

KHOA CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG
BỘ MÔN MẠNG MÁY TÍNH VÀ TRUYỀN THÔNG

LẬP TRÌNH WEB



Chương 1: CÁC KHÁI NIỆM CƠ BẢN

Biên soạn : ThS. Hà Duy An

Cập nhật : ThS. Nguyễn Cao Hồng Ngọc

Nội dung



- Tổng quan
- Các thành phần của dịch vụ web
- Giao thức HTTP
- Web Cache
- URL
- HyperText & HyperLink

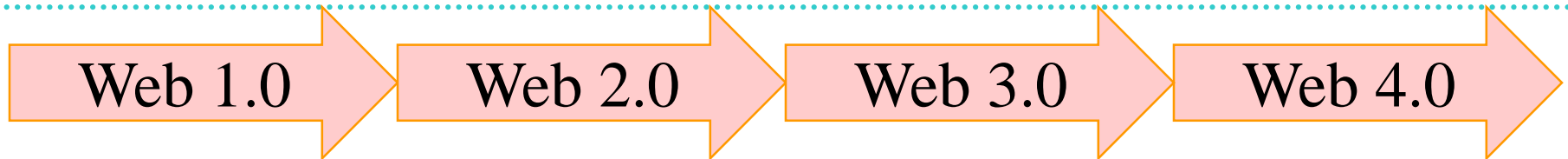
Tổng quan



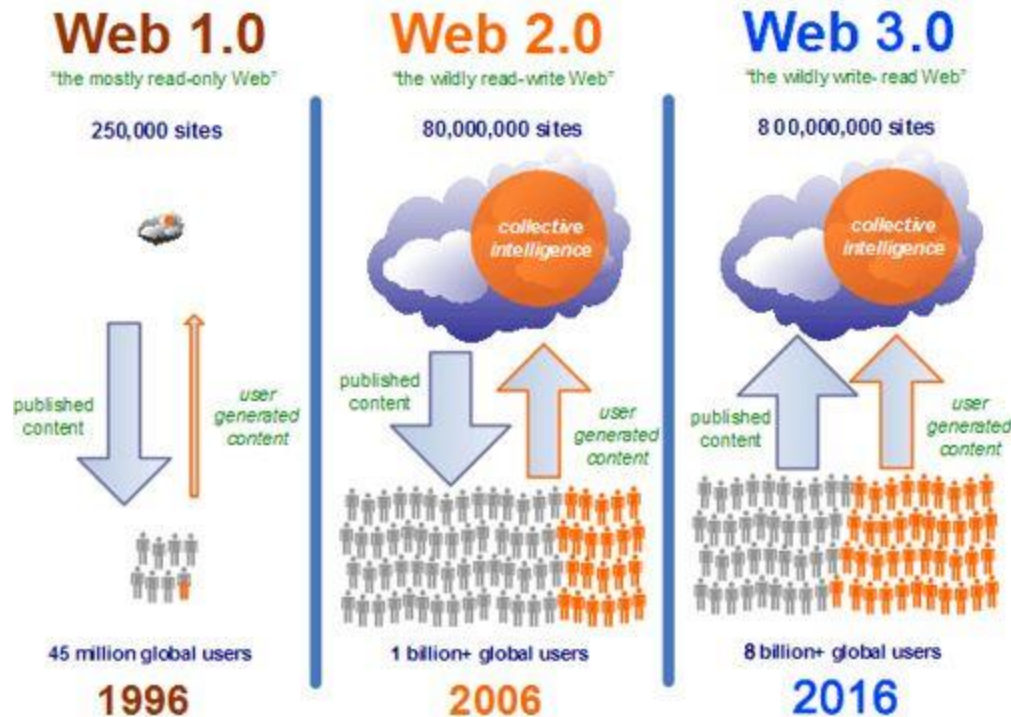
■ Web là gì?

- World Wide Web, www, web, w3 là một dịch vụ trên Internet
- Là hình thức tổ chức thông tin phổ biến và tiện lợi nhất hiện nay
- Cho phép tra cứu tài nguyên thông tin qua các siêu văn bản (Hypertext) sử dụng các siêu liên kết (Hyperlink)
- Là các tài liệu văn bản thường được lưu trữ với phần mở rộng .html, .htm
- Web được phát minh và đưa vào sử dụng vào khoảng năm 1990 bởi viện sĩ Viện Hàn lâm Anh Tim Berners-Lee tại CERN, từ đó đến nay web đã phát triển mạnh mẽ và trở thành một hệ thống siêu phương tiện.

