

Services: Apache – a patchy web server

David Morgan

© David Morgan 2003,2004

The Whole Business

“...the whole business of a web server is to translate a URL either into a filename, and then send that file back over the Internet, or into a program name, and then run that program and send its output back. That is the meat of what it does: all the rest is trimming.”

Apache, The Definitive Guide, page 1

© David Morgan 2003,2004

A Web Site = a Directory

- RedHat default: `/var/www`
- Holds “what to send” to connecting clients
 - content of files in: `/var/www/html`, or
 - output of programs in: `/var/www/cgi-bin`

© David Morgan 2003,2004

FILE CONTENT
PROGRAM OUTPUT

Subliminal message

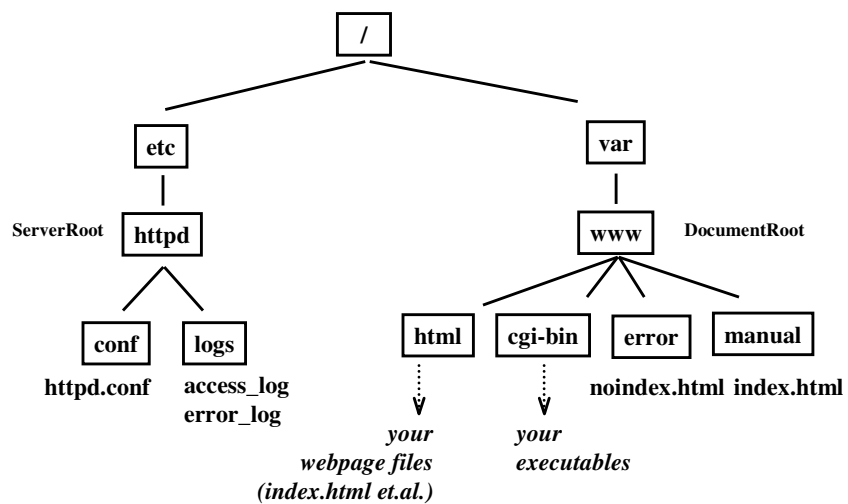
© David Morgan 2003,2004

4 important directories

- For HTML scripts /var/www/html
- For programs /var/www/cgi-bin
- For config files /etc/httpd/conf
- For logs /etc/httpd/logs -> /var/log/httpd

© David Morgan 2003,2004

Default directories and files



© David Morgan 2003,2004

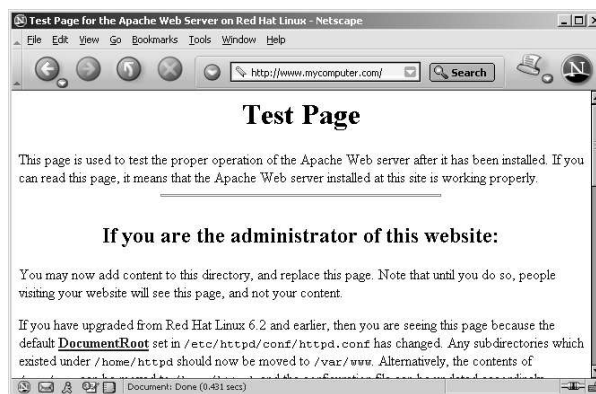
FILE CONTENT PROGRAM OUTPUT

Subliminal message

© David Morgan 2003,2004

Placeholder default test page

- /var/www/error/noindex.html
- appears as long as administrator supplies no /var/www/html/index.html



© David Morgan 2003,2004

Apache manual

- /var/www/manual/
- installed by default



© David Morgan 2003,2004

Primary config file: httpd.conf

- Houses “directives”
- Categorized
 - global directives – apache process
 - default host directives – main “site”
 - virtual host directives – other (2nd, 3rd, ...) sites
- directives number about 150

© David Morgan 2003,2004

Turning services on and off

- Services re-read configuration files when restarted
- Starting
 - `/etc/rc.d/init.d/<script for service> start` *or*
 - **`service <script for service> start`**
- Stopping
 - `/etc/rc.d/init.d/<script for service> stop` *or*
 - **`service <script for service> stop`**

© David Morgan 2003,2004

Turning apache on and off

- Re-reads `/etc/httpd/conf/httpd.conf` when restarted
- Starting
 - `/etc/rc.d/init.d/httpd start` *or*
 - **`service httpd start`**
- Stopping
 - `/etc/rc.d/init.d/httpd stop` *or*
 - **`service httpd stop`**

© David Morgan 2003,2004

Running programs - CGI

- A way to get a “page,” to send to a client
- Alternative to getting it out of a file (“.htm”)
- Instead call a program to generate it
- Write the program to output webpage-legal stuff
- CGI stands for “common gateway interface”
(which could gain from improvement, as it’s creatively greatly inhibited)

