

# 第十五章 全球資訊網系統

## 15-1 WWW 系統簡介

『全球資訊網』( World Wide Web, WWW ) 是帶動整個 Internet 網路發展中最主要的系統，已漸漸成為 Internet 的代名詞。WWW 將網路由專業領域帶到商業用途與日常生活，尤其近年來電子商務的發展，都是由 WWW 系統演變而來。首先，介紹 WWW 的發展過程：

- 1989 年瑞士日內瓦的歐洲粒子物理實驗室 CERN (European Laboratory for Particle Physics) 建立整合性資訊系統，提出 WWW 的構想。
- 1990 年 11 月在 NeXT 系統上開發出 WWW 雛型，12 月開發出文字模式的瀏覽程式。
- 1993 年 4 月 NCSA ( The National Center for Supercomputing Applications ) 發表 Mosaic ( 文字模式 )。
- 1994 年起，WWW 開始了第二次多媒體革命，導入多媒體技術。
- 1994 年 7 月 Netscape 進入 WWW 市場 ( 多媒體模式 )。
- 1995 年 6 月 Sun-soft Java Applet 增添文字以外的動態美感。
- 1995 年 12 月 Oracle 開發出 Web 伺服器連接資料庫技術。
- 1996 年 3 月 Microsoft 發表 Active X 架構。

我們可以發現在短短的十年功夫，WWW 就改變了整個網路世界，成功地連接數千萬台電腦，和電話連結演進的時效相比，真是不可思議。而且 WWW 的應用更不斷地推陳出新，除了電子商務外，我們相信它將會再進入更深層的應用。本章首先介紹 WWW 系統架構及其通訊協定，再進一步以目前最廣泛的 Apache 伺服軟體為例，介紹網頁伺服器的安裝與管理。

## 15-2 WWW 系統概念

早期發展網頁系統時，並沒有想到會有這麼大的發展空間，那時候只考慮到如何將文件可以在各種系統之間流通。當時各系統之間的文件格式並不相容(現在也一樣)，譬如，在 Apple、Windows、Unix、或其它系統所製作的文件，並無法直接在另一系統上顯示或修改處理。當時 CERN 只希望建立一套系統，可以共通顯示不同系統之間的文字，它的做法是建構一套平台來顯示文件，而這個平台可以安裝在不同系統之上。另外，由於作業系統之間的檔案結構也不盡相同，無法將一個系統所製作的文件儲存於磁碟片，再由另一系統將它讀出來，因此，共通平台的文件必需利用網路以 ASCII 格式來互相傳輸。圖 15-1 為網頁系統的基本原理，其目的是希望建構一個可以共通的顯示平台，例如，由電腦 A 所製作出來的文件，能夠透過網路傳輸給電腦 B，並可在電腦 B 能如身歷其境般的顯示出來。CERN 的想法是希望各研究單位的成果，利用網路傳輸到其它單位上，而用此共通平台來顯示，以解決不同系統之間，文件格式不相容的問題。

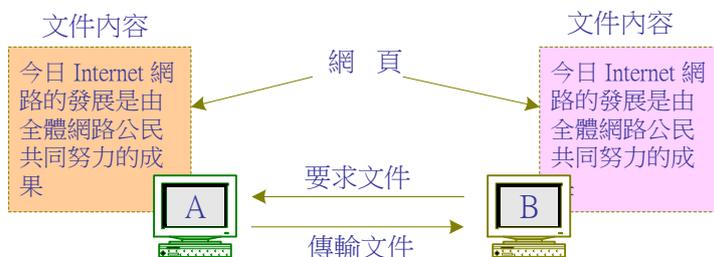


圖 15-1 網頁系統原理

這個共通平台就是目前所稱的『**瀏覽器**』，而電腦之間就以 HTTP 協定來互相通訊。文件在瀏覽器上以一頁一頁為單位來顯示，因此，將瀏覽器上顯示的文件稱之為『**網頁**』( **Web Page** )，又每一文件都有封面，就將文件的封面稱之為『**首頁**』( **Home Page** )，而將提供網頁讓瀏覽器下載的伺服器稱之為『**網頁伺服器**』( **Web Server** )。至於文件要如何製作呢？才能在瀏覽器上顯示出來，於是製定了 HTML 標準，希望所有文件都能依此標準來製作，才能在不同系統上的瀏覽器顯示，網頁系統就是這麼簡單的概念之下產生了。因此，早期網頁只能顯示文字模式(如 Mosaic)，真的沒有想到數位訊號處理技術，也正在同時如火如荼的發展中，兩者一觸即發地結合在一起，很快的將影像及聲音的數位處理技術嵌入瀏覽器之中，多采多姿的全球資訊網世界就因此而誕生了。

