

Apunts d'Apache

Juan Aguilera

7 de febrer de 2016

1 Instal·lació del paquet apache

Versió d'apache

```
[^0%]juan@cilo:~$ sudo 'apache2 -v'
[sudo] password for juan:
Server version: Apache/2.4.10 (Debian)
Server built:   Aug 28 2015 16:28:08
```

d'entre les novetats de la versió apache 2.4 respecte les versions anteriors cap destacar que la directiva *NameVirtualHost* ja no és útil i és *deprecated*.

Una altra novetat és que es poden definir variables en la configuració.

Per més info, veure les new features d'apache a la pàgina oficial[1].

2 Configuració d'apache

2.1 fitxer de configuració principal

jugant amb l'index.html Traient l'index.html de /var/www podem veure els directoris que es troben a www. Això s'ha de fer perquè deu haver-hi per defecte un “DirectoryIndex index.html” i un “Option Indexes” si esborres el fitxer es mostra tot. Això és degut a que en apache2.conf ho han definit tres directives *Directory* per configurar com s'ha de comportar el servei quan s'intenta accedir a aquesta ruta:

```

access here, or in any related virtual
<Directory />
    Options FollowSymLinks
    AllowOverride None
    Require all denied
</Directory>

<Directory /usr/share>
    AllowOverride None
    Require all granted
</Directory>

<Directory /var/www/>
    Options Indexes FollowSymLinks
    AllowOverride None
    Require all granted
</Directory>

#<Directory /srv/>
#     Options Indexes FollowSymLinks
#     AllowOverride None
#     Require all granted
#</Directory>

```

2.2 fitxer de configuració de ports

És on especifiquem els ports on ha d'escollar el dimoni.

3 Principals directives

Tota la documentació sobre qualsevol directiva es pot trobar en la pàgina de referència de directives d'apache

Per entendre les directives a apache és important saber en quins contextos es poden utilitzar. Hi han els següents:

3.1 Context

Indica en quin àmbit o àmbits s'aplica la directiva:

server config This means that the directive may be used in the server configuration files (e.g., `apache2.conf`), but not within any `<VirtualHost>` or `<Directory>` containers. It is not allowed in `.htaccess` files at all.

virtual host This context means that the directive may appear inside `<VirtualHost>` containers in the server configuration files.

directory A directive marked as being valid in this context may be used inside `<Directory>`, `<Location>`, `<Files>`, and `<Proxy>` containers in the

server configuration files, subject to the restrictions outlined in Configuration Sections.

.htaccess If a directive is valid in this context, it means that it can appear inside per-directory .htaccess files. It may not be processed, though depending upon the overrides currently active.

3.2 Taula de directives

Directiva	Context	Descripció
AllowOverride	directory	Què directives es poden posar en el fitxer <i>.htacce</i> s. Per defecte el seu valor és none.
DirectoryIndex	server config, virtual host, directory, <i>.htaccess</i>	Diu què mostrar quan l'usuari demana una URL que fa referència a un directori i acabant amb una /
Options	server config, virtual host, directory, <i>.htaccess</i>	Configura qué característica està disponible en un directory en un àmbit en particular.
NameVirtualHost		DEPRECATED en apache 2.4
DocumentRoot	Server config; VirtualHost	Les adreces demanades pel client seràn relatives a aquesta ruta. Httpd servirà fitxers relatius a la ruta.
ServerRoot	server config	indica el directoi base per la configuració d'apache.
ServerName	server config; Virtualhost	Nom de servidor i port que el servei utilitza per identificar-se a si mateix.
ServerAlias	Virtualhost	Alias per el nom del servidor.
VirtualHost	server config	context per configurar un virtualhost.
Directory	server config; VirtualHost	Enclose a group of directives that apply only to the named file-system directory, sub-directories, and their contents
Location	server config; VirtualHost	Applies the enclosed directives only to matching URLs.
Files	server config; VirtualHost; Directory; <i>.htaccess</i>	Contains directives that apply to matched filenames. Example: <Files ~ '\.(gif png)\$'>.

