

**FIB**Facultat d'Informàtica  
de Barcelona

UNIVERSITAT POLITÈCNICA DE CATALUNYA

CONCEPTES AVANÇATS DE SISTEMES OPERATIUS  
Departament d'Arquitectura de Computadors

# Servidors web sota sistemes Unix

(Seminaris de CASO)

Autors

Cabezas Rodriguez, Javier <[e6974449@est.fib.upc.es](mailto:e6974449@est.fib.upc.es)>

Cerdan Fosch, David <[e7726490@est.fib.upc.es](mailto:e7726490@est.fib.upc.es)>

## Què és un servidor web (I): Conceptes

---

- **SGML** (*Standard Generalized Markup Language*):  
Estàndard internacional per l'intercanvi d'informació amb  
computadors
- **HTML** (*Hypertext Markup Language*): És el llenguatge  
estàndard per la publicació de documents amb *hypertext*.  
És una aplicació de l'estàndard SGML.
- **Hypertext**: És una estructura d'informació que organitza un  
conjunt d'elements en forma de xarxa permanent així  
navegar per ells.

## Què és un servidor web (II): Conceptes

---

- **Internet**: És una infraestructura enorme de xarxes d'ordinadors connectades entre si, una xarxa de xarxes. Per la comunicació, es fan servir unes regles, uns llenguatges, anomenats *protocols*.
- **Protocol**: És una manera concertada de transmetre dades entre dos dispositius.
- **WWW** (*World Wide Web*): És un univers d'informació accessible mitjançant una xarxa d'ordinadors. Una manera d'accedir a la informació d'Internet. Un model d'intercanvi d'informació construït sobre Internet.

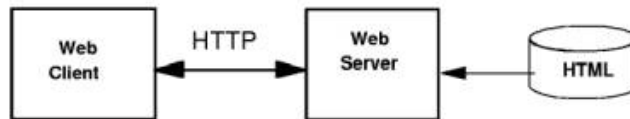
## Què és un servidor web (III): Conceptes

---

- **HTTP** (*Hypertext Transfer Protocol*): És el protocol utilitzat per la WWW. Defineix el format i la manera de transmetre els missatges, i com han d'actuar els servidors web i els navegadors.
- **Pàgina web**: Qualsevol document de la WWW. Està identificada per una URL.
- **URL** (*Uniform Resource Locator*): L'adreça dels documents i altres recursos de la WWW. La primera part indica el protocol, i la segona la adreça IP o el domini sol·licitat.

## Què és un servidor web (IV): Definició

- **Servidor web:** Computador que s'encarrega de servir les pàgines web sol·licitades. Qualsevol ordinador pot realitzar aquesta tasca si hi instal·lem el software servidor adient (*Web server sotware application*) i el connectem a *Internet*.



## Què és un servidor web (V): Breu història

- **Línia del temps de la WWW**
  - Anys 60 i 70: La recerca al DARPA (Defense Advanced Research Projects Agency) dona com a fruit la creació d'Internet, xarxa d'interconnexió d'ordinadors a nivell mundial.
    - Inicis del que seria la base de l'HTML: *hypermedia* i *hypertext*.
  - Anys 70 i 80: desenvolupament de diversos protocols d'Internet (Mail, Gopher, Finger, ...).
  - Finals dels 80: desenvolupament de la World Wide Web (per Tim Berners-Lee).



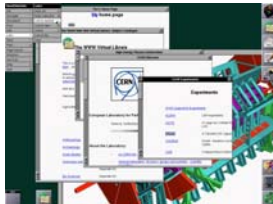
## Què és un servidor web (VI): Breu història

- Anys 90: la WWW sent la necessitat d'un protocol propi. Creació del HTTP (1991).

## Què és un servidor web (VII): Breu història

### □ Servidors i clients web

- Any 1990: El *CERN* (*European Laboratory for Particle Physics*) desenvolupa el primer servidor web, mitjançant la tecnologia orientada a objectes de *NeXT* (empresa fundada per Steve Jobs).
- Any 1991: Prototip de client web (*browser o navegador*): *WorldWideWeb* (anomenat més tard *Nexus*). Va fracassar degut a la poca difusió de la tecnologia *NeXT*.

