Concepts

Help Documentation

Concepts

Understanding Server Logs and SmarterLogs

SmarterStats was developed for a wide variety of users, from the webmaster with one or two sites to large-scale enterprises with 10,000 sites. With this philosophy in mind, SmarterStats incorporates a data structure that reduces the amount of raw log files stored on a web server.

In a nutshell, SmarterStats acts like an optimized file-based relational database by taking traditional web server log files and segregating redundant data. This step alone can reduce log file sizes up to 80%. Additionally, SmarterStats compresses the already small SmarterLogs even further, adding another 5-10% reduction in the size of the log files. This allows system administrators to easily archive, or even delete, the original log files created by the web server to conserve disk space on the server. However, if the original logs are ever needed, SmarterLogs can easily be exported in their original format.

Because the unique architecture results in smaller log files and smaller disk space requirements, the benefits include:

- Less disk space is required for log file storage
- Less backup space is required for backing up log files
- Shorter download times for clients
- Faster data queries and faster display times
- Faster backup schedules
- Lower overhead for monitoring backups
- Lower initial hardware investment

Supported Log Files

SmarterStats supports these common log file types: W3C, IIS, NCSA, NCSA Extended, Apache, Apache NCSA, and IPlanet Common Logs. Compressed logs in .GZ or .ZIP format are also supported. Rotating logs are also supported as of SmarterStats 8.x.

Understanding Report Items

SmarterStats breaks down virtually every report available within the interface into individual report items, which are considered to be the backbone of the SmarterStats application. Most of these report items display only one type of information pulled from the site's log files or Site Tuning reporting, and typically contain a single table of information and a chart, when applicable. Examples include Top Pages, Referring Sites, Issues by Type and more.

Custom reports are user-created reports that combine a number of individual report items into a single, comprehensive report to meet your business needs. For more information on custom reports, see Custom Reports .

Understanding Visits, Views, and Hits

A visit occurs when an individual or search engine spider comes to your website and looks around a bit. That person or bot may go to one page or may go to 100 pages. However, that visit happens only once per day.

A view refers to the number of pages that were seen during that visit.

A hit refers to the number of physical resources requested from the server during that visit. Pages are made up of many items like images, audio files, video files and even text. Each of those items that are presented to the visitor count as a hit, while the overall view is the presentation of the page in its entirety.

Imagine going to a convenience store and picking up a magazine to buy. You're essentially "visiting" the magazine simply by picking it up. While looking through the magazine, you flip through, looking at various pages, which are your page "views". Each page you look at has text and some pictures, and probably a few ads. Each of these items are "hit" simply by showing up on the page. Taking this all into account, you have one visit, you view many pages of the magazine and incur various hits based on the text and images you see on those pages.

So which value is more useful? It depends on the use of the information. If you're in marketing and you're striving for bragging rights or high activity numbers, use hits. However, if you are using your stats to actually identify trends and the personal impact of your site, use views and visits. Hits can always be inflated artificially by adding a few more images to a page, for example. Visits and views, however, represent people coming to your site and the number of webpages they saw, which is much more useful for analyzing usage patterns.

SmarterStats vs. Google Analytics A Comparison of Log File and Script-based Analysis for Accurate Website Statistics

Who Should Use This Document

This document provides a comparison of SmarterTools' SmarterStats web log analytics and SEO software and the Google Analytics website statistics service. The comparison is from the end user's point of view.

Overview

Regardless of whether a website is designed for a business or individual, it is important to collect and evaluate the site's analytics. At a very basic level, this involves tracking information about unique visits, page views, bounce rates and conversion rates. Analyzing this data will help you predict visitor behavior and modify your website design and/or marketing campaigns for improved success.

Methods for Gathering Website Analytics

The Web Analytics Association defines web analytics as "the measurement, collection, analysis and reporting of Internet data for the purposes of understanding and optimizing web usage." There are two main methods for collecting website analytics: log file analysis and script-based analysis.

Every time someone visits a website, the web server records information about every file request, i.e., the HTML files, CSS files, JavaScript files, graphic files, PDF documents, MP3 s, etc., and stores them in log files. These log files also include information on site and web server errors, page processing time, bandwidth used, visitor IP address, referring websites, and much more. Log file analyzers like SmarterStats interpret this data to create website statistics and reports.

Alternatively, script-based analyzers like Google Analytics require the use of tracking code (usually JavaScript) on each webpage that is to be tracked. As visitors surf the website, the code places a cookie on their computer so they have a unique identifier and can be tracked—provided the visitor is using a browser that supports JavaScript, has cookies enabled and is not using security software to protect the browser.