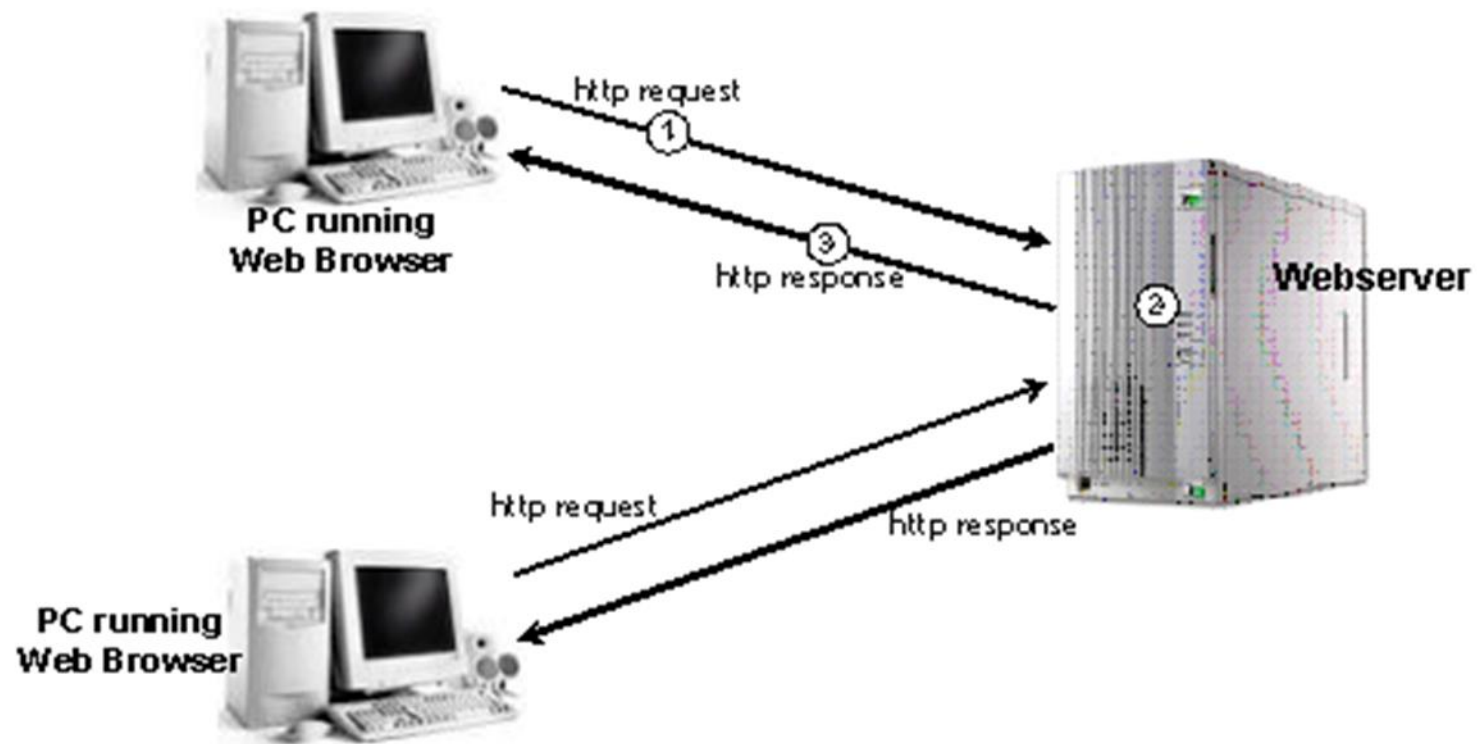
Các thể hệ web



Information web (1994 - 2000)	Social web (2000 - 2010)	Semantic web (2005 - 2010)	Intelligent web (2015 - ????)
----------------------------------	-----------------------------	-------------------------------	----------------------------------



Các thành phần của dịch vụ web



Web client & Web server



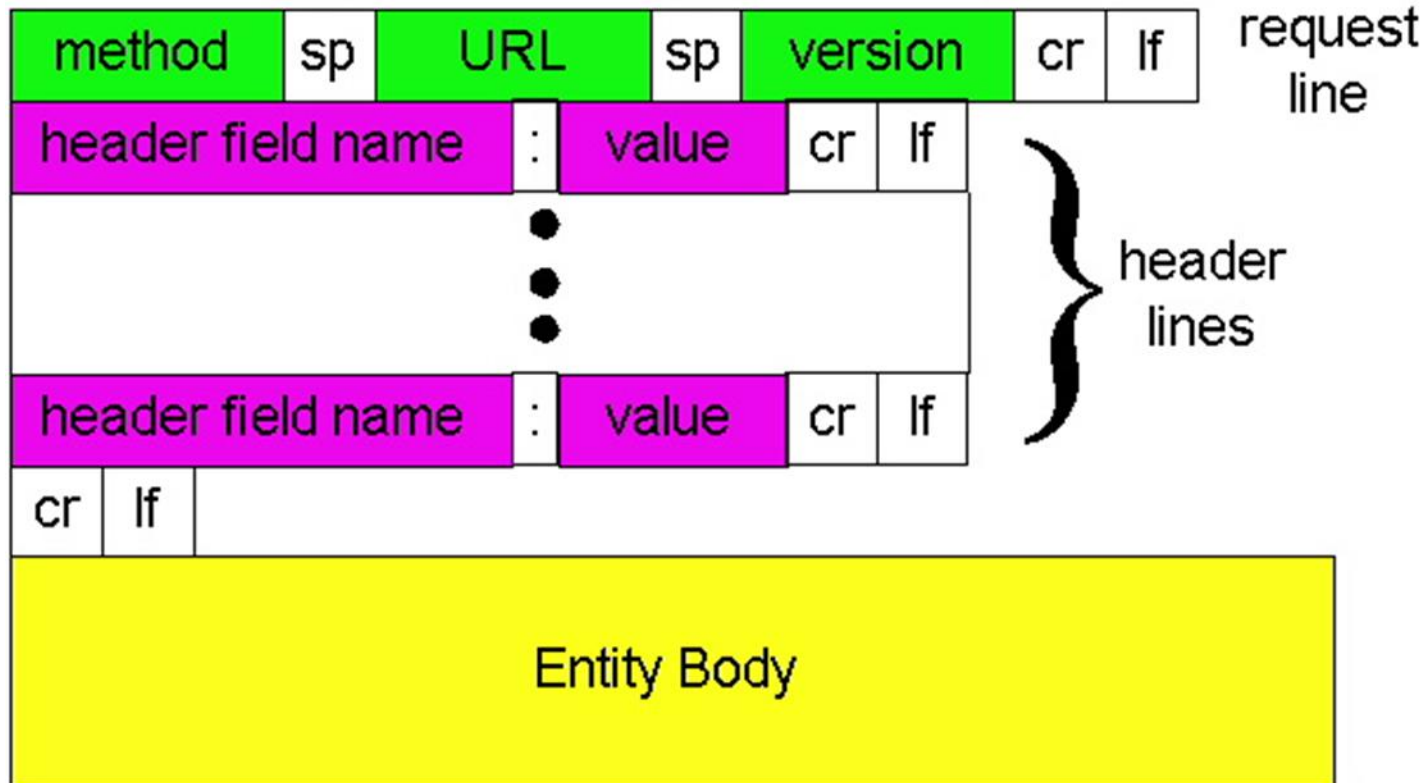
- Client: *Web Browser*: Internet Explorer, Firefox, Chrome, Safari, Opera, Netscape, Mozilla, ...
 - Được cài đặt tại máy khách
 - Gửi các yêu cầu về web đến Web Server
 - Nhận kết quả trả về từ Web Server
 - Hiển thị kết quả
- Server: *Web Server*: Apache, MS Internet Information Server, lighttpd, Sun Java System Web Server, Jiqsaw Server, ...
 - Cài đặt tại máy chủ, cung cấp một dịch vụ Web
 - Lắng nghe các yêu cầu về Web trên một cổng (80)
 - Xử lý các yêu cầu
 - Tạo kết quả và trả về cho trình duyệt Web

