



WHITE PAPER



COMPARING SITECATALYST METRICS

SiteCatalyst and Alternative Technologies

July 12, 2007

Version 2.0



1 Overview

SiteCatalyst is a browser-based, hosted data collection model. SiteCatalyst uses the most accurate data collection technology available for the web today. This technology is based on collecting data about web site pages directly from a user's browser. The data collection method is more accurate than older, web log or packet sniffing (some vendors call these “sensors”) solutions for several reasons, including the following.

- SiteCatalyst overcomes both proxy and browser caching of pages (which web logs do not measure).
- SiteCatalyst measures only pages loaded within a web browser and not automated traffic such as spiders and robots
- SiteCatalyst measures visitors more accurately than most log file based solutions, which rely upon IP addresses rather than cookies.

Most customers that move from other log file based solutions, or even other browser-based solutions, encounter some differences in metrics, including Page Views, Visits, and Unique Visitors. SiteCatalyst provides improved accuracy by eliminating false traffic generated by automated sources and increasing the visibility into actual site visitor traffic. In some cases, the business rules within SiteCatalyst may also produce different results from other products.

This document describes specific metric differences most customers should expect to encounter. Clients transitioning from one technology or system to another are challenged to communicate and educate their organizations about these differences.

1.1 Other Resources

In addition to this white paper, Omniture suggests the following resources in order to provide insight on differences from other reporting systems.

Omniture White Paper: Log Files vs. The ASP Model	This white paper provides a direct comparison between log file-based web analytic solutions and SiteCatalyst's ASP model. The white paper compares the differences in technologies, implementation effort, and data collection and accuracy.
Omniture Case Study: The American Marketing Association	This document describes a case study with an Omniture client, The American Marketing Association. AMA's CEO, Bob Wallach, discusses the differences - and added value - of SiteCatalyst compared to an older, log file-based solution that AMA previously used.
Omniture White Paper: Visits and Unique Visitors	This white paper describes the business rules used by SiteCatalyst to measure Visits and Unique Visitors. SiteCatalyst uses the most accurate methods and technologies available, including patent-pending methods for detecting a browser's ability to set cookies.
Omniture White Paper: Deployment Validation	This white paper describes step-by-step instructions to validate the SiteCatalyst implementation to ensure that the data collection is functioning correctly.

<p>Validation: Maxamine Site Quality Reporting</p>	<p>Maxamine (a third party service provider) is a site quality reporting service that includes a Web Beacon Report that can be used to scan an entire site and report any unimplemented pages. Contact Omniture Live Support for additional information about this service.</p>
<p>Omniture Validation Consulting</p>	<p>Omniture's Implementation Engineers assist customers in analyzing differences in metrics reported by SiteCatalyst and other systems. In many cases, the assistance begins with identifying potential differences (refer to <i>Comparison Strategies</i> in this document), as well as providing insight into specific differences in data by comparing limited amounts of the "raw" data collected by both systems.</p> <p>In many cases, it may also be necessary to request other vendors to provide additional information on processing rules and filters in order to completely quantify differences in metrics. Omniture cannot provide a valid comparison without detailed information about how your current reporting solution calculates its metrics, what business rules/filters are in place, and how log files are processed.</p>

1.2 Expected and Potential Differences

There are some major differences between data collection technologies (i.e. log or server-based solutions and browser-based solutions). Data collection technologies include standard web server log files, packet "sniffers" placed on web servers (or in the data center), hosted solutions that process log files, and other data collection methods that collect data at or near the web server. Data collection type results have similar differences when compared to SiteCatalyst.

<p>Overstated Pageviews</p>	<p>Pageviews may be higher in server-based solutions on "home" or entry pages (such as "http://www.yoursite.com/") due to site monitoring such as Keynote, Gomez, or other monitors, including internal IT or QA monitoring (which are included in server logs, but not in browser-based data collection). Since these automated monitors request pages from web servers very frequently (sometimes as often as every 15 seconds), and from multiple locations throughout the world, the differences on these pages can be significant. Note that in most cases "internal" pages of the site are not monitored as frequently, and therefore compare more closely to browser-based data collection. Identify all monitors, both internal and external, as a first step in comparing large differences in pageviews on key entry or "home" pages.</p>
<p>Missed Pageviews (Browser Cache)</p>	<p>Pages are cached by browsers, and may be reloaded from the cache instead of being reloaded from the server. While SiteCatalyst will measure these pageviews, servers will often not "see" these page requests. In some cases, this traffic is as much as 20-30% of traffic, especially representing the use of the Back Button in navigation of web sites.</p>
<p>Missed Pageviews (Server</p>	<p>Although less common than browser caching, servers between the</p>

