

# Web Programming

## 8) Web Server

*Emmanuel Benoist*  
Fall Term 2013-14

## State of the art

### What are the servers of the internet?

Developer	Nov 2011	Percent	Dec. 2011	Percent	Change
Apache	341,880,662	65.00%	362,267,922	65.22%	0.22
Microsoft	81,261,099	15.45%	82,521,809	14.86%	-0.59
nginx	44,731,780	8.50%	49,143,289	8.85%	0.34
Google	17,749,748	3.37%	18,464,148	3.32%	-0.05

(Source : <http://news.netcraft.com/archives/category/web-server-survey/>)

## Web Server: Principles and Configuration

- Introduction
  - Presentation of Apache
  - Principles of a Web Server
- Apache
  - Installation
- Apache Configuration
  - Network and Connexions
  - File Handling
  - Log Files
  - Control Instructions
  - Authorizations and Access Control
- Virtual Hosts

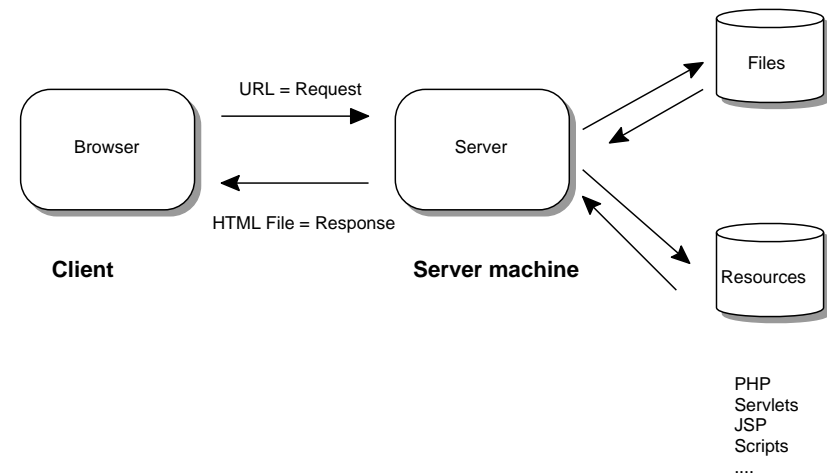
## What is APACHE ?

- ▶ **HTTP Daemon**
- ▶ **APACHE = a Patchy Server**
- ▶ **Great compatibility with HTTP/1.1**
- ▶ **Very fast to execute**
- ▶ **Real Modularity of the configuration**
- ▶ **Possibility to have virtual servers**
- ▶ **Almost 70% of the servers in the world**
- ▶ **Possibility of a SSL server**
- ▶ **Multiplatforms : Unix, Windows, Mac-OS X, OS/2, Amiga, ...**

# How does a web server run ?

- ▶ A “Server” application (daemon) listens to a communication port
- ▶ standard port: 80 (< 1024 hence only root can access)
- ▶ “Master” server: runs with root id; listens to port 80
- ▶ “Slave” servers: are lunched by the master and run with an other ID (default is nobody)
- ▶ Treatment of a communication
  1. The master receives a connection on port 80,
  2. it lanches the slave and transfer the communication canal
  3. the slave treats the request(s) and send the response back

# World Wide Web - Client Server Architecture



# Web Server vs File System

## On a standard LAMPP server (XAMPP for linux)

- ▶ A request for the directory : “/”  
goes to : “/opt/lampp/htdocs/”
- ▶ A request for the image : “/img/logo.jpg”  
goes to : “/opt/lampp/htdocs/img/logo.jpg”
- ▶ A request for the directory : “/~bie1/”  
goes to : “/home/bie1/public\_html”
- ▶ Request for “/~bie1/../../etc/passwd”

Bad Request

Your browser sent a request that this server could not understand.