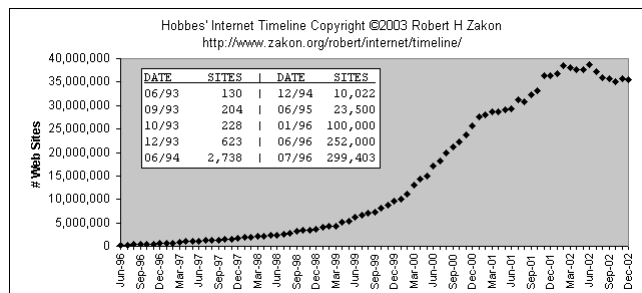


## Què és un servidor web (VIII): Breu història

- Any 1992: Es dona més portabilitat al navegador. Desenvolupament d'altres projectes de navegador: *Erwise*, *ViolaWWW*.
- Any 1993: Marc Andreessen (estudiant del *National Center for Supercomputing Applications (NCSA)*, *Univesitat d'Illinois*) desenvolupa junt amb Eric Bina un navegador més ric gràficament i el porta a altres plataformes no Unix: *Mosaic*.
- Any 1994: Marc Andreessen crea junt amb Jim Clark (fundador de *Sillicon Graphics*) l'empresa *Mosaic Communications Corp (MCOM)*. La universitat d'Illinois reclama els drets del navegador, i MCOM, per protegir-se, es converteix en *Netscape Communications*.
- A partir de 1995: comença la guerra dels navegadors (*Microsoft*).

## Què és un servidor web (IX): Breu història

### □ Creixement del nombre de pàgines web



## Què fa un servidor web?

---

- Un servidor web bàsicament emmagatzema pàgines web i les serveix als clients (normalment *navegadors*) que les sol·liciten.
- És capaç molts cops de proporcionar connexions segures per realitzar transaccions bancàries o moviments de dades personals.
- Algunes aplicacions servidores web ofereixen suport per pàgines dinàmiques (*PHP, ASP, CGI...*) o amb necessitat d'interconnexió amb bases de dades (suport per *SQL,...*).

## Com funciona un servidor web? (I)

---

- Intercanvi d'informació entre l'aplicació client i la servidora:
  - El navegador separa la URL demanada en les seves parts: protocol, adreça i *path* de la pàgina.
  - Mitjançant un DNS (*Domain Name Server*) l'adreça es tradueix a una adreça IP (numèrica).
  - El navegador examina quin protocol s'ha de fer servir.
  - El navegador envia una petició (GET REQUEST) al servidor web i espera una resposta.
  - El servidor verifica el *path*, busca els arxius, executa el codi necessari (si es tracta de codi dinàmic, s'interpreten els *scripts*), tracta les *cookies* si cal i retorna el resultat al navegador.

## Com funciona un servidor web? (II)

---

- En cas de produir-se un error, envia una pàgina d'error al navegador
- El navegador transforma la informació rebuda en HTML i l'interpreta.

### HTTP Error 404

#### 404 Not Found

The Web server cannot find the file or script you asked for. Please check the URL to ensure that the path is correct.

Please contact the server's administrator if this problem persists.

## Alguns exemples

---

- [Apache Web Server](#), usat normalment sota sistemes Unix. També disponible per les plataformes com Windows i Novell Netware
- [MIIS](#) (*Microsoft Internet Information Server*) sota WindowsNT, Windows2000
- [Roxen](#) per Linux, Solaris, MacOS X, Windows.
- [Zeus](#) disponible per la majoria de plataformes Unix
- [Tux](#) només per Linux (inclòs dintre del nucli)

## Apache Web Server: Introducció



- El servidor de pàgines web Apache és un projecte *open source* (codi obert) de la comunitat Apache (*Apache Software Foundation*).
- L'objectiu d'aquest projecte és desenvolupar i mantenir un servidor HTTP (el propi Apache) segur, eficient i extensible.
- Apache és el servidor més popular des de l'Abril de 1996.
- L'agost del 2002, Apache representava el 63% dels servidors de tot el món.