Giao thức HTTP



- HyperText Transfer Protocol
- Dùng để giao tiếp giữa Web Browser và Web Server
- Giao thức ở tầng ứng dụng trong mô hình OSI, hoạt động trên nền giao thức TCP/IP
- Có 2 kiểu thông điệp: request (webbrowser), response (webserver).
- HTTP Server hoạt động mặc định trên cổng 80
- Là giao thức “không trạng thái” (*stateless*)
- Có thể dùng để truyền tải bất kỳ kiểu dữ liệu nào
- Thông tin trao đổi chủ yếu là các tài liệu bằng ngôn ngữ HTML
- HTTP lúc công bố đến nay đã trải qua nhiều phiên bản 1.x (1.0, 1.1, 1.2, 1.3), đến ngày nay là phiên bản 2.x

HTTP request message



HTTP request message (tt)



- Trong đó:
 - Cr (carriage return) và fl (feed line) là những ký tự xuống dòng, đánh dấu kết thúc thông điệp
 - Method: là phương thức gửi bao gồm:

Method	Description
OPTIONS	Yêu cầu thông tin về các tùy chọn hiện có.
GET	Lấy về tài liệu được xác định trong URL
HEAD	Lấy về thông tin thô về tài liệu được xác định trong URL
POST	Cung cấp thông tin cho server
PUT	Tải tài liệu lên server và đặt ở vị trí được xác định trong URL
DELETE	Xóa tài liệu nằm ở vị trí URL trên server

HTTP request message (tt)



request line
(GET, POST,
HEAD commands)

header
lines

```
GET /somedir/page.html HTTP/1.0
User-agent: Mozilla/4.0
Accept: text/html, image/gif, image/jpeg
Accept-language: fr
```

(extra carriage return, line feed)

Carriage return,
line feed
indicates end
of message

HTTP response message



status line
(protocol
status code
status phrase)

header
lines

data, e.g.,
requested
html file

HTTP/1.0 200 OK

Date: Thu, 06 Aug 1998 12:00:15 GMT

Server: Apache/1.3.0 (Unix)

Last-Modified: Mon, 22 Jun 1998 ...

Content-Length: 6821

Content-Type: text/html

data data data data data ...

HTTP response message (tt)



Mã	Loại	Lý do
1xx	Thông tin	Đã nhận được yêu cầu, đang tiếp tục xử lý
2xx	Thành công	Thao tác đã được tiếp nhận, hiểu được và chấp nhận được
3xx	Chuyển hướng	Cần thực hiện thêm thao tác để hoàn tất yêu cầu được đặt ra
4xx	Lỗi client	Yêu cầu có cú pháp sai hoặc không thể được đáp ứng
5xx	Lỗi server	Server thất bại trong việc đáp ứng một yêu cầu hợp lệ

Hoạt động của giao thức HTTP



- Ví dụ: Giả sử người dùng Web Browser kích hoạt URL:

- *www.someSchool.edu/someDepartment.html*

1a. http client khởi tạo nối kết TCP tới http server tại *www.someSchool.edu* qua cổng 80

2. http client gán thông điệp yêu cầu (http request) có chứa URL vào TCP socket

5. http client nhận thông điệp đáp ứng có chứa tập tin html – hiển thị html – phân tích tập tin html và thấy có 10 đối tượng jpeg được tham chiếu đến

6. Các bước 1-5 được lặp lại cho mỗi đối tượng jpeg

1b. http server tại host *www.someSchool.edu* lắng nghe nối kết TCP trên cổng 80 – “chấp nhận” nối kết – thông báo client

3. http server nhận thông điệp yêu cầu – tạo thông điệp đáp ứng (http response) chứa đối tượng được yêu cầu (*someDepartment.html*) – đưa thông điệp vào socket