Directiva	Context	Descripció
Allow, Deny, Order	Directory; .htaccess	DEPRECATED. Substituit pels <i>Require</i> , <i>RequireAny</i> — <i>All</i> — <i>None</i>
RequireAll	Directory; .htaccess	totes les directives <i>Require</i> dintre d'aquesta s'han de complir.
RequireAny	Directory; .htaccess	Alguna de les directives <i>Require</i> dintre d'aquesta s'han de complir. Si hi han directives <i>Require</i> que no estan dintre de les <i>Require*</i> aleshores s'interpreten com a que es troben en el <i>RequireAny</i>

Options Pot ser *None* o més d'entre molts valors d'entre els que destaquen:

- *FollowSymlinks*: Es permeten les peticions a enllaços simbòlics que hi hagin en el directori.
- *Indexes*: Si demanem una URL corresponent a un directori, mostra el que haguem escollit amb la directiva *DirectoryIndex* (en cas que no hi hagi *DirectoryIndex*).
- *All*: Tots.

Nota: Amb els signes *+* o *-* davant dels diferents valors que pot prendre *Options* voldrà dir que afegim o traiem llurs valors (si n'hi ha algun amb signe o han d'estar tots).

DirectoryIndex Lista de recursos que el cliente puede encontrar en un directorio dado cuando el cliente solicita la URL de ese directorio acabada en */*. En la lista podemos poner ficheros que estan en el directorio o rutas relativas que estan dentro del directorio. Al cliente se le devolverá el primero de los recursos que se encuentre en esa lista.

Está en estos contextos: server config, virtual host, directory, .htaccess.

Exemple:

```
DirectoryIndex index.html
then a request for
http://example.com/docs/
would return
http://example.com/docs/index.html
```

if it exists, or would list the directory if it did not.

Exemple (Note that the documents do not need to be relative to the directory):

```
DirectoryIndex index.html index.txt /cgi-bin/index.pl
```

would cause the CGI script

```
/cgi-bin/index.pl
```

to be executed if neither index.html or index.txt existed in a directory.

Si no encuentra ninguno de esos archivos, y está establecida la opción `Options Indexes` para ese directorio, el servidor generará y devolverá una lista, en formato HTML, de los subdirectorios y archivos del directorio. El valor predeterminado, almacenado en `/etc/apache2/apache2.conf`, es «index.html index.cgi index.pl index.php index.xhtml». Por tanto, si Apache2 encuentra un archivo en un directorio solicitado que se ajusta a alguno de esos nombres, se mostrará el primero de todos.

3.3 Taula de contextes

Context	Directives que podem posar-hi
Server config	Directives sobre el dimoni i el servei; ports d'escolta; Includes d'altres fitxers; Directoris generals; Què fer amb determinats fitxers genèrics.
.htaccess	Directives de tota mena que apliquen al directori on es troba el fitxer.
VirtualHost	<pre><VirtualHost *:80> DirectoryIndex index.html index.txt Option Indexes DocumentRoot /var/www/Bash ServerName www.bash.aguileraj.net ServerAlias www.aguileraj.net ServerPath error.html <Directory ps1> . . . </Directory> </VirtualHost></pre>
Directory	<pre><Directory ps1> Order Allow,Deny Allow from 192.168.199.0/24 Deny from 192.168.199.1 DirectoryIndex index.html Option Indexes <Files ~ ‘\.(gif jpe?g png)\$’> . . . </Files> </Directory></pre>

4 Control de l'accés als directoris

When multiple Require directives are used in a single configuration section and are not contained in another authorization directive like `RequireAll`, they are implicitly contained within a `RequireAny` directive. Thus the first one to authorize a user authorizes the entire request, and subsequent Require directives are ignored.

5 Apache com a proxy

Apache tant pot desempènyer la funcionalitat de proxy directe o proxy (*Forward proxy*) com la de proxy invers o gateway (*Reverse proxy*).

Principals mòduls

Per el mínim de funcionalitat:

- `mod_proxy`: mòdul necessari per a fer que Apache sigui un proxy.
- `mod_proxy_balancer`: per a que faci de balancejador de càrrega.
- `mod_cache` si el que volem és un proxy que faci també de servidor cau.
Amb aquest mòdul farà falta també el mòdul

Protocols que pot implementar com a proxy:

5.1 Forward Proxy

Principals directives

- `ProxyRequests`
ha d'estar a On. si el proxy ha de ser un proxy per llocs http o llocs ftp, necessitem els mòduls `mod_proxy_http` i el `mod_proxy_ftp`. Si els llocs són HTTPS, a més a més necessitem habilitat el `mod_proxy_connect`.
- `<Proxy>`
Per aplicar directives relatives al contingut que es vol accedir a través del proxy i que està indicat com a paràmetre al costat de la directiva. Per exemple: en la directiva `<Proxy http://cnn.com/*>` configurarem què s'ha de fer per accedir a `cnn.com` a través del proxy. Un altre exemple:

```
<Proxy '*'*>  
    Require host internal.example.com  
</Proxy>
```

Per accedir a qualsevol lloc a través del proxy cal ser del domini `internal.example.com`. Fixeu-vos que després, en el firewall, pot haver-hi una regla que només deixi sortir les consultes HTTP que surten del proxy.