## Apache Web Server: Inicis



- El febrer de 1995 el servidor http més popular era el "*HTTP daemon*", desenvolupat per Rob McCool al NCSA
- Però desde la sortida de McCool del NCSA, a mitjans de 1994, el projecte va quedar gairebé aturat
- Diversos webmasters programaven extensions i *patches*
- El 1995 es van coordinar per unir esforços i van crear l'*Apache Group*
- Després de diverses modificacions i versions, finalment el 1999 es van consolidar i van formar la *Apache Software Foundation*

## Apache Web Server: Versions



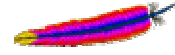
- Actualment es mantenen dues versions:
  - Apache 1.3
  - Apache 2.0
- La versió 1.3 és la versió estable per excel·lència fins fa poc
- La versió 2.0 incorpora noves prestacions però no permet compatibilitat amb els mòduls dissenyats per la versió 1.3

## Apache Web Server: Característ. (I)



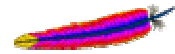
- **Característiques generals de la versió 1.3**
  - potent i flexible, compleix l'estàndard HTTP/1.1
  - implementa les últimes versions del protocol HTTP
  - altament configurable i extensible mitjançant mòduls
  - personalitzable amb mòduls mitjançant l'*Apache Module API*
  - és de codi obert
  - corre sota WindowsNT/9x, Netware 5.x, OS/2 i tot tipus de Unix
  - està continuament en desenvolupament
  - accepta noves idees, *patches* i avisos de *bugs*
  - implementa nombroses funcions sol·licitades freqüentment
  - és multiprocés

## Apache Web Server: Característ. (II)



- Els mòduls es carreguen només quan és necessari (*Dynamic Shared Object, DSO*)
- Possibilitat de tenir més d'un servidor virtual en una màquina, amb un sol Apache (*virtual hosts*)

## Apache Web Server: Característ. (III)



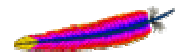
- **Noves característiques de la versió 2.0**
  - a sistemes amb suport de threads POSIX, pot executar-se en un mode multiprocés/multithread
  - renovat el sistema de configuració
  - suport per servir múltiples protocols
  - més segur i ràpid sota plataformes no Unix
  - nova API pels mòduls
  - suport per IPv6, Apache suporta ara escolta amb sockets d'aquesta versió IP
  - els mòduls actuen ara com a filtres (per filtrar per exemple la sortida d'un CGI)

## Apache Web Server: Característ. (IV)



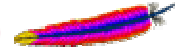
- missatges d'error disponibles en diversos llenguatges
- simplificació de la configuració (crides *Port* i *BindAddress* eliminades; ús més extens de *Listen*)
- suport Unicode per WindowsNT
- actualització de la llibreria d'expressions regulars (ús de Perl 5)

## Apache Web Server: Funcionament (I)



- Com fa servir Apache els processos/fluxes? I els sockets?
- Depèn de la versió
  - versió 1.3 Unix: multiprocés
  - versió 1.3 Windows: un procés, diversos fluxes
  - versió 2.0: multiprocés/multithread/multifibra
- Cicle de vida d'un procés Apache
  - inici loop que serveix les peticions
  - aten peticions
  - mor si: ja ha servit *MaxRequestsPerChild* peticions, es produeix un error fatal o el seu pare l'elimina

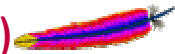
## Apache Web Server: Funcionament (II)



### □ Visió general de la versió 1.3

- un procés pare crea processos fill per atendre les peticions dels clients
- cada fill aten una petició
- quan un fill acaba una petició, es prepara per rebre'n una altra
- el conjunt de fills s'anomena *farm* ó *process pool*
- Apache crea un nombre inicial de fills només iniciar-se, per poder atendre clients el més aviat possible i tenir una capacitat inicial considerable. És el que s'anomena *preforking*
- a Windows, enlloc de tenir un conjunt inicial de fills, té un procés fill amb múltiples fluxes

## Apache Web Server: Funcionament (III)



### □ Visió general de la versió 2.0

- Apache, a la versió 2.0, pot funcionar en un mode híbrid multiprocés/multithread/multifibra, configurable en temps de compilació
- les fibres són threads a nivell d'usuari, implementades fora del nucli de S.O. Són multitasca no preemptiva, només canvien de context en E/S ó a parades explícites
- la opció més eficient (teòricament) és un sol procés, amb N threads i M fibres
- les fibres són les que atenen les peticions