Cache)	browser and the web site may also be caching pages. This is especially frequent on large ISPs or corporations that may cache a page locally, and use it for any additional page requests occurring within the network.
Accurately Identifying Visitors	SiteCatalyst uses cookies to measure Unique Visitors (by time period). Log file-based solutions often rely upon IP addresses (although some rely upon cookies - check with your vendor). IP addresses are often "pooled" (or shared) between many users. In fact, a single user might show up as more than one IP address during a single session. This is especially true for large ISPs, such as AOL, or large corporations (which translate internal networks addresses into a single public IP once outside the corporate LAN). SiteCatalyst generally measures fewer unique visitors than these solutions due to the use of cookies.
Time Zones	Ensure that the time zone setting within SiteCatalyst matches the time zone setting in the previous reporting solution. Additionally, ensure that all servers are sychronized and that all web logs processed by the system are time-synchronized prior to processing, especially if multiple data sources are imported.
Campaigns or Search Engines	Search engines, affiliate programs, banners ads, and other types of campaigns may report clickthrough metrics that do not match SiteCatalyst reports. Note that SiteCatalyst counts a "pageview" for each campaign code, search engine, or keyword for each time it is loaded. Often the entry page combined with the back button will report the same value multiple times for the same visitor. Use the "Visits" metric to determine the number of visits associated with the campaign or search keyword. Alternately, if SiteCatalyst is under-reporting a value, ensure that the campaign is correctly implemented, and that all possible "entry" pages for the campaign are correctly implemented and capturing the campaign codes.
Session (Visit) Lengths	SiteCatalyst uses an industry standard 30-minute Visit timeout length from the last pageview (the Visit timeout period may be customized by Omniture clients upon request). Other products may use a different session length, or may apply the timeout rule differently.
Effects of Testing or QA	SiteCatalyst does not record orders, page views, or processes that are automatically placed or executed. If the automated processes are completed in a standard web browser, this can result in overstating orders or other success events on the web site. This type of "internal" traffic can be excluded from reports by using SiteCatalyst's IP Filtering features. In some cases, business or analytic departments may not be fully aware of internal testing procedures in place that may be included in reporting. Make certain to verify with QA, design, and development departments how these types of testing are excluded from reporting and analytics.
Technology Differences	In some cases, SiteCatalyst may not measure a small percentage of pages that are requested from servers because they do not fully load in the browser and therefore do not generate the image request that transmits the data to Omniture. To minimize, place the SiteCatalyst

	code as high on the page as possible. to measure only pages that are fully loaded in the browser, place the code near the bottom of the page (or contact your Omniture Implementation Engineer for assistance).
Pages not Implemented	A web server log file measures requests for pages, images, and content that are requested from the server. SiteCatalyst measures all pages that are implemented, and may also be measuring third-party content not included in web server logs. In some cases, pop-ups or other "minor" content pages may get missed in the implementation process. Ensure that all pages are implemented (contact your Omniture Implementation Engineer for recommendations for automated auditing services).
Inaccurate Filters	Existing analytic solutions may not be fully filtering content, or may be including some content that SiteCatalyst does not, by default, include in "pageviews," which includes graphics, Word documents, PDF files, or other types of non-HTML content.
IP Exclusion	Ensure that IP exclusion filters between the two systems are consistent. Exclusions within one system only could account for significant differences.
VISTA Rules	If you have requested the implementation of VISTA rules within SiteCatalyst, traffic may be segmented or reported differently than in existing systems. Document and communicate all VISTA rules that may affect how traffic is reported. Additionally, you may also need to document the dates that VISTA rules were put into place for trending and comparison purposes.
Firewall Settings	Ensure that firewalls are not excluding traffic to the SiteCatalyst servers, especially if the site is primarily internal or an Intranet.
Site Changes	If detected differences have increased, use the trending feature in SiteCatalyst on affected pages or variables in order to determine the date of change. Compare to site changes made during the same period, and determine if the SiteCatalyst implementation has been affected (removed tagging, caused a JavaScript error, deleted the .JS file, etc.). The Deployment Validation white paper and checklist can be used to re-verify the implementation.

1.3 Comparison Strategies

Accounting for all differences between systems can be very time-consuming because the raw data is difficult to compare since there is so much of it. Detailed comparisons cannot always fully determine the exact differences. One reason is because a single page request can appear differently between SiteCatalyst and other systems, including the IP address. There is no automated tool to compare log files with data collected by SiteCatalyst, making the comparison a difficult and tedious process. In many cases, educating executives and analytic users about the expected differences may be sufficient for transitioning to SiteCatalyst. In other cases, limited audits may accomplish the objectives of understanding the differences in metrics. For detailed comparisons, limited data sets facilitate the ability of compare data from SiteCatalyst and other systems.

1.3.1 Overall Accuracy

SiteCatalyst is extremely accurate, especially compared to server-based solutions, which might be including as much as 20-50% automated traffic (which is misleading for site analysis and reporting), and may be missing as much as another 15-20% of cached pageviews. SiteCatalyst measures between 96-99% of all data when implemented following standard best practices, and when the code is placed near the top of the HTML page. For pages where even greater accuracy is desired, especially SSL (secure) pages in the checkout or conversion processes, a server-generated image request can generate data accuracy as high as 99.5% or greater, based on comparisons with existing Omniture clients.

While pageviews may be measured by the different systems in many different ways, transactional metrics such as "registrations" or "orders" are often easier to compare between systems than "pageviews". It is often helpful to focus on validating these metrics first. Generally, these metrics should be between 97-99% accurate if the same time zones are compared, the code is near the top of the HTML page, and the code is correctly implemented.