5.1.1 Proxy com a servidor cau

Els objectius d'un servidor són els d'evitar total o parcialment les peticions que fan els clients als servidors web per a disminuir el tràfic i augmentar l'ample de banda. Si no millora significativament el rendiment un servidor cau no serveix de res. Hem de tenir clar que muntar un servidor cau és costós. És necessari que aquest tingui un molt bon accés a internet per a respondre amb celeritat a les peticions dels clients.

A tenir en compte en el tràfic web:

- Round trips (anades i tornades) entre client i servidor web
- Ample de banda
- latència: desfase o temps que triga el servidor en respondre o en processar una petició HTTP un cop l'ha rebuda.

Mòduls i directives

Per Apache com a servidor cau, el mòdul que hem de fer servir és `mod_cache`, `mod_cache_disk` i/o `mod_file_cache` (servidor cau enmagatzema el contingut a la memòria per servir-lo més ràpidament) junt amb els mòduls que necessita com a proxy. Principals directives del mòdul `mod_cache`:

- `CacheRoot`: directori arrel on estarà el cau
- `CacheEnable`: quin contingut guardem al cau i a on.
- `CacheDisable`: quin contingut no hem de guardar.
- `CacheIgnoreCacheControl`: a on, ignora la petició de no utilitzar la cache per a aquest contingut.
- `CacheIgnoreHeaders`: No guardem les capçaleres de les respostes al cau.
- `CacheIgnoreNoLastMod`: Normalment, les respostes sense la capçalera Last Modified no són guardades al cau. Amb aquesta directiva a on ignorem aquest fet.
- `CacheIgnoreQueryString`: Normalment, les respostes amb diferent query string es consideren diferents. Amb aquesta directiva a on ignorem aquest fet i donem de resposta qualsevol contingut encara que defereixi de la query string.
- `CacheMaxExpire`: el màxim temps en segons que guardem un document.
- `CacheMinExpire`: el mínim temps en segons que guardem un document.
- `CacheStorePrivate|NoStore|Expired`: guardem a la cache contingut encara que tingui la directiva private, no store o que sigui obsolet.

I aquí enumerem algunes de `mod_cache_disk`:

- `CacheDirLevels`: Nombre de nivell de subdirectoris.

- **CacheDirLength:** Longitud del nom dels directoris.
- **CacheMax|MinFileSize:** tamany màxim — mínim del fitxer.
- **CacheRoot:** Ruta del directori de la cache.

Exemples de configuració

Configuració de proxy.conf. Per accedir a qualsevol lloc a través del proxy s'ha de tenir com a ip la 192.168.1.0:

```
ProxyRequests On
<Proxy *>
    AddDefaultCharset off
    Require ip 192.168.1
    Require all denied
</Proxy>
```

HTTP Headers que intervenen en un servidor cau

- **Cache-control:** Sobreesciu els algorismes de cau sobre un contingut en particular. Pot estar en el client com en el servidor. Pot portar molts valors o directives. Entre elles, la directiva *max-age*.
- **Age:** Capçalera que inclou un servidor cau en la seva resposta si el contingut el serveix ell informant de l'edat d'aquest. O sigui, si en una resposta apareix aquesta capçalera, vol dir que el contingut no ve de primera mà. NO obstant això, aquesta capçalera només la inclouen els servidors que aconsegueixen el protocol HTTP/1.1, per lo que l'absència d'aquesta capçalera també pot voler dir que el contingut ha sigut donat per un servidor cau HTTP/1.0.
- **Date:** Des de la resposta, és el temps en el que aquesta fou generada.
- **Expires:** Data en la qual el contingut es considera obsolet.