Invalid URI in request GET /~bie1/../../etc/passwd HTTP/1.0

Apache/1.3.9 Server at localhost Port 80

# How to install Apache Server

- ▶ Download sources on the [www.apache.org](http://www.apache.org) server
- ▶ Uncompress files in a temporary directory
- ▶ *Read the documentation*
- ▶ Run configure and Set up the configuration of your server
- ▶ Compile
- ▶ Configure your server (its httpd.conf file)
- ▶ Run the server and see the page : “It Worked”

Or more simply: **download and install XAMPP**

# Configure the server

- ▶ **Configuration of the HTTP Daemon** : httpd.conf
- ▶ Port, user ID, root of the server, timeouts, virtual servers, number of clients, etc.
- ▶ Virtual Hosts can be included or imported from separate files

# httpd.conf : network

- ▶ **ServerName** : name returned to the client (must be in the DNS);
- ▶ **Port** : port used by the master (80 for the standard configuration);
- ▶ **BindAddress** : Used to limit IP addresses (if more than one);
- ▶ **Listen** : replaces Port and BindAddress if more than one IP address;
- ▶ **HostnameLookups** : *on* or *off*, reserch of the *name* of the client;
- ▶ **User and Group** : user ID and group ID for slave processus;
- ▶ **ServerAdmin** : email of the administrator.

# http.conf : connections

**Timeout** : maximum delay of a seponse (cgi, php, network access, nfs);

**Keepalive** : *on* or *off*;

- ▶ **MaxKeepAliveRequests** : number of requests for a connection;
- ▶ **KeepAliveTimeout** : maximum delay between two request in a connection;

## Server-Pool Size Regulation (MPM specific)

- ▶ **StartServers**: number of server processes to start
- ▶ **MinSpareServers**: minimum number of server processes which are kept spare
- ▶ **MaxSpareServers**: max nb of server processes which are kept spare
- ▶ **MaxClients**: max nb of server processes allowed to start
- ▶ **MaxRequestsPerChild**: max nb of requests a server process serves

# http.conf : connections (Cont.)

```
<IfModule prefork.c>
StartServers      5
MinSpareServers   5
MaxSpareServers   10
MaxClients        150
MaxRequestsPerChild  0
</IfModule>
```

## httpd.conf : files

- ▶ **ServerRoot** : root directory of the server  
`ServerRoot "/home/bie/"`
- ▶ **DocumentRoot** : root directory of the files  
`DocumentRoot "htdocs"`  
or  
`DocumentRoot "/home/bie/htdocs"`
- ▶ **UserDir** : root of user home (corresponds to `http://<server>/~<user>`)  
`UserDir www`  
or  
`UserDir public_html`
- ▶ **PidFile** : where does httpd put its *pid* when started.  
`<IfModule !mpm_netware.c> PidFile logs/httpd.pid  
</IfModule>`

## Log files : access\_log

**File** : `/var/apache/logs/www.hes-be.ch_transfer.log`

```
128.173.188.78 - - [13/Apr/2000:13:39:12 +0200]
"GET /portrait/facts.html HTTP/1.0" 200 17898
"http://www.hes-be.ch/portrait.html"
"Mozilla/4.61 [en] (WinNT; U)"

128.173.188.78 - - [13/Apr/2000:13:39:13 +0200]
"GET /portrait/navportrat/fachhochschulenn.gif HTTP/1.0" 200 180
"http://www.hes-be.ch/portrait/facts.html"
"Mozilla/4.61 [en] (WinNT; U)"

216.35.116.65 - - [13/Apr/2000:14:25:35 +0200]
"GET /page32/f.html HTTP/1.0" 404 276
"--" "Slurp.so/1.0 (slurp@inktomi.com;
      http://www.inktomi.com/slurp.html)"

147.87.65.34 - - [13/Apr/2000:14:37:41 +0200]
"GET /~bie HTTP/1.0" 404 267
"--" "Mozilla/4.7 [en] (X11; I; SunOS 5.7 sun4u)"
```

## httpd.conf : Log files

### Log files

- ▶ **ErrorLog** : name of the file containing the log of errors ;  
`ErrorLog logs/error_log`
- ▶ **LogLevel** : Level of the message for the error\_log (debug, info, notice, warn, error, crit, alert, emerg)  
`LogLevel warn`
- ▶ **LogFormat** Defines a new format for log files  
`LogFormat "%h %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"\n\n\" \"%{User-Agent}i\"\" combined`  
`LogFormat "%h %l %u %t \"%r\" %>s %b" common`  
`LogFormat "%{Referer}i -> %U" referer`  
`LogFormat "%{User-agent}i" agent`
- ▶ **CustomLog** : name and format of access files  
`CustomLog logs/access_log common`