4. http server đóng nối kết TCP

Web cache

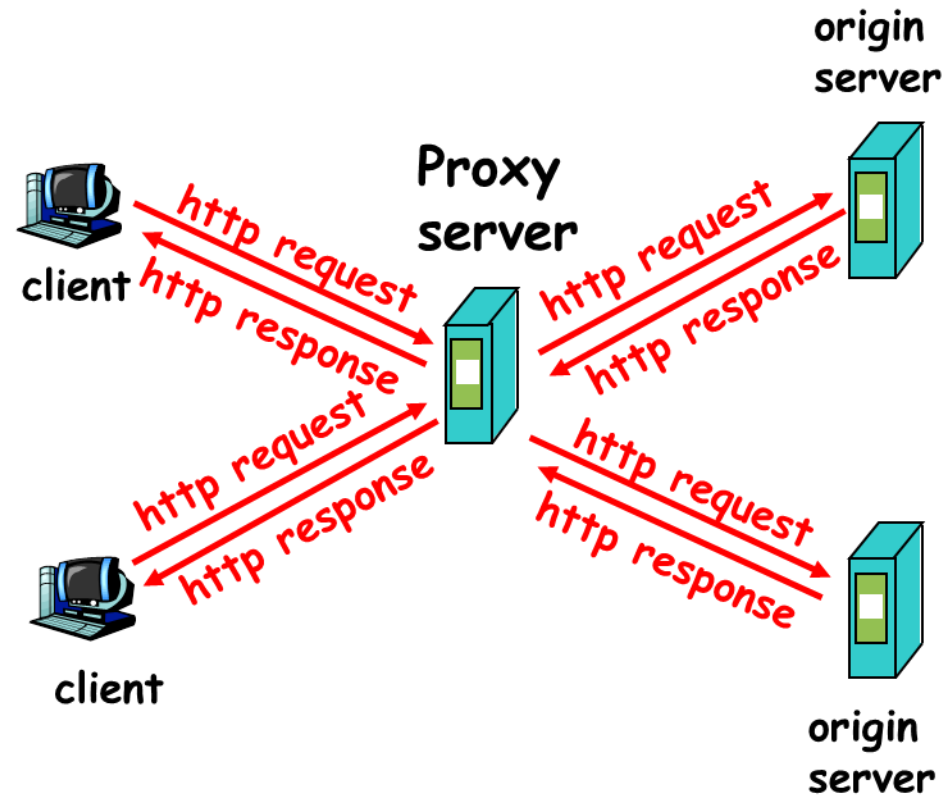


- Web cache: một cơ chế lưu trữ tạm bản sao của các tài nguyên web (các trang html, hình ảnh, ...) nhằm đáp ứng cho client mà không cần truy xuất đến server gốc
- Lợi ích của web cache:
 - Giảm thời gian đáp ứng cho client vì không cần truy xuất đến server gốc
 - Giảm tải cho web server
 - Giảm lưu thông trên mạng, tiết kiệm băng thông => hạn chế tình trạng nghẽn (bottleneck) đường truyền
- 3 loại web cache chính: Browser cache, Proxy cache, Reverse Proxy cache