Com decidir si un contingut és encara fresc o si és obsolet

Hem de saber l'edat, *Age* d'un element i si aquesta edat supera el temps de vida en el que es considera que un contingut és fresc, *freshness lifetime*.

Com calcular l'edat d'un contingut

El servidor original envia la capçalera *Date* en les seves respostes donant el temps en el que la resposta fou generada. Per tant, l'edat és el temps en el que porta el contingut a la cache des de la resposta del servidor original.

Com calcular *freshness lifetime*

La directiva *max-age* té prioritat sobre la capçalera *Header*. Per lo que si aquesta directiva existeix:

`freshness_lifetime = max_age_value`

En canvi, si la capçalera *Expires* es troba en la resposta i *max-age* no:

`freshness_lifetime = expire - date`

Si la directiva ni la capçalera existeixen aleshores el servidor cau calcularà mitjançant un càlcul heurístic, un valor per el *freshness lifetime*.

Finalment, determinem si un contingut és obsolet o no amb la següent comprovació:

`response_is_fresh = (freshness_lifetime > current_age)`

Què pot ser contingut susceptible d'estar en cau

1. Caching must be enabled for this URL. See the CacheEnable and CacheDisable directives.
2. The response must have a HTTP status code of 200, 203, 300, 301 or 410.
3. The request must be a HTTP GET request.
4. If the response contains an *Authorization:* header, it must also contain an *s-maxage*, *must-revalidate* or *public* option in the *Cache-Control:* header, or it won't be cached.
5. If the URL included a query string (e.g. from a HTML form GET method) it will not be cached unless the response specifies an explicit expiration by including an *Expires:* header or the *max-age* or *s-maxage* directive of the *Cache-Control:* header, as per RFC2616 sections 13.9 and 13.2.1.
Nota: *s-maxage* és com el max-age però el sobreescriu si hi és.
6. si en *Cache-control* hi ha *private*, en principi no és *cachejable* a no ser que tingui Apache la directiva *CacheStorePrivate*.
7. si en *Cache-control* hi ha *no-store*, en principi no és *cachejable* a no ser que tingui Apache la directiva *CacheStoreNoStore*.
8. A response will not be stored if it includes a *Vary:* header containing the match-all “*”.

En el cau s'ha de tenir en compte també si el contingut és *Multiviews*. Per exemple, si té diferents idiomes. Amb la capçalera *Vary* es determina quines són les capçaleres que s'han de tenir en compte. Per exemple:

Vary: negotiate,accept-language,accept-charset

Cau de tres estats vs cau de dos estats

En un servidor cau de dos estats tenim contingut fresc o obsolet i que, per tant, s'esborra del disc. En un servidor de tres estats un contingut, en canvi, pot ser:

- **Fresh:** EL servidor cau el pot servir directament del disc perquè és suficientment nou.
- **Stale:** Contingut massa vell. El servidor cau ha de connectar amb el servidor original i comprovar si el contingut és encara fresc abans de servir-lo al client. El servidor original respondrà amb el contingut nou o amb un missatge dient que el contingut que té la cache es pot servir.
- **Non-existent:** S'ha de demanar el contingut al servidor original.

Com s'enmagatzema el contingut en l'Apache Es pot decidir si guardar el contingut en disc o en memòria. No obstant això, encara que haguem decidit guardar el contingut en disc, els continguts més sol·licitats estaran temporalment en memòria per a un ràpid accés.

Per Una URL guardada a disc es crea un directory amb un hash de la url que inclou també els valors depenents de la capçalera *Vary*.

CacheDirLevels indica els nivells de subdirectoris i el *CacheDirLength* indica la longitud dels noms dels directoris.

htcacheclean comanda per controlar l'espai del contingut al cau. Es pot executar com a comanda o com a dimoni. Si s'usa com a comanda, pot executar-se periòdicament amb cron o posant-la en el fitxer */etc/rc.local*.