## Log files : error\_log

**File** : `/var/apache/logs/www.hes-be.ch_error.log`

```
[Thu Apr 13 07:22:59 2000] [error] [client 204.123.13.65]
File does not exist:
/usr/local/www/apache/vhosts/www.hes-be.ch/data/f.html

[Thu Apr 13 08:23:22 2000] [error] [client 62.53.27.209]
File does not exist:
/usr/local/www/apache/vhosts/www.hes-be.ch/data/e.html

[Thu Apr 13 14:25:35 2000] [error] [client 216.35.116.65]
File does not exist:
/usr/local/www/apache/vhosts/www.hes-be.ch/data/page32/f.ht
```

# File system

**Alias** : enables to use a directory that is not under the document root

```
Alias /icons/ "/home/bie/webserver/icons/"
```

# Default

- ▶ **DirectoryIndex** : default index file (for instance index.html)
- ▶ **FancyIndexing** how to display a directory  
IndexOptions FancyIndexing VersionSort
- ▶ **AddIconByType** : Add icon to show for different mime type of file  
AddIconByType (TXT,/icons/text.gif) text/\*  
AddIconByType (IMG,/icons/image2.gif) image/\*
- ▶ **Addicon** : Tells wich icon to use for a filename extension  
AddIcon /icons/tar.gif .tar
- ▶ **DefaultIcon** : When no type is found  
DefaultIcon /icons/unknown.gif
- ▶ **IndexIgnore** : files that will not be displayed (e.g. .toto toto~)  
IndexIgnore .??\* \*~ \*# HEADER\* README\* RCS CVS \*,v \*,t

# Document types

- ▶ **AddType** : allows to tweak mime.type without editing it (useful for php)  
# since LAMPP 1.0RC1  
AddType application/x-httpd-php .php .php3 .php4
- ▶ **DefaultType** : default type of an unknown file  
DefaultType text/plain
- ▶ **AddEncoding** : allows you to have browsers uncompress on the fly  
AddEncoding x-compress Z  
AddEncoding x-gzip gz tgz
- ▶ **AddLanguage** : to specifie the language of a document  
AddLanguage fr .fr  
AddLanguage he .he  
AddLanguage hr .hr

# Treatments

**Redirect** : to redirect a directory to a server : /toto/ will be redirected to http://serv.toto.ch/

```
Redirect old-URI new-URL
```

**AddHandler** : allows to map file extensions to "handlers"

```
AddHandler cgi-script .cgi  
AddHandler imap-file map
```

**Action** : define media types that will execute a script whenever a matching file is called.

```
Format: Action media/type /cgi-script/location  
Format: Action handler-name /cgi-script/location
```

**ErrorDocument** : Customizable error response

```
ErrorDocument 404 /missing.html  
ErrorDocument 404 /cgi-bin/missing_handler.pl  
ErrorDocument 402 http://some.other_server.com/subscribe_info.htm
```

## Control on the operations

- ▶ Some orders are only valid for a specific directory

```
<Directory dir_name> orders </Directory>

<Directory "/opt/lampp/htdocs">
  Options Indexes FollowSymLinks ExecCGI Includes
  AllowOverride All
  Order allow,deny
  Allow from all
</Directory>
```
- ▶ Or if a module was loaded (they are only valid for this module)

```
<IfModule module_name> orders </IfModule>

<IfModule dir_module>
  DirectoryIndex index.html index.html.var index.php
</IfModule>
```

## Control on the operations (Cont.)

- ▶ Or for files matching a specific regular expression

```
<FilesMatch regExp> orders </FilesMatch>

<FilesMatch "~\.ht">
  Order allow,deny
  Deny from all
</FilesMatch>
```

## Authorizations

- ▶ **Options** : what can we do – This may be "None", "All", or any combination of "Indexes", "Includes", "FollowSymLinks", "ExecCGI", or "MultiViews".

Options Indexes FollowSymLinks MultiViews

- ▶ **AllowOverride** : what .htaccess files can override.

AllowOverride None

## Access Control

**Order** : what to test first

**Deny**: Machines that are not allowed

**Allow**: Machine that are allowed

```
<Directory /home/*/public_html>
  AllowOverride FileInfo AuthConfig Limit Indexes
  Options MultiViews Indexes SymLinksIfOwnerMatch IncludesNoExec
  <Limit GET POST OPTIONS PROPFIND>
    Order allow,deny
    Allow from all
  </Limit>
  <LimitExcept GET POST OPTIONS PROPFIND>
    Order deny,allow
    Deny from all
  </LimitExcept>
</Directory>

<Files ~ "\.ht">
  Order allow,deny
  Deny from all
</Files>
```

## Protection with password

- ▶ **In the httpd.conf**

AccessFileName : name of the file for access control (in each directory)

```
AccessFileName .htaccess
```