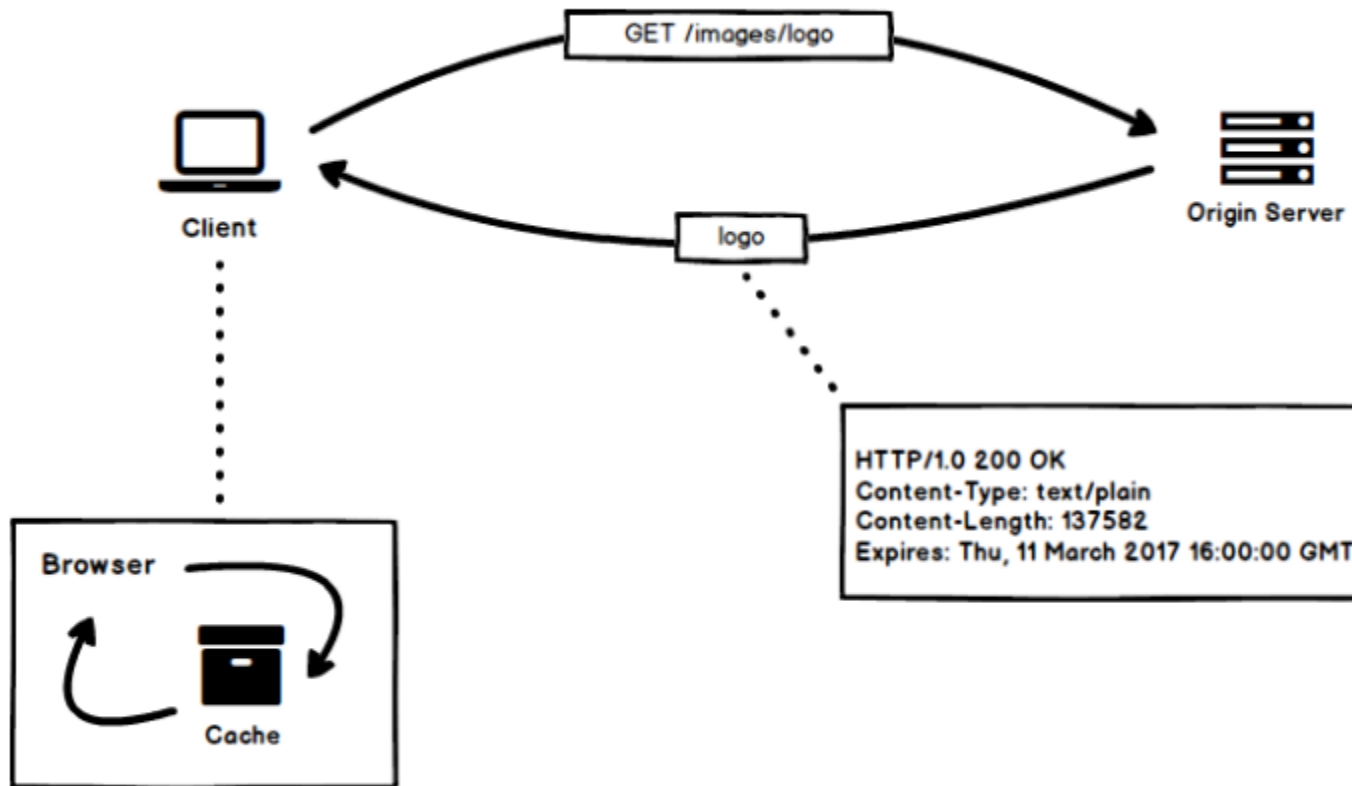
Hoạt động của web cache



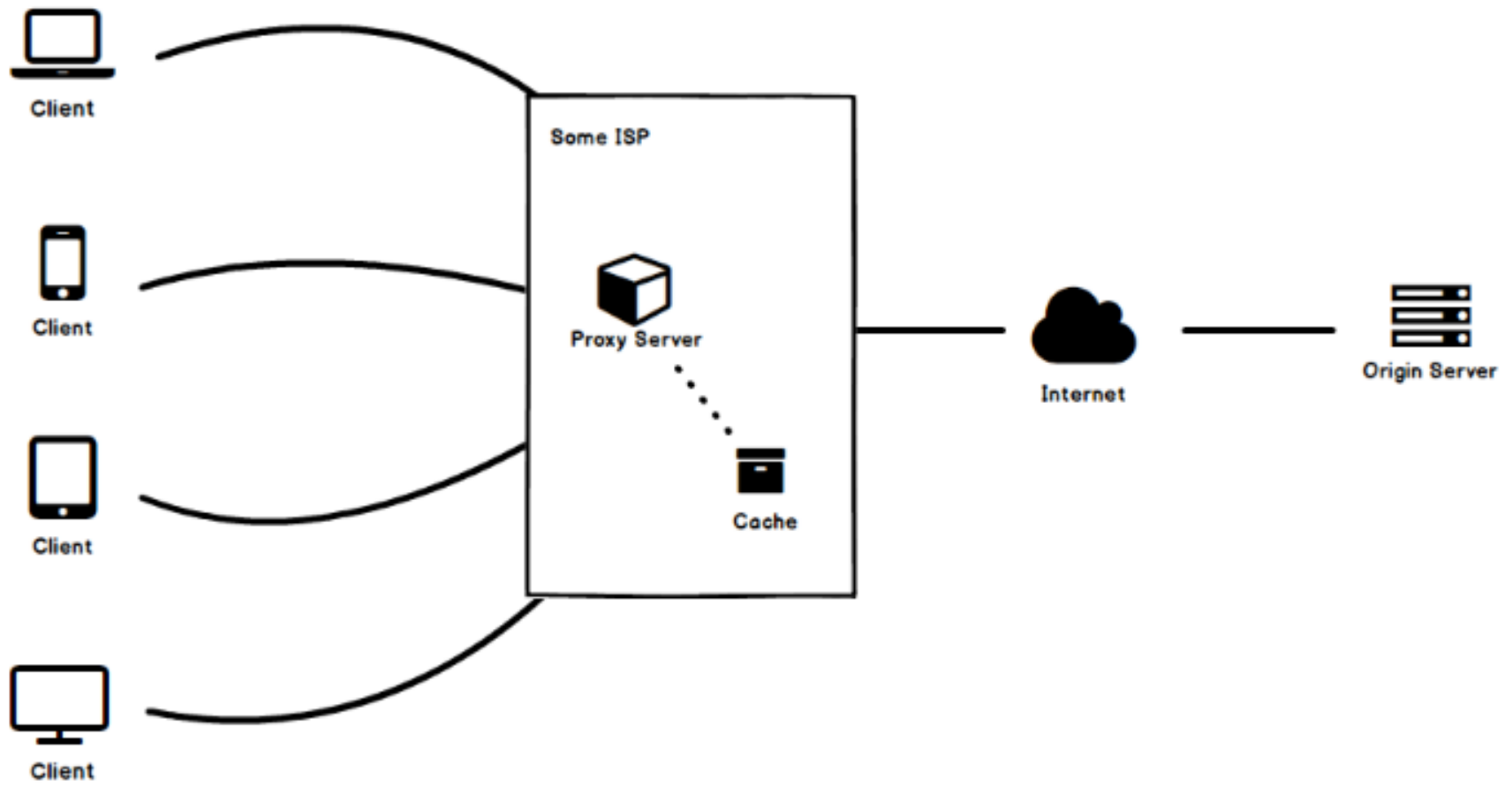
- Tất cả yêu cầu http của client được gửi đến web cache
 - Nếu đối tượng yêu cầu đã được lưu trữ tại web cache, web cache sẽ lập tức đáp ứng cho client
 - Ngược lại, web cache sẽ truy xuất đến server gốc, nhận đáp ứng, chuyển đến client, đồng thời cũng lưu một bản sao trên web cache