Log Analyzers vs. Script-based Analyzers

Each method for gathering website analytics has its benefits and drawbacks. For example, script-based analyzers only record data from pages that have the JavaScript tracking code; any pages without the code will not have statistics available. This is also one of the reasons log file analyzers report higher traffic (views, visits and hits) than Google Analytics does, as the web server log files track every interaction visitors have with a website.

Similarly, because log file analyzers process log file data instead of relying on forward-facing JavaScript, they have access to information that script-based analyzers don't (such as data relating to traffic from spiders and bots, server errors and bandwidth statistics).

Please refer to the following tables for further comparison on the strengths and weaknesses of log file analyzers and script-based analyzers.

Log File Analyzers		
Pros	Cons	

• Access to server-side information (404 pages, 500 errors, time taken, etc.) • Every resource is counted (images, RSS feeds, etc.) • Bandwidth information is available • Since logs are always stored, privacy software is limited in the information it can mask (browser type and referrer) • Provides the most accurate view of what is actually happening on the web server. • Visits by automated bots are tracked, which can reveal security problems or hack attempts, as well as search engine spider activity.	• Not as good at counting "live" users that may visit the website via proxies • No intrinsic ability to report on browser-side data (resolution, number of colors, etc.)	

Script-based Analyzers		
Pros	Cons	

• More accurate "live" visitor count (for webpages only), if all pages are tagged correctly • Access to browser-side data	• Anti-spyware software and security software now block JavaScript callback methods, leading to untracked users • No server-side information about the website can be collected • Requires more maintenance to get the site set up for analytics • Potential for forgetting to tag some pages with the tracking code, resulting in lost/incomplete analytics • No method for tracking items like downloads or RSS traffic • No information is stored on the server • No way to detect abusers/hack attempts • No access to historical data. Statistics start on the date the tracking code was implemented; data prior to that date is unavailable	

SmarterStats and Google Analytics Feature Comparison

Please refer to the chart below for a comparative list of key features available in SmarterStats and Google Analytics:

Features	SmarterStats	Google Analytics
Install and manage on your own server(s)	•	
Can be used on firewall-protected corporate Intranet	•	
Ability to reprocess historical data (from log files)	•	
Ability to process/reprocess log files locally	•	
Ability to collect information through tags		•
Reports on spider and bot activity	•	

Reports on server errors and status codes	•	
Tightly integrated with Google AdWords		•
Ability to report on paid search campaigns	•	•
Ability to report on banner and traditional marketing campaigns	•	•
Geo-targeted reporting	•	•
Visitor session/navigation path analysis	•	•
Access to raw data for custom report-building	•	
SEO reports to analyze search engine ranking	•	
Competitor analysis reports	•	
Free	• 1	• 2

1 SmarterStats Free edition is available for download from the SmarterTools website. However, SmarterStats Free is limited to a single site and a single SEO campaign.

2 Google Analytics limits the number of hits (pageviews) it will track for free. Larger sites with high traffic counts are required to link an active Google Adwords account that has at least one active and running campaign that is subject to a minimum budget of $1 per day. That boosts the cost of Google Analytics to a minimum of $30 per month for larger sites that want to continue collecting analytics once they reach the maximum number of monthly hits.

Accuracy of Statistics from SmarterStats and Google Analytics

As previously discussed, users of both analytics program may notice reporting discrepancies, particularly in regards to visitor counts. Because of the limitations of script-based analytics, Google Analytics will underreport or misrepresent some website statistics. Consider:

• When comparing data on website visits, it is important to realize that SmarterStats and Google Analytics process data at different intervals. Google Analytics uses a 30-minute window while SmarterStats uses a 20-minute idle timeout window. For this reason, SmarterStats will report a higher number of visits for visitors that keep the browser window open but do not navigate to new pages within the window.

• Look at the referral reports in SmarterStats. Are there any referrals listed that are part of the base domain (i.e., IP addresses or aliased domains)? These referrals are not available in Google Analytics reporting.

• Hits from people "borrowing" content from your website. Are people linking directly to documents or images on your website? SmarterStats counts any request for those files as visits

and hits; Google never reports them because JavaScript can't be placed on files, just HTML pages.

In the end, SmarterStats provides more accurate website statistics than Google Analytics because SmarterStats reports all requests to the server, not just what is tagged with tracking code.