### 15-3 WWW 系統架構

有了上述的系統概念後，緊接著我們來看圖 15-2 的 WWW 系統架構。WWW 也是主從式架構，伺服器端( **Web Server** )提供資源( **HTML 文件** )讓客戶端( **瀏覽器** )下載，它們之間是以 HTTP

通訊協定來傳輸。伺服器端使用 URL 的定址方式，客戶端可以依照 URL 位址找到所要的網站，所以 URL 又稱為『網址』。我們分別簡述其功能如下：

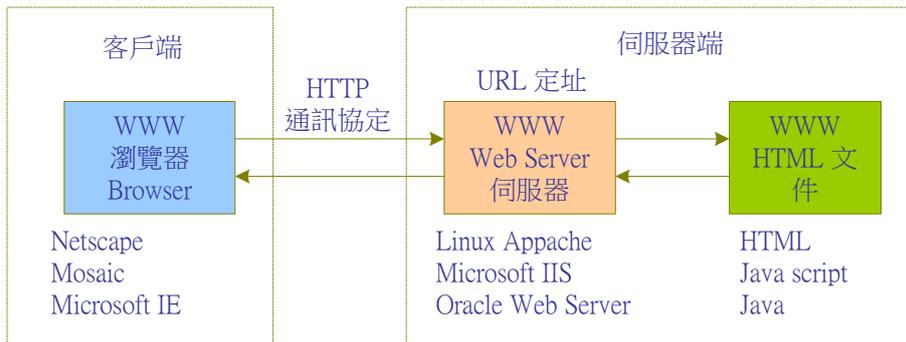


圖 15-2 全球資訊網簡介

### 15-3-1 瀏覽器

客戶端就是瀏覽器（如 IE 或 Netscape），它的功能是由伺服器端上接收 HTML 程式後，再將其執行並顯示成文件，此文件型態就稱為網頁（Web Page）。所以，客戶端以顯示大量文件（或網頁）為主要工作。每一網頁上的文字或圖樣可以指向其它相關頁來連結，頁和頁之間的連結可以無止境的延伸，此連結方法就稱為『超鏈結文件』（HyperText）。不僅可以連結網頁，還可以在網頁上任何文字或圖樣設定連結到其它網站，稱之為『超連結』（Hyperlink）。因此，在客戶端上可以行走全世界任何一個網站，觀看網站上的網頁，所以稱之為『瀏覽器』（Browser）。

基本上，瀏覽器是屬於直譯器（Interpreter）的功能，它由伺服器上下載網頁程式（HTML）將其翻譯並執行後，再將結果顯示到螢幕上。另一方面，它也是屬於 HTTP 伺服器的客戶端，負責和 HTTP 伺服器之間的通訊。但隨著瀏覽器應用的方便性，人們期望將不同的網路功能都附加到瀏覽器上，使它具有其它網路系統的功能，譬如，FTP，BBS，Mail 等功能。因此，再將其它網路系統的客戶端功能加到瀏覽器上，如圖 15-3 所示。

