

LinuxChix

Apache

Serving Webpages

The layer 7 protocol (HTTP) is what our browsers talk to get us the websites we can't seem to live without.

HTTP is written as a server/client protocol; the client is called the browser, the server is called the web server.

Different servers exist: we shall use Apache – but lighthttpd, zope, etc are examples for UNIX, IIS is an example one for windows.

Apache: an intro

- Written as a patch to the original NSCA httpd webserver.
- Evolved into a huge project and the “patch” is now huger than the original code.
- Now runs on Windows NT environments as well as UNIX.
- Netcraft surveys show that Apache is the most common (announced) web server.. see http://news.netcraft.com/archives/web_server_survey.html

Installing Apache

- We shall do this from ports. Two different branches are being maintained: 1.3 and 2.x. 1.3 is deprecated so we shall use 2.x
- FreeBSD ports maintain different versions so that people can migrate their applications when they are ready.
 - # cd /usr/ports/www/apache22
 - # make all install clean
- Edit /etc/rc.conf to turn apache on.
apache_enable=“YES”

httpd.conf

- This is the file which tells apache how to behave. It has the general format of directives and their values separated by spaces.
- Lines beginning with '#' are ignored so are blank lines.
- Directives can be made more specific to files, directories, conditionals, etc using sections which have a format similar to html tags viz:
 <section_name options>
 SpecificDirective thevalue
 </section_name>
- The configuration syntax is well documented in the file as well as <http://httpd.apache.org/docs/2.2/>

Sections & Directives of interest

- ServerRoot: This is the directory your server's system lives in; also configuration details which don't specify a full path will be relative to this path.
- DocumentRoot: When someone requests for a webpage, the server needs to know how to map that to a folder in the filesystem. This directive tells it where to start.
- ServerName: set this to the dns name of the server
- ServerAdmin: email address of the webmaster
- ErrorLog and CustomLog: paths to log files. Very usefull for debugging problems with apache setup etc

Sections & Directives ctd:

- DirectoryIndex: tells apache which file/object to serve by default when a client requests for a website if none is specified. Usual setting is index.html but can be anything. Multiple files are separated by spaces and tried in order (left to right)
e.g
 DirectoryIndex index.html index.html home.htm

Common Pitfalls with configuring apache

- Spelling errors (use apachectl configtest after making changes)
- Wrong hostname (set in /etc/rc.conf)
- Improper file permissions (See the ErrorLog and CustomLog)
- With that in mind, modify the ServerName and ServerAdmin variables, and start apache using:
 # /usr/local/etc/rc.d/apache22.sh start
- From the command line do
 \$ lynx localhost
 If all is well, you should be able to see an "It works" page

UnderThe Hood: Brief http protocol discussion

- HTTP messages consist of:
 - Headers
 - Blank line
 - Body
- When a client requests for a web object (page/image/...) the server sends a response. Both have the same format shown above; difference is the allowed contents of the Headers and the Body.
- As soon as the response is sent, the webserver closes the connection. As such, HTTP is said to be “stateless”.

Sample session

```
$ telnet inst.ws.linuxchix.or.ke 80
GET / HTTP/1.1
Host: inst.ws.linuxchix.or.ke
<blank_line>
```

```
HTTP/1.1 200 OK
Date: Thu, 15 Mar 2007 13:33:05 GMT
Server: Apache/2.2.0 (FreeBSD) mod_ssl/2.2.0 OpenSSL/0.9.7e-p1
DAV/2
Last-Modified: Sat, 20 Nov 2004 20:16:24 GMT
ETag: "41286b-2c-4c23b600"
Accept-Ranges: bytes
Content-Length: 44
Content-Type: text/html
```

```
<html><body><h1>It works!</h1></body></html>Connection closed by
foreign host.
```