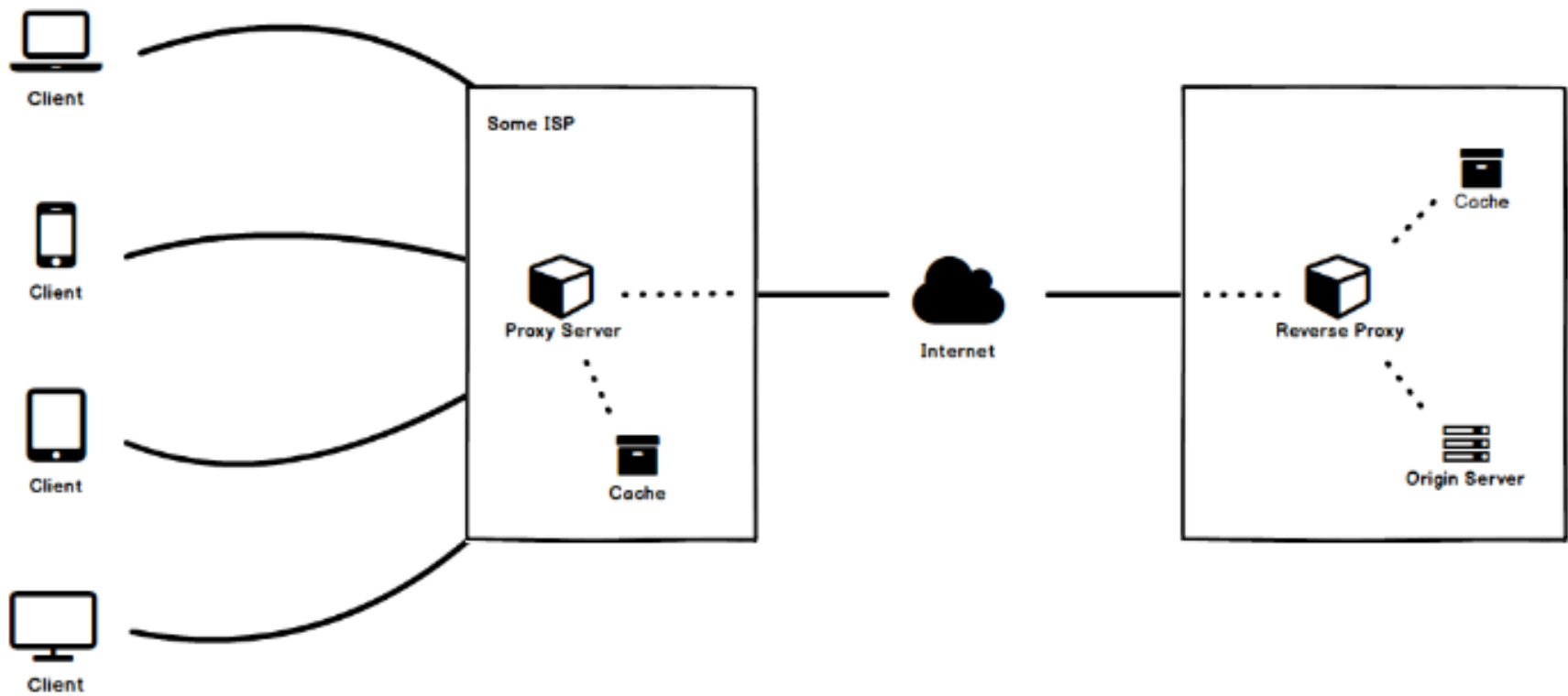
Browser cache



Proxy cache



Reverse Proxy cache



URL



- URL (Uniform Resource Locator): bộ định vị tài nguyên đồng dạng là địa chỉ dùng để định vị các nguồn tài nguyên trên Internet

- Cú pháp tổng quát:

<giao thức>://<địa chỉ server>[:port]<đường dẫn>

- Ví dụ:

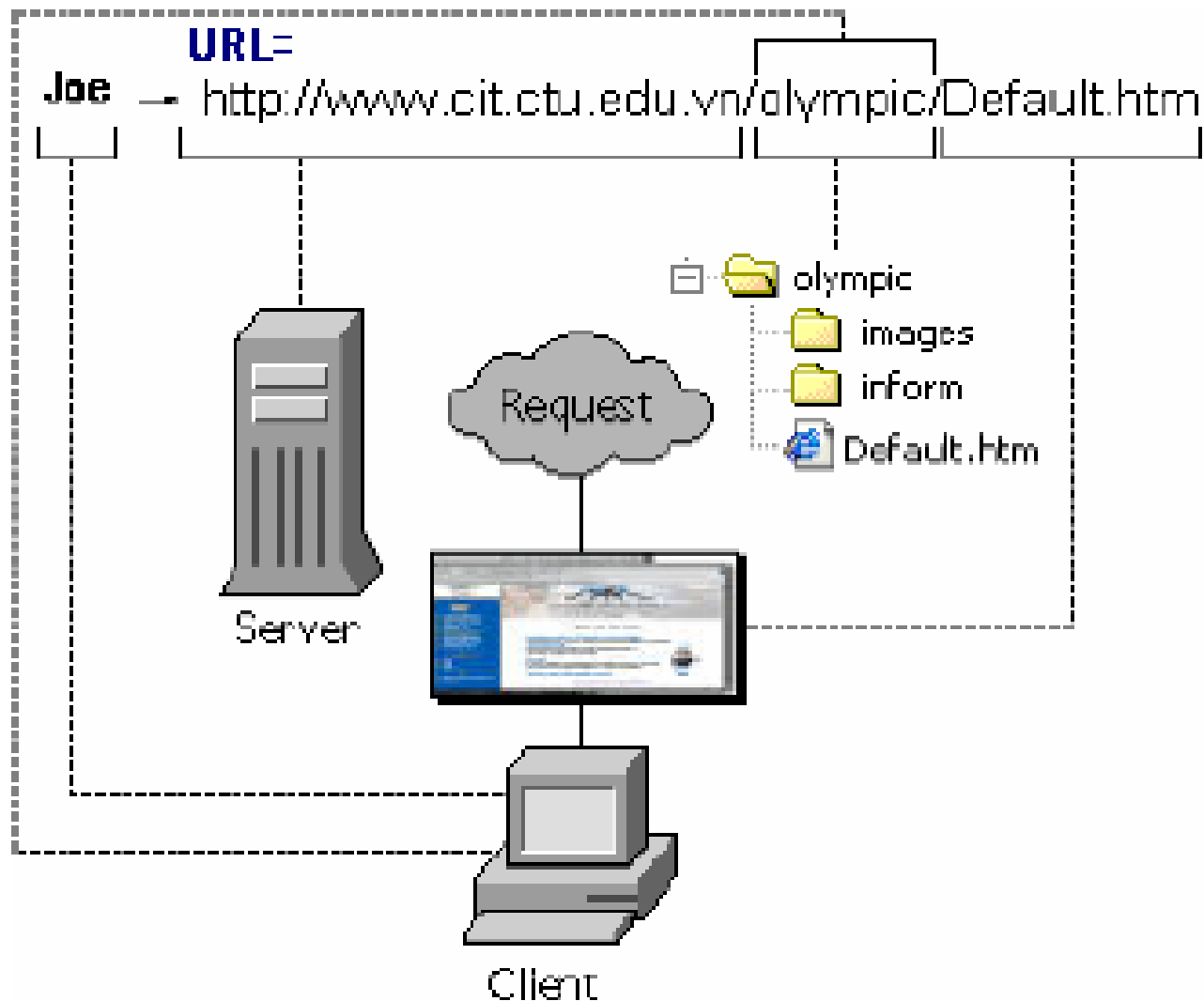
- <http://www.w3.org/Consortium/siteindex.html>