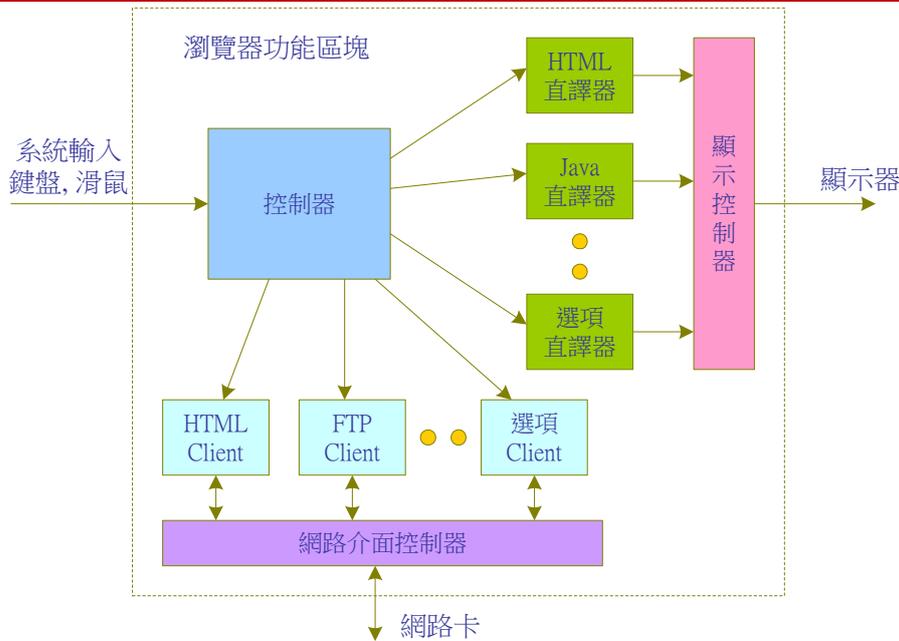


圖 15-3 瀏覽器的功能圖

由圖 15-3 可以發現瀏覽器不再僅是網頁系統的客戶端，也可當作 FTP、Mail、New 等系統的客戶端，甚至可以瀏覽本機上的檔案 (File) 內容。另一方面，如果僅有 HTML 直譯程式，那麼瀏覽器只能觀賞文字或圖形，如此並不能滿足使用者的需求，我們希望能隨使用者的喜好載入其它程式，來增加瀏覽器的功能。例如，插入 Java 直譯器後，瀏覽器就可以執行 Java 所開發的程式，不但可以整合多媒體 (文字、聲音、影像)，還可以表現動畫，提高網站的可看性。也就是說，我們將各種不同的直譯程式，嵌入瀏覽器之中，使它能編譯執行不同的程式語言，來增強瀏覽器的功能。總而言之，瀏覽器不但必須具備有各種通訊協定的功能，也包含各式各樣的直譯程式，因此，可以將瀏覽器視為全球資訊網的工作平台並不為過，也促使瀏覽器成為最複雜的軟體套件，相對的，網路應用系統的最大安全漏洞，也出現在瀏覽器上。

### 15-3-2 一致性資源定址

既然瀏覽器可以接受不同協定的傳輸，也可以處理不同語言所編寫的程式，但它如何判斷該以何種模式來工作，這必須由使用者命令它處理。然而使用者應該以何種模式和瀏覽器溝通，這就是『一致性資源定址』(Uniform Resource Locators, URL) 的制定目的，當使用者以 URL 通知瀏覽器工作時，必須標明以下三項資訊：(a) 連接該網站使用何種通訊協定 (http 或 ftp)；(b) 網站位址在哪裡 (主機的 DNS 名稱)；(c) 該網頁的檔案名稱 (或檔案格式)。例如：

<http://www.tsnien.idv.tw/index.html>

URL 的三個部分是：通訊協定( http )、主機位址( www.tsnien.idv.tw )、網頁的檔案名稱( index.html )。通訊協定有：http ( 超連結文件，HTML )、ftp ( FTP 檔案傳輸協定 )、file ( 本地檔案 )、news ( 新聞文章 )、gopher ( Gopher 文件協定 )、mailto ( 傳送郵件協定 )。檔案名稱的副檔名 ( 如，.html ) 是用來標示，該檔案是由何種程式語言編寫而成，以啟動相對應的直譯程式 ( 如，HTML 直譯程式 )。

並非只有使用者下達命令時，才會以 URL 模式通知瀏覽器，在網頁上的超連結，也是使用 URL 位址模式來連結到另一個網頁 ( 或網站 )，在連結中可能跨越不同的通訊協定，譬如，在瀏覽網頁 ( HTTP 協定 ) 時可超連到檔案傳輸 ( FTP 協定 ) 功能，也可超連結來傳遞郵件 ( SMTP 協定 )。