© David Morgan 2003,2004

FILE CONTENT PROGRAM OUTPUT

Subliminal message

© David Morgan 2003,2004

Running a CGI program

- Put the program file in /var/www/cgi-bin
- apache interprets files in cgi-bin as “runnables”
- Run by naming its URL in
 - “location/address” in client/browser
 - ACTION=<url> within an HTML FORM statement

© David Morgan 2003,2004

Virtual (multiple) servers

- IP-based
 - separate IP address for each virtual server
- Name-based
 - just one address for all virtual servers

© David Morgan 2003,2004

IP-based: multiple IP addresses

- multiple physical interfaces, each with its own IP
 - replicate everything, or
- multiple “ip aliases” on a single physical interface
 - replicate the virtual, not the physical

© David Morgan 2003,2004

IP-based: number of daemons

- Multiple daemons
 - allows separate configurations
 - performance-expensive
- Single daemon
 - must share single configuration
 - performance-efficient

© David Morgan 2003,2004

IP-based virtual servers

```
<VirtualHost www.smallco.com>
ServerAdmin webmaster@mail.smallco.com
DocumentRoot /groups/smallco/www
ServerName www.smallco.com
ErrorLog /groups/smallco/logs/error_log
TransferLog /groups/smallco/logs/access_log
</VirtualHost>
```

```
<VirtualHost www.baygroup.org>
ServerAdmin webmaster@mail.baygroup.org
DocumentRoot /groups/baygroup/www
ServerName www.baygroup.org
ErrorLog /groups/baygroup/logs/error_log
TransferLog /groups/baygroup/logs/access_log
</VirtualHost>
```

in /etc/httpd/conf/httpd.conf © David Morgan 2003,2004

IP-based virtual servers

- Requires distinct name resolution
 - www.smallco.com → 111.22.33.44
 - www.baygroup.org → 111.22.33.45
- Both addresses for the webserver machine
- Apache differentiates by address
- Destination address embedded in IP header

© David Morgan 2003,2004

Name-based virtual servers

```
NameVirtualHost 111.22.33.44

<VirtualHost 111.22.33.44>
  ServerName www.domain.tld
  DocumentRoot /www/domain
</VirtualHost>

<VirtualHost 111.22.33.44>
  ServerName www.otherdomain.tld
  DocumentRoot /www/otherdomain
</VirtualHost>
```

in /etc/httpd/conf/httpd.conf

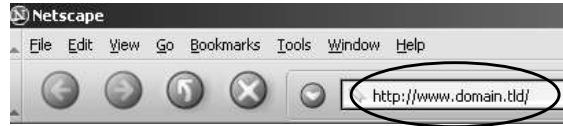
© David Morgan 2003,2004

Name-based virtual servers

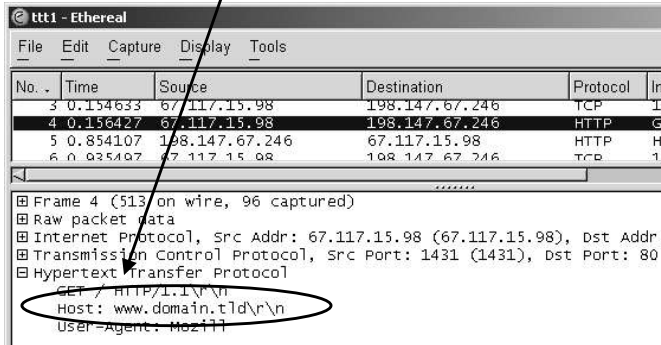
- Requires common name resolution
 - www.domain.tld → 111.22.33.44
 - www.otherdomain.tld → 111.22.33.44
- Apache differentiates by name
- Destination name embedded in HTTP request

© David Morgan 2003,2004

Differentiation by name

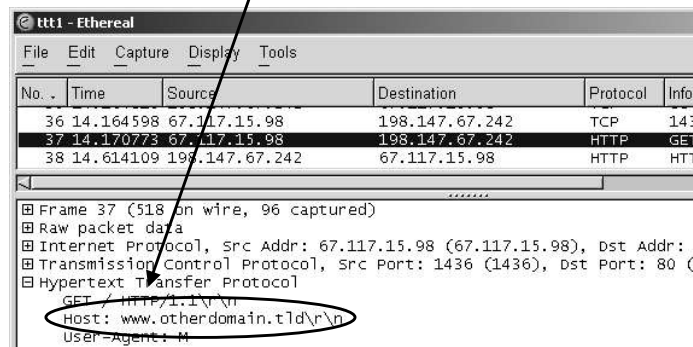
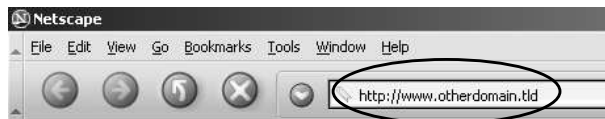