Giao thức

Địa chỉ server

Đường dẫn đến tài nguyên

Ví dụ về URL



Một số URL phổ biến



■ HTTP:

`http://server_name.domain[:port]/directory/subdirecrory [/document]]`

Ví dụ:

`http://www.microsoft.com`

`http://www.ctu.edu.vn:8080/cong/home.htm`

■ FTP:

`ftp://[User_account[:password]@]server_name.domain[:port]/
directory[/document]]`

Ví dụ:

`ftp://ftp.cit.ctu.edu.vn/giaotrinh/`

`ftp://ttinternet:ttinternet@172.18.211.19/thuctap/file1.txt`

Một số URL phổ biến (tt)



- Email:

`mailto:email_address`

Ví dụ:

`mailto:dtngghi@cit.ctu.edu.vn`

- File trên đĩa:

`file://server_name/directory/subdirecrory/document`

`file://<ô đĩa:>/directory/subdirecrory/document`

Ví dụ:

`file://internet_server/course/index.htm`

`file://c:/course/test.html`

URL tuyệt đối & URL tương đối



■ URL tuyệt đối

- Là địa chỉ đầy đủ của một tài nguyên
- Bao gồm giao thức, vị trí mạng, đường dẫn và tên tập tin
- Ví dụ: <http://www.cit.ctu.edu.vn/student/index.htm>

URL tuyệt đối & URL tương đối

■ URL tương đối

- Là một địa chỉ không đầy đủ của một tài nguyên
- Bao gồm đường dẫn (có thể không có) và tên tập tin
- Ví dụ 1:
 - Người dùng đang đọc trang web:
`http://www.cit.ctu.edu.vn/student/homepage.php`
 - Địa chỉ URL tương đối: `chuyenmuc.php`
 - Trình duyệt tự xác định URL tuyệt đối:
`URL= http://www.cit.ctu.edu.vn/student/chuyenmuc.php`

URL tuyệt đối & URL tương đối (tt)

■ URL tương đối (tt):

● Ví dụ 2:

- Người dùng đang đọc trang web:

`http://www.cit.ctu.edu.vn/student/homepage.php`

- Địa chỉ URL tương đối: `../index.php`

- Trình duyệt tự xác định URL tuyệt đối:

URL= `http://www.cit.ctu.edu.vn/index.php`

● Chú ý:

- Ký hiệu `../` được trình duyệt hiểu như trở về thư mục cấp trên.

Hyperlink và Hypertext

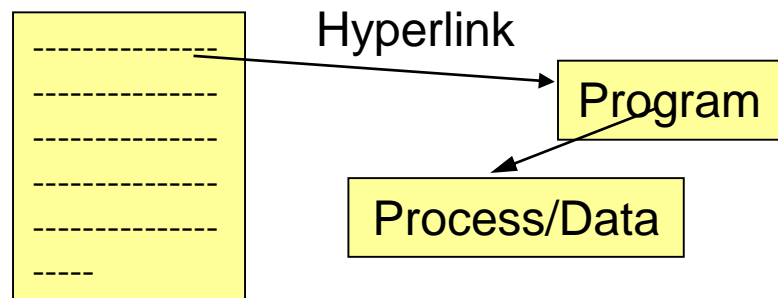
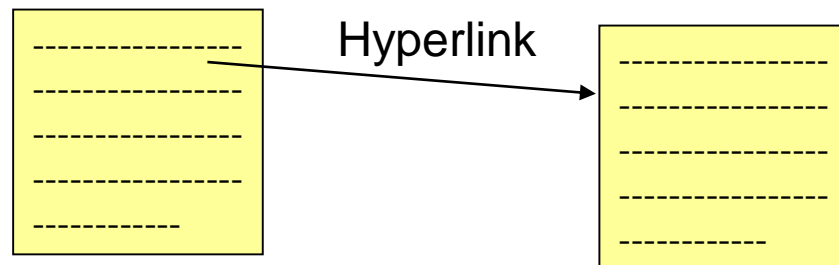
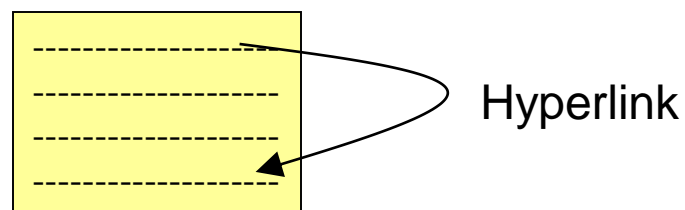


- **HyperText**: (siêu văn bản) là văn bản của 1 tài liệu có thể được truy tìm không theo tuần tự. Người đọc có thể tự do đuổi theo các dấu vết liên quan qua suốt tài liệu đó bằng các mối liên kết xác định sẵn do người sử dụng tự lập nên.
- **HyperLink**: (siêu liên kết) Là mối nối kết giữa phần tử thông tin này với phần tử thông tin khác. Phần tử thông tin có thể là:
 - Văn bản, siêu văn bản, website
 - Âm thanh, hình ảnh
 - Tập tin, các đối tượng ActiveX (Word, Excel,...)
 - Những chương trình có thể thực thi viết bằng các ngôn ngữ như Java, Java Applet, ASP, ASP.NET, PHP,...