### 15-3-3 網頁伺服器

『**網頁伺服器**』( **Web Server** ) 是用來儲存 HTML 文件，讓瀏覽器下載執行的伺服器。它和客戶端之間是以 HTTP 通訊協定溝通，又稱為『**HTTP 伺服器**』( **HTTP Server** )，傳輸埠口大多架設在 80/tcp 位置。網頁伺服器是目前最炙手可熱的設備，它也是一套非常複雜的系統。隨著網站需求的大量增加，一部網頁伺服器只能架設一個網站已漸不符所需了，我們希望在同一部網頁伺服器上建構更多的網站，才能符合經濟價值。因此，它必須透過虛擬主機技術，來建構許多虛擬網站，乃至個人網站。這方面我們將以 Apache 伺服器為範例，來介紹網頁伺服器的管理，以及虛擬網站和個人網站的架設技巧。

## 15-4 HTML 語言

早期設計網頁瀏覽器的觀念非常簡單，僅希望整合不同文書處理工具所產生文件，並於共同的瀏覽器上顯示。不同的文書處理有：MS-Word、vi、PE2 等等，各種文書處理的文件都有其資料表示方式。例如，在 MS-Word 上的文書儲存方式，對於某一段文字的大小或粗細體 ( 如 14 號字及粗體 )，會在這一段文字的頭尾加入標記符號，文章被開啟時，再依照標記符號將文字型態顯示在螢幕上。網頁的設計就是利用這種觀念，希望在文件上，以標準的標記符號來標示，使能在不同工作平台上的瀏覽器開啟，這個標準文件標記語言就是『**超文字標記語言**』( **HyperText Markup Language, HTML** )。

HTML 為 ISO 8879 · SGML ( Standard Generalized Markup Language ) 的一種應用，設計使用於超連結文件並為 Web 所採用。如前所述，HTML 是一種標記 ( Markup ) 語言，即是用於描述文件格式的語言。標記( Markup )是放置文件之中，告知排版軟體顯示文件的格式。例如，在 HTML 中，<B> 表示開始粗體模式，</B> 表示離開粗體模式。以記號語言所寫的文件和 WYSIWYG( What You See Is What You Get ) 文書處理器產生的文件非常類似，如 MS-Word 一樣。一些較常用的標記如表 15-1 所示。

**表 15-1 HTML 語言之標記**

標 記	描 述
<HTML> ..... </HTML>	宣告網頁將以 HTML 編寫。
<HEAD> .... </HEAD>	定義網頁的檔頭。
<TITLE> ... </TITLE>	定義標題 ( 並不在網頁上顯示 ) 。
<BODY> ... </BODY>	框註內為網頁主體。
<Hn> ... </Hn>	n=1~6，框註內六個階層的標題字大小。
<B> ... </B>	設定框註內文字為粗體。
<I> .... </I>	設定框註內文字為斜體字。
<UL> ... </UL>	框註內為無序串列 ( 註標式 ) 。
<OL> ... </OL>	框註內為編號串列。
<MENU> ... </MENU>	框註內<LI>項目行成選單。
<LI>	串列項目的開始 ( 並無 </LI> ) 。
 	強迫分離。
<P>	區段開始。
<HR>	水平線。
<PRE> ... </PRE>	已格式化文字。
<IMG SRC= "...">	在此載入影像圖形。
<A HREF= "..."> ... </A>	定義超連結。

我們用圖 15-4 為例子來說明 HTML 程式的編寫方式。HTML 編寫程式相當類似於 MS-Word 的文件儲存格式，瀏覽器再依照文件上的標記來顯示。反過來，如果我們使用類似 MS-Word 的文件編輯方式，將於 MS-Word 上編輯（如 WYSIWYG 方式）的文件，儲存時再依照標記文件方式儲存。藉由此概念，使用一般文件編輯方式來設計網頁，儲存時再依照 HTML 語言型態儲存，就可節省許多程式設計的時間，這個網頁編輯工具就是類似 FrontPage 的網頁開發工具。



**圖 15-4 HTML 編寫方式**

另外為了提高網頁的可看性，HTML 結合了 MIME ( Multipurpose Internet Mail Extensions ) 協定( RFC 1314 )。MIME 是將網頁( 或程式 )定義成標準化格式，網頁的包裝格式分為表頭( Header )和主體 ( Body )。一般表頭包括網頁格式 ( HTML )、網頁頁頭、網頁標題、網頁主題及粗體文字等等 ( 依照 HTML 定義 )。主體是指某一個表頭的內容。