Destination domain name embedded in HTTP request



© David Morgan 2003,2004

Differentiation by name



© David Morgan 2003,2004

Some noteworthy directives

- Global
 - ServerRoot
 - Listen
- Host
 - User, Group
 - ServerAdmin
 - ServerName
 - DocumentRoot
 - UserDir
 - DirectoryIndex

© David Morgan 2003,2004

ServerRoot

```
# ServerRoot: The top of the directory tree under which the server's
# configuration, error, and log files are kept.
#
ServerRoot "/etc/httpd"
```

from /etc/httpd/conf/httpd.conf

© David Morgan 2003,2004

Listen

```
# Listen: Allows you to bind Apache to specific IP addresses and/or
# ports, in addition to the default. See also the <VirtualHost>
# directive.

#Listen 12.34.56.78:80
Listen 80
```

from /etc/httpd/conf/httpd.conf

© David Morgan 2003,2004

User, Group

```
# If you wish httpd to run as a different user or group, you must run
# httpd as root initially and it will switch.
#
# User/Group: The name (or #number) of the user/group to run httpd as.
#
User apache
Group apache
```

from /etc/httpd/conf/httpd.conf

```
desktop:x:80:80:desktop:/var/lib/menu/kde:/sbin/nologin
postgres:x:26:26:PostgreSQL Server:/var/lib/pgsql:/bin/bash
apache:x:48:48:Apache:/var/www:/sbin/nologin
postfix:x:89:89::/var/spool/postfix:/sbin/nologin
squid:x:23:23::/var/spool/squid:/dev/null
```

from /etc/passwd

© David Morgan 2003,2004

ServerAdmin

```
# ServerAdmin: Your address, where problems with the server should be
# e-mailed. This address appears on some server-generated pages, such
# as error documents. e.g. admin@your-domain.com
#
ServerAdmin webmaster@linnet.edu
```

from /etc/httpd/conf/httpd.conf

© David Morgan 2003,2004

ServerName

```
# ServerName gives the name and port that the server uses to identify itself.
# This can often be determined automatically, but we recommend you specify
# it explicitly to prevent problems during startup.
#
# If this is not set to valid DNS name for your host, server-generated
# redirections will not work. See also the UseCanonicalName directive.
#
# If your host doesn't have a registered DNS name, enter its IP address here.
# You will have to access it by its address anyway, and this will make
# redirections work in a sensible way.
#
ServerName hostz.linnet.edu:80
```

from /etc/httpd/conf/httpd.conf

© David Morgan 2003,2004

DocumentRoot

```
# DocumentRoot: The directory out of which you will serve your
# documents. By default, all requests are taken from this directory,
# but
# symbolic links and aliases may be used to point to other locations.
#
DocumentRoot "/var/www/html"
```

from /etc/httpd/conf/httpd.conf

© David Morgan 2003,2004

UserDir

```
# UserDir: The name of the directory that is appended onto a user's home
# directory if a ~user request is received.
#
# The path to the end user account 'public_html' directory must be
# accessible to the webserver userid. This usually means that ~userid
# must have permissions of 711, ~userid/public_html must have permissions
# of 755, and documents contained therein must be world-readable.
# Otherwise, the client will only receive a "403 Forbidden" message.
UserDir public_html
```

from /etc/httpd/conf/httpd.conf

© David Morgan 2003,2004

DirectoryIndex

```
# DirectoryIndex: sets the file that Apache will serve if a directory  
# is requested.  
#  
DirectoryIndex index.html
```

from /etc/httpd/conf/httpd.conf

© David Morgan 2003,2004

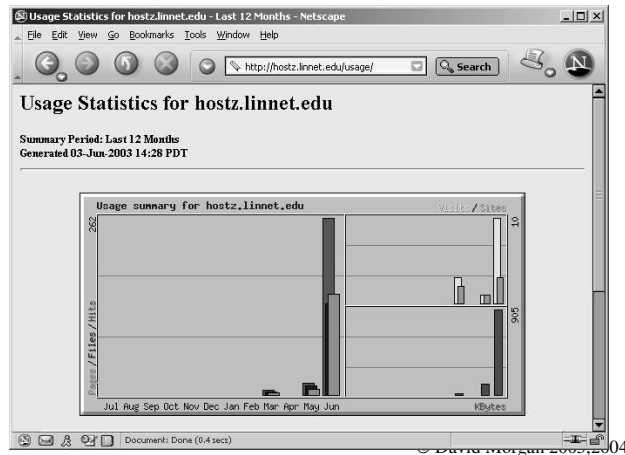
Status codes

- returned in status line of httpd reply messages
- found in httpd.h
- well-known examples
 - 200 – OK
 - 404 – not found
 - 500 – internal server error

© David Morgan 2003,2004

Webalizer

- run “webalizer” on server
- deposits analytics in /var/www/html/usage/
- view by browser



FILE CONTENT PROGRAM OUTPUT

Subliminal message

© David Morgan 2003,2004