

# Web Usage Statistics Definitions

## Hits

Any request made to the server which is logged, is considered a 'hit'. The requests can be for anything...html pages, graphic images, audio files, CGI scripts, etc... Each valid line in the server log is counted as a hit. This number represents the total number of requests that were made to the server during the specified report period.

## Files

Some requests made to the server, require that the server then send something back to the requesting client, such as an html page or graphic image. When this happens, it is considered a 'file' and the files total is incremented. The relationship between 'hits' and 'files' can be thought of as 'incoming requests' and 'outgoing responses'.

## Pages

Pages are, well, pages! Generally, any HTML document, or anything that generates an HTML document, would be considered a page. This does not include the other stuff that goes into a document, such as graphic images, audio clips, etc... This number represents the number of 'pages' requested only, and does not include the other 'stuff' that is in the page. What actually constitutes a 'page' can vary from server to server. The default action is to treat anything with the extension '.htm', '.html' or '.cgi' as a page. A lot of sites will probably define other extensions, such as '.phtml', '.php3' and '.pl' as pages as well. Some people consider this number as the number of 'pure' hits...Some other programs (and people :) refer to this as 'Pageviews'.

## Sites

Each request made to the server comes from a unique 'site', which can be referenced by a name or ultimately, an IP address. The 'sites' number shows how many unique IP addresses made requests to the server during the reporting time period. This **DOES NOT** mean the number of unique individual users (real people) that visited, which is impossible to determine using just logs and the HTTP protocol (however, this number might be about as close as you will get).

## Visits

Whenever a request is made to the server from a given IP address (site), the amount of time since a previous request by the address is calculated (if any). If the time difference is greater than a pre-configured 'visit timeout' value (or has never made a request before), it is considered a 'new visit', and this total is incremented (both for the site, and the IP address). The default timeout value is 30 minutes (can be changed), so if a user visits your site at 1:00 in the afternoon, and then returns at 3:00, two visits would be registered. **Note:** Visits only occur on PageType requests, that is, for any request whose URL is one of the 'page' types defined with the PageType option. Due to the limitation of the HTTP protocol, log rotations and other factors, this number should not be taken as absolutely accurate, rather, it should be considered a pretty close "guess".

## KBytes

The KBytes (kilobytes) value shows the amount of data, in KB, that was sent out by the server during the specified reporting period. This value is generated directly from the log file, so it is up to the web server to produce accurate numbers in the logs (some web servers do stupid things when it comes to reporting the number of bytes). In general, this should be a fairly accurate representation of the amount of outgoing traffic the server had, regardless of the web servers reporting quirks.

**Note:** A kilobyte is 1024 bytes, not 1000 :)