- ▶ **File .htaccess**

```
AuthName Any_Name
```

```
AuthType Basic
```

```
AuthUserFile <absolute path>/.htpasswd
```

```
AuthGroupFile <absolute path>/.htgroup
```

```
<Limit GET POST>
```

```
require group group1 group2
```

```
require user emmanuel jean
```

```
</Limit>
```

## Protection with password (Cont.)

- ▶ **File .htgroup**

```
group1 toto1 toto2 toto3
```

```
group2 titi1 titi2
```

- ▶ **File .htpasswd** : generated by the program *bin/htpasswd*

```
toto1:GZomFT8JmJGsc
```

```
toto2:FQ3UzRLn52/sw
```

```
toto3:<coded password>
```

```
titi1:<coded password>
```

```
titi2:<coded password>
```

```
emmanuel:<coded password>
```

```
jean:<coded password>
```

## Virtual Hosts

### Two domains for one server

```
<VirtualHost www.smallco.com>
```

```
ServerAdmin webmaster@mail.smallco.com
```

```
DocumentRoot /groups/smallco/www
```

```
ServerName www.smallco.com
```

```
ErrorLog /groups/smallco/logs/error_log
```

```
CustomLog logs/smallco.com-access_log common
```

```
</VirtualHost>
```

```
<VirtualHost www.baygroup.org>
```

```
ServerAdmin webmaster@mail.baygroup.org
```

```
DocumentRoot /groups/baygroup/www
```

```
ServerName www.baygroup.org
```

```
ErrorLog /groups/baygroup/logs/error_log
```

```
CustomLog logs/baygroup.org-access_log common
```

```
</VirtualHost>
```

## Virtual Hosts : in biel

### httpd.conf

```
### Section 3: Virtual Hosts
```

```
#
```

```
NameVirtualHost 147.87.65.7
```

```
Include vhosts/enigma.hta-bi.bfh.ch/vhost.conf
```

```
Include vhosts/www.bfh.ch/vhost.conf
```

```
Include vhosts/www.hes-be.ch/vhost.conf
```

```
Include vhosts/www.abegg.bfh.ch/vhost.conf
```

```
Include vhosts/www.hgkk.bfh.ch/vhost.conf
```

```
Include vhosts/www.hta-bi.bfh.ch/vhost.conf
```

```
Include vhosts/sims.hta-bi.bfh.ch/vhost.conf
```

```
Include vhosts/webaccess.hta-bi.bfh.ch/vhost.conf
```

```
Include vhosts/sun-file.hta-bi.bfh.ch/vhost.conf
```

# Virtual Hosts : in biel school of engineering (Cont.)

File : vhosts/enigma.hta-bi.bfh.ch/vhost.conf

```
<VirtualHost 147.87.65.7>
  ServerName      enigma.hta-bi.bfh.ch
  ServerAdmin     sungroup@hta-bi.bfh.ch
  RewriteEngine   On
  RewriteLogLevel 9
  ErrorLog        /var/apache/logs/enigma.hta-bi.bfh.ch_error.log
  CustomLog       /var/apache/logs/enigma.hta-bi.bfh.ch_transfer.log combined
  RewriteLog      /var/apache/logs/enigma.hta-bi.bfh.ch_rewrite.log
  DirectoryIndex  index.html e.html
  AccessFileName  .htaccess
  RewriteRule     ~/icons/(.*) /usr/local/www/apache/icons/$1 [L]
  RewriteRule     ~/(.*) /usr/local/www/apache/vhosts/enigma.hta-bi.bfh.ch/data/$1 [L]
  <Directory /usr/local/www/apache/vhosts/enigma.hta-bi.bfh.ch/data>
    Options       Indexes FollowSymLinks
    AllowOverride None
    Order         allow,deny
    Allow         from all
  </Directory>
</VirtualHost>
```

# Conclusion

## ▶ Web Server Configuration

- ▶ Similar for Tomcat or other servers
- ▶ Matching of Request with Resource?
- ▶ Which Response must be given for a given file?

## ▶ Lots of possible configuration

- ▶ Virtual servers
- ▶ HTTPS servers (with double direction recognition)
- ▶ Integration of numerous modules
- ▶ mod\_access, mod\_actions , mod\_alias, mod\_asis, mod\_auth, mod\_auth\_anon, mod\_auth\_db, mod\_auth\_dbm, ...