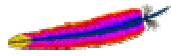
## Apache Web Server: Funcionament (IV)



### □ Sockets

- Apache fa servir la crida `select(2)` per atendre les peticions dels clients
- tots els fills que no estan atenent cap petició comproven (a diferents sockets) si hi ha alguna connexió disponible
- problema: quan diversos fills es disposen a atendre una petició, els que no l'atenen es queden bloquejats esperant una connexió d'aquest socket i no de cap altre
- una sol·lució és fer els sockets no bloquejants
- l'altra sol·lució, que implementa l'Apache, és serialitzar l'entrada al bucle d'`accepts`, per garantir que els fills entrin i facin el `select` un a un (exclusió mútua)

## Apache Web Server: Funcionament (V)



### □ Problema, bloqueig de múltiples fills

```
for (;;) {  
    for (;;) {  
        fd_set accept_fds;  
  
        FD_ZERO (&accept_fds);  
        for (i = first_socket; i <= last_socket; ++i) {  
            FD_SET (i, &accept_fds);  
        }  
        rc = select (last_socket+1, &accept_fds, NULL, NULL, NULL);  
        if (rc < 1) continue;  
        new_connection = -1;  
        for (i = first_socket; i <= last_socket; ++i) {  
            if (FD_ISSET (i, &accept_fds)) {  
                new_connection = accept (i, NULL, NULL);  
                if (new_connection != -1) break;  
            }  
        }  
        if (new_connection != -1) break;  
    }  
    process the new_connection;  
}
```

## Apache Web Server: Funcionament (VI)



### □ Sol·lució al problema

```
for (;;) {
    accept_mutex_on 0;
    for (;;) {
        fd_set accept_fds;

        FD_ZERO (&accept_fds);
        for (i = first_socket; i <= last_socket; ++i) {
            FD_SET (i, &accept_fds);
        }
        rc = select ((last_socket+1, &accept_fds, NULL, NULL, NULL));
        if (rc < 1) continue;
        new_connection = -1;
        for (i = first_socket; i <= last_socket; ++i) {
            if (FD_ISSET (i, &accept_fds)) {
                new_connection = accept (i, NULL, NULL);
                if (new_connection != -1) break;
            }
        }
        if (new_connection != -1) break;
    }
    accept_mutex_off 0;
    process the new_connection;
}
```

## Apache Web Server: Configuració (I)



- Apache és molt flexible i configurable
- Les opcions de configuració reben el nom de *directrius*
- Existeixen 3 arxius de configuració
  - *httpd.conf*: fitxer principal, conté les directrius relatives al funcionament del servidor (autenticació, modes, logs...)
  - *srm.conf*: fitxer per configurar els recursos del sistema
  - *acces.conf*: fitxer que conté informació sobre l'accés als directoris del sistema
- Altres arxius de configuració, els *.htaccess*: a certs directoris, per determinar el comportament de les comandes referides al directori

## Apache Web Server: Configuració (II)



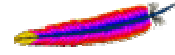
- Directrius: instruccions que indiquen al servidor com ha d'actuar. Tipus
  - *Context de configuració*: apareixen a qualsevol arxiu de configuració excepte a *contenidors* o *.htaccess*
    - Ex: `ServerType Standalone` # Forma d'execució del servidor
  - *Context de contenidor*: parell d'etiquetes que agrupa diferents directrius, restringint-ne així el seu abast
    - Ex. `<Directory /var/www>` # Opcions afecten al directori
      - `Options Indexes FollowSymLinks`
      - ...
      - `</Directory>`

## Apache Web Server: Configuració (III)



- *Context de directori*: arxius *.htaccess*
- *Context de mòdul*: directrius que permeten especificar el funcionament dels mòduls compilats per l'Apache
- Per comprovar la correctesa de l'arxiu de configuració:  
apachectl configtest

## Apache Web Server: Anàlisi (I)



### □ Principals punts forts d'Apache

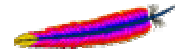
- és estable
- té moltes funcionalitats (gran diversitat de mòduls)
- és molt ràpid
- és portable, i està disponible per moltes plataformes
- és molt modular, i això permet ampliar-lo fàcilment i personalitzar-lo
- és de codi obert, i per tant, el seu cost és nul, i a més a més és possible modificar-lo segons les necessitats
- la seva reputació en seguretat és envejable

