How To Read Your Server's Access Log

by
Will Bontrager
Willmaster.com

Every request for a file on your server is logged. Web pages, images, CSS files – every file. Whether by browser, spider, robot, or other software, the request is logged.

I'll show you how to read the log file where all requests are logged. It is called the access log.

The server's access log described here is the log file created by the Apache server. Most Unix/Linux domain hosting computers use Apache. Other operating systems may also use Apache.

The Server's Access Log

The server can maintain many different log files. What we're working with here is only the access log.

If you don't know where your server's access log is located, your hosting company can tell you.

Some servers are set up to maintain 3 logs with every request: An access log. A referrer log. And a user-agent log. Other servers maintain all that information in one log file, a combined access log.

This article assumes the combined access log. The same information generally can be found in the 3 mentioned individual log files if your server is set up that way.

Each line of the combined access log file contains certain information. Each item of information is separated with a space. If an item of information is unknown, a hyphen is printed in its place.

1. IP address

Every request is accompanied by an IP address. The IP address is where the response is sent to.

2. Machine Identity

The identity of the client's machine (the computer the browser/spider/etc. software is running on) is unreliable and unlikely to be used on your server unless it is part of a tightly controlled internal network.

3. User

The user if logged in through Apache authentication.

4. Date and Time

The date, time and time zone between square brackets.
`[day/month/year:hour:minute:second zone]`

5. Request

The request from the client (browser, spider, etc.) between double quotes. The request line contains three space-separated items of information.

(I) The request method – GET, POST, or HEAD.

(II) The location of the requested resource and, if applicable, the query string – the URI of file, script, or other file being requested. If a query string accompanies the request, it is appended to the URI.

(III) The request protocol – generally HTTP/1.0 or 1.1, but can also be FTP or other protocol.

6. Status Code

The status code sent back to the client (browser, spider, etc.)

7. *Size*

The size of the object (source code file, image, etc.) sent back to the client (browser, spider, etc.)

8. *Referrer*

The referrer information accompanying the request between double quotes, if available. This is the URL the browser reports it was referred by, generally the URL of the web page with the link that was clicked or the URL of the web page that requires an external file (like a CSS or image file).

9. *User-Agent*

The user-agent information accompanying the request between double quotes, if available. This is the identifying information that the client (browser, spider, etc.) reports about itself.

How To Read The Log

Obtain a copy of your server's access log file and open it in a plain text processor, one that will not wrap the log entries but maintain the lines as they are in the file.

If you do not know where your server's access log file is located, ask your hosting company.

NOTE: Log files can become huge. If the file is too large to load into your text processor, use the data mining software available free at the [Server Log Treasures](#) article.

When the log file has been opened in your text processor, you'll see a line for each request. The lines contain the information listed in the previous section.

The IP address is an important bit of information. Because each request is logged, a browser loading one web page can result in many log entries. The web page is only one request. The browser makes a separate request for every image on the page. Same with every other external file required to render the page, such as CSS or JavaScript files.

With the IP address, the separate requests related to one web page can be perceived as a group.

Here are 6 lines from the lightfocus.com log file, all by the same IP address. (Because the log entries are so long, their lines are wrapped here, with a blank line between each, for ease of reading.)

```
74.216.112.159 - - [25/Sep/2010:14:00:27 -0500]
"GET /recipes/How-to-Make-a-Very-Good-Pie.php HTTP/1.1" 200 12044
"http://www.google.ca/url?q=how%20long%20to%20let%20a%20pie%20cool"
"Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; InfoPath.2)"


74.216.112.159 - - [25/Sep/2010:14:00:28 -0500]
"GET /lfstyle.css HTTP/1.1" 200 3965
"http://lightfocus.com/recipes/How-to-Make-a-Very-Good-Pie.php"
"Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; InfoPath.2)"


74.216.112.159 - - [25/Sep/2010:14:00:28 -0500]
"GET /images/wmlogolightfocus_sm.gif HTTP/1.1" 200 2001
"http://lightfocus.com/recipes/How-to-Make-a-Very-Good-Pie.php"
"Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; InfoPath.2)"


74.216.112.159 - - [25/Sep/2010:14:00:28 -0500]
"GET /images/dot.gif HTTP/1.1" 200 44
"http://lightfocus.com/recipes/How-to-Make-a-Very-Good-Pie.php"
"Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; InfoPath.2)"


74.216.112.159 - - [25/Sep/2010:14:00:29 -0500]
"GET /terra/thblue-midnite.jpg HTTP/1.1" 200 2992
"http://lightfocus.com/recipes/How-to-Make-a-Very-Good-Pie.php"
"Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; InfoPath.2)"


74.216.112.159 - - [25/Sep/2010:14:00:30 -0500]
"GET /favicon.ico HTTP/1.1" 200 3638
"-"
"Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; InfoPath.2)"
```

In the first of the 6 log entries, the /recipes/How-to-Make-a-Very-Good-Pie.php web page is requested. Google is the referrer, a click from "how long to let a pie cool" search results. (Because the query string of the referrer URL was so long, all but the pertinent information was deleted).