Các loại siêu liên kết



- **Liên kết trong:** liên kết trong một tài liệu chỉ đến một phần tử thông tin ngay trong chính tài liệu đó.
- **Liên kết ngoài:** liên kết đến một tài liệu khác bên ngoài tài liệu đang tra cứu.
- **Liên kết có thể thực thi được:** liên kết ngoài, thực thi một chương trình xử lý dữ liệu theo yêu cầu người dùng Web, và cho ra thông tin kết quả.



Các loại siêu văn bản



■ Web tĩnh

- Nội dung tĩnh (như nhau với mọi người dùng)
- Nội dung trang web sẽ được gửi về web browser chính xác như nó được lưu trên web server
- Chỉ bao gồm HTML và CSS
- Web server tổ chức lưu trữ các trang web và đáp ứng trực tiếp các yêu cầu của client, không thông qua môi trường xử lý nào

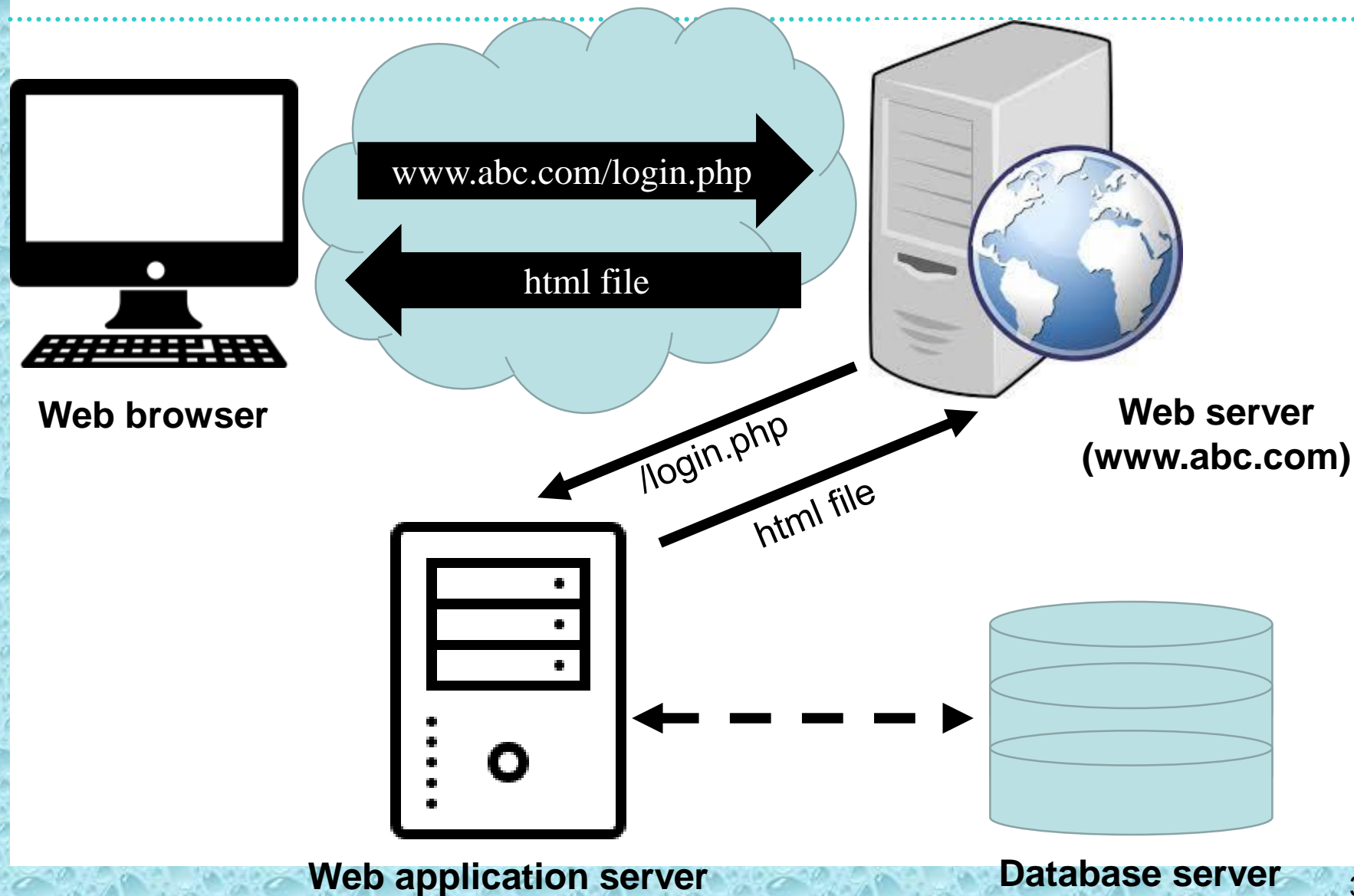
Các loại siêu văn bản (tt)



■ Web động

- Nội dung thay đổi tùy ngữ cảnh
- Nội dung trang web có chứa đoạn mã của một ngôn ngữ script.
- Đoạn mã script này sẽ được ứng dụng web ở máy chủ xử lý trước khi gửi về trình duyệt hoặc được xử lý bởi chính trình duyệt và có thể tạo ra các nội dung khác nhau.
 - Ngôn ngữ script thực thi tại máy chủ (server-side script): PHP, ASP, JSP
 - Ngôn ngữ script thực thi tại trình duyệt (client-side script): Javascript, VBscript

Mô hình web động



HẾT