## Apache Web Server: Anàlisi (II)

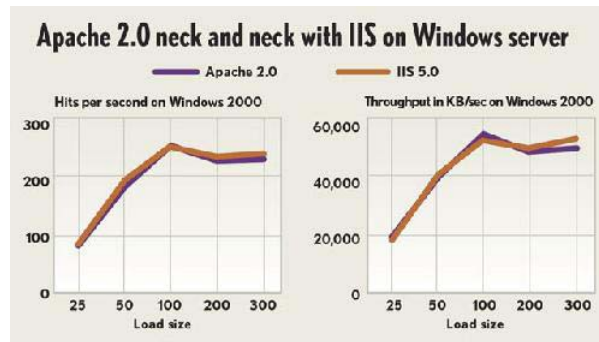


- Tot i que existeixen *benchmarks* que assegurin que altres servidors web de pagament són encara més ràpids, no s'ha demostrat encara que els tests realitzats siguin una bona manera de mesurar-ne la velocitat
- En la majoria de casos, i fins i tot en *sites* que reben mil·lions de peticions diàries, la rapidesa d'Apache és suficient
- El cost nul d'Apache fa que sigui preferible abans que un servidor de pagament més ràpid

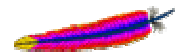
## Apache Web Server: Anàlisi (III)



- Tot i que en principi va ser desenvolupat per plataformes Unix, amb la inserció de codi natiu de Windows a la versió 2.0 s'ha aconseguit igualar la velocitat respecte IIS



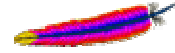
## Apache Web Server: Anàlisi (IV)



- Resum de les característiques d'Apache (<http://webcompare.internet.com>)

Server Features	
<input checked="" type="checkbox"/> Also acts as an HTTP proxy server	<input type="checkbox"/> Remote maintenance
<input checked="" type="checkbox"/> Can require password (Authorization: user)	<input checked="" type="checkbox"/> Scripting languages built-in or as modules (e.g. PHP, mod_perl)
<input checked="" type="checkbox"/> Can write to multiple logs	<input type="checkbox"/> Search engine
<input type="checkbox"/> Comes with SNMP agent	<input checked="" type="checkbox"/> Supports IPv6
<input type="checkbox"/> GUI-based maintenance	<input type="checkbox"/> Supports Microsoft ISAPI
<input type="checkbox"/> GUI-based setup	<input checked="" type="checkbox"/> Supports Non-IP Intensive Virtual Servers
<input checked="" type="checkbox"/> Has built-in scripting language	<input type="checkbox"/> Supports SNMP (1, 2c or secure v3) for management
<input checked="" type="checkbox"/> Includes full source code for server	<input type="checkbox"/> Supports SSL encryption in hardware
<input checked="" type="checkbox"/> Includes own API	<input type="checkbox"/> Supports SSL v. 3
<input type="checkbox"/> Integrated certificate server	<input checked="" type="checkbox"/> Supports WebDAV

# Apache Web Server: Anàlisi (V)



## □ Ranking servidors web (March 2003 Netcraft Survey Results)

March Results			
Server	No. of Servers in Use	Market Share	No. of Servers With .com domains
Apache	24,486,857	62.51%	10,817,915
Microsoft-IIS	10,748,324	27.44%	6,104,729
Zeus	794,889	2.03%	384,375
Unknown	591,441	1.51%	325,819
Netscape-Enterprise	408,116	1.04%	205,593
Rapidsite	399,136	1.02%	271,435
tigershark	291,124	0.75%	211,720
thttpd	237,113	0.61%	158,324
WebSTAR	89,905	0.23%	51,086
Lotus Domino	82,379	0.21%	30,501

## Bibliografia

### □ Links d'interès

- Consorci World Wide Web
  - <http://www.w3.org>
- Història d'Internet
  - <http://www.netvalley.com>
- Glossari de termes
  - <http://www.webopedia.com>
- Pàgina oficial Apache
  - <http://www.apache.org>