製作網頁並非只有 HTML 語言，我們還可將許多直譯程式嵌入瀏覽器之中，來執行 Java，Active X、JavaScript、VBScript 等語言所發展的程式，甚至可加入 CGI( Common Gateway Interface ) 介面，使網頁系統不再只是單方向的傳輸，而成為雙方向的『動態網頁』( **Dynamic Page** ) 或『活動伺服器網頁』( **Active Server Page, ASP** ) 等等。本書因限於篇幅，這方面不再詳細介紹，請讀者參考其它書籍。

## 15-5 HTTP 傳輸協定

『超文件傳輸協定』( **HyperText Transfer Protocol, HTTP** ) 是針對 Web 設計的傳輸協定。目前使用的協定大多是 HTTP/1.0( RFC 1945 ) 以上的版本，而在 HTTP/1.1( RFC 2068、RFC 2616 )

版中增加了虛擬主機的功能，相信會有更新版本陸續被發展出來。在 HTTP 協定下，伺服器端和客戶端之間也是以 NVT ASCII (第十二章介紹) 終端機模式來互相通訊，屬於『要求/回應』的存取方式。也就是說，客戶端以 ASCII 文字的通訊命令，來要求伺服器端執行，伺服器端將執行結果以 ASCII 文字方式回應給客戶端，基本上，所有請求命令都由客戶端要求，而伺服器端只是被動地接受命令來工作。

### 15-5-1 HTTP 命令與回覆

基本上，HTTP 協定只負責伺服器和客戶端之間文件的傳輸工作，並不理會文件格式 (如，HTML、VBScript、JavaScript 等等)，文件格式是由瀏覽器上相對應的直譯程式來處理。也就是說，無論何種格式都以 ASCII 編碼來傳輸，既然，客戶和伺服器之間是以 NVT 終端機模式通訊，我們也可以利用 Telnet 程式登入該傳輸埠口，如下：

```
[tsnien@linux-2 tsnien]$ telnet www.cu.edu.tw 80
Trying 163.15.2.30...
Connected to linux-2.cu.edu.tw (163.15.2.30).
Escape character is '^'.
get /index.html HTTP/1.1 (輸入 HTTP 命令)
```

雖然我們可以直接輸入 HTTP 命令，但傳回文件時，因顯示平台 (不是瀏覽器) 不符，而會發生錯誤的情形。HTTP 協定包含下列幾種基本命令：

- (1) **GET**：請求讀取網頁。瀏覽器以 GET 命令，請求伺服器送出網頁，以 MIME 方式編碼。
- (2) **HEAD**：請求讀取網頁的檔頭 (header)。客戶端僅請求訊息檔頭，而非實際網頁。這方法可取得網頁最後修改時間，可用在建立或測試 URL 的有效性。
- (3) **PUT**：請求儲存網頁。客戶端寫入網頁，提供客戶端建立網頁的功能。
- (4) **POST**：附加一個名稱資源。客戶端將資料附加到某一資源的資料之後。
- (5) **DELETE**：刪除網頁。客戶端要求刪除某一網頁。
- (6) **LINK**：建立超連結。客戶端要求加入超連結。
- (7) **UNLINK**：刪除超連結。客戶端要求刪除超連結。

如同其它應用系統一樣，HTTP 伺服器也以 3 個數字來回應執行的結果，第一個數字表示回應種類，如 1×× 表示訊息，目前未使用；2×× 表示命令執行成功 ( Success )；3 ×× 為命令執行轉向 ( Redirection )；4×× 為客戶端錯誤 ( Client Error )；5×× 為伺服器錯誤 ( Server Error )，常用的回應訊息如下：

- **200**：執行正確 ( OK )
- **201**：開啟服務 ( Created )
- **202**：接受服務 ( Accepted )
- **204**：無此文件 ( No content )
- **301**：永久移除 ( Moved Permanently )
- **302**：暫時移除 ( Moved Temporarily )
- **400**：命令錯誤 ( Bad Request )
- **401**：未經授權 ( Unauthorized )
- **403**：請求禁止 ( Forbidden )
- **500**：伺服器錯誤 ( Internal Server Error )
- **501**：伺服器未實現 ( Not Implemented )
- **502**：閘門錯誤 ( Bad Gateway )

