

Measuring Your Website

Understanding Web Metrics

Web Logs

Every time you visit a website your every move is being recorded.

Although no personally identifiable information is gathered during this process (your name, email address etc), your entry point, the paths you travel, the pages you browse, the duration of your stay, and your exit point are all saved in a web log file for later analysis.

Server Log Format

Web logs are usually saved in a 'raw' text form that is all but impossible to read without the aid of a log analyzer.

The job of a web analyzer is to demystify these logs and present the statistical information in meaningful ways. Logs present their information in a 'string' of text, one for each file you request.

Server Log Format

There are a number of different log file formats and it's crucial you use a log analyzer that can read the appropriate format delivered by your server. The most common formats are:

- Common Access Log Format
- W3C Extended
- Apache/NCSA Combined
- Microsoft IIS
- or a Custom format specific to a particular server

What is Recorded?

- **Host** — the ip address of the computer making the request. (this is almost always the ip address instead of the host name because looking up the host name slows down server performance).
- **Date and Time** — when the request was made.

What is Recorded?

- **Request Line** — What the client told the server it wanted

Method: Get or Post

Protocol: HTTP

Resource: absolute URL of page, graphic or query requested

What is Recorded?

- **Status Code** — the code corresponding to the final status of the request. These are some that are common:

200s: request was successful.

300s: redirect to another address

400s: Error, file not found or unauthorized or forbidden

- **Size** — bytes Served

Server Log Entry

```
68.142.250.29 - - [01/Apr/  
2006:03:24:26 -0600] "GET /students/  
daljase/2006/02/07/here-comes-the-  
bride/ HTTP/1.0" 200 14692
```

What is Recorded?

- **Optional Parameters** — these are not required, but are tracked because they provide useful data for measuring your audience:

Referer: the site from which the browser says it came.

User-Agent: The browser kind (mozilla, safari, internet explorer,), language, operating system, and version

User Agent Entry

Mozilla/5.0 (Macintosh; U; PPC
Mac OS X; en) AppleWebKit/
418 (KHTML, like Gecko)
Safari/417.9.2

Log File Analyzers

- There are both free and commercial programs that will analyze your server logs.
- The better ones provide charts and graphs that make it easy to interpret the data.
- Some use cookies to track returning visitors and other statistics not found in the server log file.

Common Measurements

- “Hits” are not an accurate measurement since each graphic counts as a hit on each web page.
- “Page views” are considered more accurate since they count individual pages served.

Common Measurements

- “Unique visitors” is one way to count how many people visited your site, but it is not fool proof. It represents requests from specific IP addresses. Multiple computers can use an IP address and multiple people can use each computer. Still, it is the best metric we have for measuring how many “people” have visited your site.

Why use logs?

- Find out what the most popular pages on your site are.
- Find out where your visitors are coming from.
- Identify errors and fix them.
- Track promotions
- Improve content based on page views.

What you should know:

- The main types of information captured in server log files: host, date and time, request (method, protocol and resource), status code and size
- Main optional parameters that can be captured: referrer and user-agent
- A log analyzer is used to summarize an entire log file or multiple log files that cover a specified period of time.

What you should know:

- The difference between "hits" and "page views"
- What "unique visitors" represent