

Splitting lighttpd Logs With vlogger And Creating Statistics With Webalizer

By Falko Timme

Published: 2008-01-29 18:43

Splitting lighttpd Logs With vlogger And Creating Statistics With Webalizer

Version 1.0

Author: Falko Timme <ft [at] falkotimme [dot] com>

Last edited 01/14/2008

Vlogger is a little tool with which you can write lighttpd logs broken down by virtual hosts and days. With vlogger, we need to put just one *accesslog.filename* directive into our global lighttpd configuration, and it will write access logs for each virtual host and day. Therefore, you do not have to split lighttpd's overall access log into access logs for each virtual host each day, and you do not have to configure lighttpd to write one access log per virtual host (which could make you run out of file descriptors very fast).

At the end of this tutorial I will show you how to use webalizer to create statistics from the lighttpd access logs.

I do not issue any guarantee that this will work for you!

1 Preliminary Note

I have tested vlogger on a Debian Etch system where lighttpd is already installed and working.

2 Installing And Configuring vlogger

To install vlogger, we simply run

```
apt-get install vlogger
```

Afterwards, we have to modify the *accesslog.filename* line in */etc/lighttpd/lighttpd.conf* and add an *accesslog.format* line that works with vlogger:

```
vi /etc/lighttpd/lighttpd.conf
```

```
[...]  
##### accesslog module  
accesslog.filename = "|/usr/sbin/vlogger -s access.log /var/log/lighttpd"  
accesslog.format = "%v %h %V %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-Agent}i\""  
[...]
```

Please disable all other *accesslog.filename* and *accesslog.format* directives in your lighttpd configuration, especially in the vhost configurations!

The advantage of writing just one access log is that this lowers the load on the server a lot, especially if you have some high-traffic sites on your server.

Now restart lighttpd:

```
/etc/init.d/lighttpd restart
```

Vlogger will now create subdirectories in the */var/log/lighttpd* directory, one per virtual host, and it will create access logs that contain the current date in the file name. It will also create a symlink called *access.log* that points to the current log file.

Let's assume we have two virtual hosts, *www.example.com* and *www.test.tld*. Then this is how the */var/log/lighttpd* directory will look like:

```
/var/log/lighttpd/  
    www.example.com/  
        06042007-access.log  
        06052007-access.log  
        06062007-access.log  
        access.log -> 06062007-access.log  
    www.test.tld/  
        06042007-access.log
```

```
06052007-access.log
06062007-access.log
access.log -> 06062007-access.log
```

To learn what other vlogger command line directives you can put into the `accesslog.filename` line, take a look at

```
man vlogger
```

3 Creating Statistics With webalizer

In this chapter I will show you how you can create statistics from the splitted log files with webalizer. Again, I'm assuming that you have two virtual hosts, `www.example.com` and `www.test.tld`, and these virtual hosts have the document roots `/var/www/www.example.com/web` and `/var/www/www.test.tld/web` (it's important that the server names are in the document root paths, otherwise the following procedure won't work). I'd like to put the statistics into the directories `/var/www/www.example.com/web/stats` and `/var/www/www.test.tld/web/stats`, so these must already exist.

First, let's install webalizer:

```
apt-get install webalizer
```

Take a look at

```
man webalizer
```

to see how webalizer works. Basically, to create statistics for `www.example.com` from yesterday's access log, you can use this command:

```
/usr/bin/webalizer -c /etc/webalizer/webalizer.conf -n www.example.com \  
  
-s www.example.com -r www.example.com -q -T -o /var/www/www.example.com/web/stats \  
  
/var/log/lighttpd/www.example.com/`/bin/date -d "1 day ago" +%m%d%Y`-access.log
```

(`/etc/webalizer/webalizer.conf` is the location of Debian's default `webalizer.conf`. `/bin/date -d "1 day ago" +%m%d%Y` prints yesterday's date exactly the way we need it so that we can pass yesterday's `access.log` to webalizer without needing to know the exact date.)

Of course, we don't want to run such a command manually for each virtual host, therefore we write a little shell script that reads the `/var/log/lighttpd` directory and creates statistics for each virtual host that has logs in that directory. I name the script `webstats` and place it in the `/usr/local/sbin` directory:

```
vi /usr/local/sbin/webstats
```

```
#!/bin/sh

logdir=/var/log/lighttpd
webalizerconf=/etc/webalizer/webalizer.conf
yesterdaysdate=`/bin/date -d "1 day ago" +%m%d%Y`

cd ${logdir}
for directory in *
do
  if [ -d ${directory} ]; then
    /usr/bin/webalizer -c ${webalizerconf} -n ${directory} \
    -s ${directory} -r ${directory} -q -T -o /var/www/${directory}/web/stats \
    ${logdir}/${directory}/${yesterdaysdate}-access.log
  fi
done

exit 0
```

We must make that script executable:

```
chmod 755 /usr/local/sbin/webstats
```

Finally, we create a cron job that calls the `/usr/local/sbin/webstats` script every night at 04.00h:

```
crontab -e
```

```
0 4 * * * /usr/local/sbin/webstats &> /dev/null
```

After the cron job has run for the first time, you can go to `http://www.example.com/stats` and `http://www.test.tld/stats` to see the statistics in your browser.

4 Links

- vlogger: <http://n0rp.chemlab.org/vlogger>
- webalizer: <http://www.mrunix.net/webalizer>
- Debian: <http://www.debian.org>