除了回應 3 個數字來代表處理結果外，在數字後面也會緊接著一連串文字 ( ASCII )，來說明處理結果的訊息。

## 15-5-2 網頁快取功能

如果客戶端想要瀏覽某一網頁時，都必須到該網頁的伺服器上索取，這將使全球資訊網的效率大打折扣。試想，我們想要瀏覽某一網站，而這網站可能位於世界的任何一個角落，客戶端和伺服器之間可能需要經過多個網路端點才能到達，然而，網路上使用量高時，所產生的高負載，將可能導致整個網路癱瘓掉，還好有網頁快取功能來彌補這個缺憾。

『代理伺服器』( Proxy Server ) 是網頁快取功能中最主要的設備，它有兩個主要功能，其一為代理功能，用來接受客戶端的要求，連接到網頁伺服器，並索取網頁來回應給客戶端；其二為快取功能，它將所索取到的網頁儲存到快取資料庫內，以備其它客戶端有同樣需求時，再直接給予回應。整個快取功能如圖 15-5 所示，當客戶端需要索取某一網頁時，首先到快取紀錄上查詢(本機上)是否有該網頁，如果有，便直接由快取紀錄上索取網頁來顯示(編號(1))；如果網頁不在快取紀錄上，瀏覽器便連結到 Proxy Server，請求 Proxy Server 來代理存取網頁，此時，Proxy Server 首先到快取資料庫上查詢是否有該筆網頁，如果有便直接回應給客戶端(編號(2))；如果網頁資料也不在快取資料庫內，Proxy Server 便啟動網路連線到該網頁的伺服器上索取，並將索取回來的網頁存入快取資料庫上，再回應給客戶端(編號(3))。

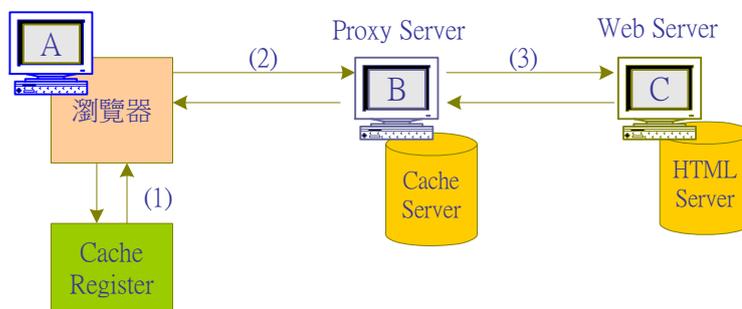
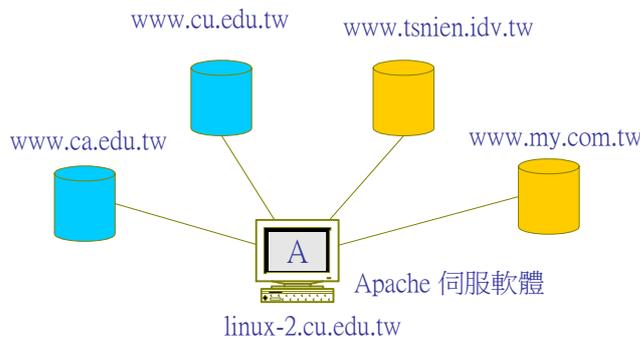


圖 15-5 網頁快取功能

我們可以來想像一種情況，如果有一部 Proxy Server，被許多客戶端(瀏覽器)所指定，它對 Internet 網路上的網頁伺服器的存取機會將非常大，相對的，其快取資料庫上的網頁必然非常豐富，當客戶端要求網頁時，便能提高快取資料的機率，因此連結到此 Proxy Server 的瀏覽器，所能抓取網頁的速度最快，相對的，造成其它客戶端更喜歡連結到此 Proxy Server，因此，網路上也常會出現許多著名的 Proxy Server。但從另一觀點來看，要使 Proxy Server 能接受許多客戶端的連線要求，除了 Proxy Server 主機的處理能力要強外，網路傳輸效率(或網路速度)更不容忽視。

### 15-5-3 虛擬主機功能

何謂『虛擬主機』( Virtual Host )？就是在一部主機(或一套伺服軟體)上，可以架設多個網頁伺服器，除了原來預設(主要)伺服器外，增加出來的伺服器都稱為『虛擬主機』。如圖 15-6 所示，我們安裝了一套 Apache 伺服軟體在 linux-2 主機上，除了建立一部主要伺服器(www.cu.edu.tw)外，還可增加多部虛擬主機(www.tsnien.idv.tw 等)，對客戶端而言，存取主要伺服器上的網頁或由虛擬主機上存取，並沒有兩樣。建構虛擬主機方式有下列三種類型：