Evitar massives peticions al servidor original per un mateix contingut obsolet

Amb Apache això es pot evitar marcant un contingut com a *Lock* i tenir-lo en un directori per a aquests tipus de continguts. Aleshores, si arriba una altra petició per el mateix contingut ja no es torna a enviar una nova petició al servidor demanant per una actualització d'aquest i així només es fa una vegada. Les peticions que arriben després de la primera esperen aleshores a que arribi el refresc i es donen aleshores del cau. D'això se'n diu *Thundering Herd*. Directives associades:

- **CacheLock:** Ha d'estar a on.
- **CacheLockPath:** Ruta del directori on es deixa el contingut bloquejat.
- **CacheLockMaxAge:** Temps màxim en segons en el que el contingut pot estar bloquejat.

registres del servidor cau

```
CustomLog 'cached-requests.log' common env=cache-hit
CustomLog 'uncached-requests.log' common env=cache-miss
CustomLog 'revalidated-requests.log' common env=cache-revalidate
CustomLog 'invalidated-requests.log' common env=cache-invalidate
```

Problemes amb el servidor cau

Com hem vist abans, les decisions que pren el servidor cau per saber si un contingut és fresc o no és a través de les capçaleres de resposta. O sigui, que si aquestes capçaleres no estan ben informades el rendiment del servidor cau no serà òptim. És més, podria ser que fins i tot relanteixi el servei a contingut per internet.

En aquest cas és necessari tenir la capacitat de poder reescriure les capçaleres de les respostes, i això amb Apache es pot fer amb els mòduls `mod_headers` i `mod_expires`.

<http://www.websiteoptimization.com/speed/tweak/cache/> <http://stackoverflow.com/questions/9933012/how-to-use-mod-headers-and-mod-expires-to-cache>

Apache traffic server: <https://docs.trafficserver.apache.org/en/stable/getting-started/index.en.html> <https://trafficserver.apache.org/>

5.2 Reverse Proxy o gateway

mira wiki en español

Principals directives

ProxyPass ProxyRequests ha d'estar a Off.

A FAQ i Definicions generals

Els nombres màgics El fitxer `/etc/apache2/magic` és el llistat de nombres màgics que té apache per identificar tipus de fitxers. Un nombre màgic és un nombre posat en un lloc al començament d'un fitxer i que serveix per identificar el tipus de fitxer.

En instal·lar el paquet Apache no està escoltant peticions ipv4

És el que passa en el cas de Debian i versions d'Apache 2.2 i 2.4 aparentment, perquè realment sí que està rebent peticions de ipv4 (es mapejen les direccions ipv4 en format ipv6).

Si el que vull és explícitament que apache escolti per un determinat port ipv4, faré:

```
Listen 0.0.0.0:80
```

Veure la documentació d'apache referent al binding[1]

URI vs URL Les URL són un tipus de URI. Les URL defineixen un recurs d'internet mitjançant la seva localització. Per exemple, tota URI que vagi precedida per "http:" s'una URL perquè denotarà una localització d'un recurs a internet. Un altre tipus de URI són les URN. Aquestes defineixen el nom d'un recurs a internet. Per exemple, l'isbn a internet podria estar representat per una URN del tipus, per exemple, isbn:n-nn-nnnnnn-n. Això no denota cap lloc, sinó un nom. Veure més a aquest article de w3c[2].

Cada petició HTTP d'un client genera una nova connexió Aquesta característica és pròpia de HTTP1.1.

Diferències entre forward proxy i reverse proxy La diferència principal és que mentre el forward proxy s'encarrega de fer d'intermediari dels clients i de protegir la identitat d'aquests, el reverse proxy s'encarrega de fer d'intermediari d'un o d'un conjunt de servidors i de protegir la identitat d'aquests.

En un forward proxy és el proxy el que envia la petició al servidor i en l'invers els clients envien les peticions al proxy en comptes de fer-ho directament als servidors. El proxy, per la seva banda, té la capacitat de seguir el fluxe de peticions i respostes des dels clients cap als servidors i al revés.

Dir també que amb un proxy reverse podem fer un balancejador de càrrega entre servidors que poden oferir idèntic servei.

Veure més a aquest article de Jscape[3].

B VirtualHost FAQ

Se puede hacer un virtualhost como `www.bash.net` y otro como `www.asix.net` o ha de ser `www.bash.aguileraj.net` y `www.asix.aguileraj.net`?
Sí. El que el Virtualhost estigui en un domini totalment diferent o no que el mateix servidor no afecta al funcionament. El procés és el següent:

1. preguntamos a dns por FQDN y este busca por dominio
2. Nos dan una ip
3. Vamos por ip
4. Apache mira url y entonces determina qué virtualhost es

Referències

[1] <http://httpd.apache.org/docs/>

[2] <http://www.w3.org>

[3] <http://www.jscape.com/blog>