The next log entry is a request for the CSS file for the web page. The referrer is the web page requested in the first log entry.

The next 3 log entries are requests for images belonging on the page. The referrer for each image request is the web page requested in the first log entry.

The last log entry is for the favicon.ico file, made after all the web page elements were requested. The favicon.ico image is used in the browser's address bar and in bookmarks if the web page is bookmarked.

Each log entry represents one request. The elements of the log entry are those listed in the previous section.

Besides browsers asking for web pages, other interesting things can be seen in the server's access log. The next section talks about a few.

Aggregators, Indexers, and Facebook User Links

The example log entries are all from lightfocus.com. As above, the log entries are wrapped, with a blank line between each, for ease of reading.

Here are access log entries from requests by several blog indexers and aggregators.

```
173.192.238.41 - - [25/Sep/2010:03:04:37 -0500]
"GET /index.xml HTTP/1.1" 304 - "-"
"Mozilla/5.0 (X11; U; Linux x86_64; en-US; rv:1.9.0.19;
aggregator:Spinn3r (Spinn3r 3.1); http://spinn3r.com/robot)
Gecko/2010040121 Firefox/3.0.19"


72.14.199.90 - - [25/Sep/2010:03:10:44 -0500]
"GET /index.xml HTTP/1.1" 304 - "-"
"Feedfetcher-Google; (+http://www.google.com/feedfetcher.html; 4
subscribers; feed-id=1180860485387498323)"


66.235.116.128 - - [25/Sep/2010:03:29:20 -0500]
"GET /index.xml HTTP/1.1" 304 - "-"
"Bloglines/3.1 (http://www.bloglines.com; 3 subscribers)"
```

Legitimate spiders, including indexers and aggregators, generally provide a valid URL within the user-agent information. The web page at the URL has information about the spider.

You'll find a URL in each of the above log entries that contain more information about that particular spider.

The following 3 log entries show that 3 search engine spiders made requests for web pages – Google's Googlebot, Yahoo!'s Yahoo! Slurp, and Bing's msnbot.

```
66.249.71.75 - - [25/Sep/2010:03:11:25 -0500]
"GET /lightstories/ HTTP/1.1" 200 7634 "-"
"Mozilla/5.0 (compatible; Googlebot/2.1;
+http://www.google.com/bot.html)"

67.195.111.176 - - [25/Sep/2010:03:18:24 -0500]
"GET /2002travel-log/Park-Rapids-MN.php HTTP/1.0" 200 6563 "-"
"Mozilla/5.0 (compatible; Yahoo! Slurp/3.0;
http://help.yahoo.com/help/us/ysearch/slurp)"

207.46.13.96 - - [25/Sep/2010:03:42:48 -0500]
"GET /robots.txt HTTP/1.1" 200 213 "-"
"msnbot/2.0b (+http://search.msn.com/msnbot.htm)"
```

You'll see a URL in each of the user-agent information fields of the above log entries. They contain information about the particular spider that made the request.

Here is a log entry made when Facebook's external hit spider requested a page.

```
66.220.149.251 - - [25/Sep/2010:03:42:02 -0500]
"GET /recipes/Vegetable-Beef-Soup-Recipe.php HTTP/1.1" 206 6280 "-"
"facebookexternalhit/1.1
(+http://www.facebook.com/externalhit_uatext.php)"
```

Following the URL in the user-agent field, you'll see a web page that says a Facebook user sent a link to other Facebook users.

Mining Data

The [Server Log Treasures](#) article contains free software source code to extract server access log entries containing certain information.

Examples:

1. Count the number of requests for a specific web page with a certain referring URL.

2. Find out how many times a specific web page was requested with a certain code. (A code can be appended to a URL as a query string – a question mark character followed by the code. Such as http://example.com/tools.html?ad_21. Do not use this method if the page at the URL already expects a query string to process, as the extra query string may cause unexpected behavior.)

3. Find out how often the Yahoo! Slurp spider visits your web site.

Pick up the free data mining software at the [Server Log Treasures](#) article.

About [Willmaster.com](#)

Willmaster.com has been creating [free and commercial software](#) for Internet web sites since late 1997. The [Possibilities](#) ezine has been published almost that long.

One thing people comment on is the good support we provide.

Another thing commented on is the high quality of the software. It just works. Like it's supposed to.

The willmaster.com web site has many hundreds of pages of [library](#) and [blog](#) articles. You'll find tips and how-to's for web site owners. You'll find code for software written in Perl, PHP, and JavaScript, ready to copy and paste. With implementation instructions. Use the search box on the site to find what you want.

In short, [Willmaster.com](#) has lots of good stuff for web site owners.

See you there.

Will Bontrager
[Willmaster.com](#)