**圖 15-6 虛擬主機架構**

- ▼ **IP-based 虛擬主機**：有些作業系統允許在一部主機（或稱網路卡）上，可設定多個虛擬 IP 位址。IP-based 的做法是將所增加的伺服器安裝在這些虛擬 IP 位址上，而達到虛擬主機的功能。目前 IP 位址漸漸不足，此種做法雖然容易，但會浪費 IP 空間。
- ▼ **Port-based 虛擬主機**：一般主要伺服器是連接到 80/tcp 埠口，我們只要將所增加的伺服器連接到其它埠口（1023 以後），而仍使用原來的 IP 位址，便可達到虛擬主機的功能，譬如，www.cu.edu.tw:8000。
- ▼ **Name-based 虛擬主機**：以網站名稱來分辨的虛擬主機，它的做法看似很簡單，只要在 DNS 伺服器上增加虛擬主機名稱對應到原 IP 位址，而在網頁伺服器上，規劃每一網域名稱所存取的網頁文件位置，便可達成虛擬主機的功能，但必須有新的協定來支援（HTTP/1.1）。

其中 IP-based 和 Port-based 的建構較沒有問題，只要用不同的 IP 位址或傳輸埠口來區分不同的網站即可。但在 Name-based 方面，要在同一部伺服器上，以不同的網域名稱來分辨不同的網站，就較為困難。記得當我們指定某一個 URL 位址（www.cu.edu.tw）時，其實該主機名稱（www）也是一個別名（linux-2），而它會固定連接到埠口 80（163.15.2.30:80）。當我們用虛擬主機名稱（www.tsnien.idv.tw）連結到同一伺服器時，經過 DNS 解譯後位址也是 163.15.2.30，連接到伺服器同樣是第 80 埠口，那麼伺服器如何來分辨 www.cu.edu.tw 或 www.tsnien.idv.tw 兩個不同網站呢？

HTTP/1.1 協定為了解決這個問題，在『要求網頁』（Get）的命令格式當中，加入一個主機名稱（Host:）選項欄位，伺服器接收此命令後，便可利用所加入的主機名稱，來分辨到底是存取哪一網站的網頁，利用此方法來達成 Name-based 虛擬主機的功能。Get 命令格式（HTTP/1.1）如下：

**命令 1**：Get <SP> / <SP> HTTP/1.1 <CR><LF> Host:www.cu.edu.tw <CR><LF>

**命令 2：Get <SP> / <SP> HTTP/1.1 <CR><LF> Host:www.tsnien.idv.tw <CR><LF>**

第一個命令表示是索取 [www.cu.edu.tw](http://www.cu.edu.tw) 網站上的網頁；第二個命令是針對 [www.tsnien.idv.tw](http://www.tsnien.idv.tw) 網站存取的。然而這兩個網站都屬於同一網頁伺服器所管轄，對使用者而言，也分不出哪一個是主要網站或虛擬主機。

## 習題

1. 請敘述建構 WWW 系統的基本理念。
2. 請繪圖說明 WWW 系統架構，並說明包含哪些元件。
3. 何謂『**瀏覽器**』( **Browser** )？並說明應具有的功能。
4. 何謂『**一致性資源定址**』( **Uniform Resource Locators, URL** )？
5. 何謂『**HTML 語言**』？並請說明其特性。
6. 何謂『**HTTP 伺服器**』( **HTTP Server** )？
7. 請說明 HTTP 協定中，Client/Server 之間的通訊模式。
8. 請繪圖說明『**網頁快取**』的功能，它和哪一個網路應用系統非常相似？
9. 何謂『**代理伺服器**』( **Proxy Server** )？功能為何？
10. 何謂『**虛擬主機**』( **Virtual Host** )？有哪幾種實現方式？
11. 請利用封包攔取軟體 ( 附錄 A )，來攔取一個 HTTP 協定封包，並分析 IP 和 TCP 封包格式。
12. 同上題，請攔取一個 Get 命令封包 ( 由瀏覽器抓取網頁 )，來分析 Get 命令訊息的封包格式。
13. 同上題，請分析由伺服器傳回的 Response 訊息之封包格式。