縮寫,譯成中文就是非同步 JavaScript 和 XML 技術,AJAX 是由 HTML 和 CSS、XML、XML DOM 及 XMLHttpRequest 等物件組成。

- ASP (Active Server Pages)
  - 英文字面上是一種讓網頁在伺服器上動起來的技術,能夠將 Script 語言內嵌 HTML 標籤的網頁,在伺服器端產生動態網頁內容,這是一種在伺服器端以直譯方式執行的網頁技術。
  - ASP.NET 是繼 ASP 3.0 後,微軟開發的伺服端網頁技術,以 CLR (Common Language Runtime) 架構建立的 .NET 程式設計平台,可以讓開法者運用該語言在伺服器端建立Web應用程式。
  - 可以在使用微軟的 IIS 加設網頁伺服器的系統常看到 .asp 或 .aspx 的網頁檔

- JSP (Java Server Pages)
  - JSP 是 Java 家族中能和 ASP 一較長短的伺服器端網頁技術,結合 HTML 和 Java Servlet;其中 Java Applet 是下載到客戶端執行的程式檔,Java Servlet 則是在伺服端執行提供服務。

#### • Node.js

• 以往 JavaScript 只能用於使用者端,但 Node.js 提供了同時運用在伺服器端和使用者端的方式,可支援密集和高互動性的應用情境。可透過 NPM 管理器取得其他開源的第三方程式整合到自己所設計的網頁中,搭配框架及模組化設計提升網頁開發與建構的速度。

- PHP (PHP: Hypertext Preprocessor)
  - 同樣是開放原始碼的伺服器端 Script 語言,可以直接內嵌於 HTML 網頁,特別適用在 Web 網站開發,主要是使用在 Unix 作業系統的伺服端網頁技術,目前也支援 Windows 作業系統。
  - PHP 可以用來取代 CGI 類型的文字應用設計 (例如將 Perl CGI 實作的功能改用 PHP 來建立), PHP 語言只需配合 Web 伺服器和瀏覽器的開發環境,就可以透過瀏覽器執行 PHP 語法,並且讓使用者的瀏覽器顯示執行結果的網頁內容。
  - PHP 亦可配合伺服端的資料庫系統建立網頁資料庫 (Web Database),直接連接和存取資料庫的記錄資料在瀏覽器顯示。
  - Linux + Apache + MySQL(MariaDB) +PHP 通稱 LAMP, 能提供基礎的 web 開發服務環境。目前 最新的主流版本雖然是 PHP8, 但仍然可以看到許多系統仍繼續停在 PHP7 環境。

## 共通匣道介面

- CGI (Common Gateway Interface)
  - 共通匣道介面提供 Web 伺服器執行外部程式的管道,CGI 應用程式是一種外部程式的執行檔, 能夠使用各種程式語言來開發,例如: Visual Basic、C、C++ 和 Perl,程式需要編譯成執行檔案,以便在伺服器端執行。
  - CGI 能夠讓 web service 提供使用者執行類似 console application (透過命令提示字元運作的應用) 之程式。CGI 可以在 web server 上實現動態 HTML 頁面的展現 (例如隨著 user 透過網頁輸入的值或者一天的 24 小時有所變化)。
  - 使用 CGI 的 web 應用程式設計中須包含 CGI 指令碼。通常情況下,一次的 client request 會 對應一個 CGI 指令碼的執行,並產生相對應的 HTML 以 response。
  - 現在已經較少看到 CGI 的描述,很多網頁應用情境會說明其使用了 Rest API 或 HTML5/Web Socket (提供更豐富的雙向互動) 的方式來實作。

# Q & A

Thanks for your time and attention.