1.3.2 Communication

Using the tools and resources referenced in this document, and by communicating the differences that apply to your company and situation, users can have confidence in using SiteCatalyst metrics. Note that in most cases your Implementation Engineer will need your assistance in identifying internal situations and configurations that impact the metric reporting solution currently deployed. Verify the testing and monitoring that is occurring internally - these can greatly impact accuracy.

1.4 Technical Review of Differences

There are several differences between using log files and SiteCatalyst to capture and track web analysis data. The following sections outline the differences between using Log Files and using SiteCatalyst.

The following table describes the data that SiteCatalyst receives, but server-based solutions do not.

Browser Caching	Many web browsers will only make a request from the server if a page is not already stored in the disk cache within the browser. This reduces traffic, but a request may not be made to the server for a viewed page. The web logs will not record the page view.
Proxy Server Caching	ISPs and other institutions with a large user base use proxy servers to reduce network traffic. Pages that are requested and loaded into a browser, by way of the proxy server, are stored in the proxy server disk cache. Once the page is in the proxy server cache, the proxy server rather than the web server will serve it. The web logs will not show the page view in this case.

The following table describes the data that log files receive, but SiteCatalyst does not.

Site Monitoring Services	<p>Many companies pay for a monitoring service of their web site to measure performance and up time. Companies might also monitor web sites in an attempt to solicit their services if the site goes down, or if the response is slow. Because the monitor makes a request, but never renders a page, an entry will show on the web server logs, but not on client-side logging.</p> <p>Sample web log entry:</p> <pre>23.251.135.162 - - [23/Jun/2003:10:34:31 -0400] "GET / HTTP/1.1" 200 62077 "-" "GomezAgent2.0" "-"</pre>
--------------------------	---

<p>Crawlers</p>	<p>There are a variety of crawlers that will "spider" a site and create web log entries. Many of these are filtered by popular log analyzer products. New crawlers being created all the time, however, and require constant updating of filters. Many are homegrown and others are disguised to appear like a browser to harvest emails. SiteCatalyst's data collection technology eliminates the spiders' traffic because spiders and bots normally do not execute JavaScript nor make image requests.</p> <p>Sample web log entry:</p> <pre>80.213.1.115 - - [23/Jun/2003:10:24:00 -0400] "GET / HTTP/1.1" 200 62077 "-" "Mozilla/4.0 (compatible; grub-client-1.3.7; Crawl your own stuff with http://grub.org)" "-"</pre>
<p>Link Checkers</p>	<p>There are services that check for broken links from other web pages. A request is made by another server to see if the link is still valid. Also, many companies check links on sites to</p> <p>Sample Log Entry:</p> <pre>216.2.193.1 - - [23/Jun/2003:10:07:40 -0400] "HEAD / HTTP/1.1" 200 0 "-" "Xenu Link Sleuth 1.2d" "-"</pre>
<p>Internal Watch Dogs</p>	<p>Many companies have internal hits from watchdogs. These will be recorded by server-based solutions, but not in SiteCatalyst because a page was never rendered. Normally these will come from an internal IP address. In the example below the IP is internal and the requests are timed 5 minutes apart.</p> <p>Sample Web Log:</p> <pre>23.215.56.155 - - [23/Jun/2003:10:00:16 -0400] "GET / HTTP/1.1" 200 62077 "-" "Mozilla/4.0 (compatible; MSIE 4.01; Windows NT)" "-" 23.215.56.155 - - [23/Jun/2003:10:05:16 -0400] "GET / HTTP/1.1" 200 62077 "-" "Mozilla/4.0 (compatible; MSIE 4.01; Windows NT)" "-" 23.215.56.155 - - [23/Jun/2003:10:10:16 -0400] "GET / HTTP/1.1" 200 62077 "-" "Mozilla/4.0 (compatible; MSIE 4.01; Windows NT)" "-" 23.215.56.155 - - [23/Jun/2003:10:15:16 -0400] "GET / HTTP/1.1" 200 62077 "-" "Mozilla/4.0 (compatible; MSIE 4.01; Windows NT)" "-" 23.215.56.155 - - [23/Jun/2003:10:20:16 -0400] "GET / HTTP/1.1" 200 62077 "-" "Mozilla/4.0 (compatible; MSIE 4.01; Windows NT)" "-"</pre>
<p>Internal Program Calls</p>	<p>Some companies have an environment that makes requests to their own server from internal programs. The user agent as seen below in the example identifies these internal programs.</p> <p>Sample Web Log:</p>

	<pre>23.215.56.199 - - [23/Jun/2003:10:14:26 -0400] "GET /cgi/recruit.cgi?user=XXXX HTTP/1.0" 200 16227 "-" "Java1.2.2" "-"</pre>
<p>User Refresh Before Page is Loaded</p>	<p>Many times users refresh the page before the page is completely loaded. This tends to happen with users whose geography is far from the geography of the web server (e.g., a European user accessing a server in the U.S). Notice the short time frame in the example below. These requests will appear in the web logs, but not in SiteCatalyst.</p> <p>Example Web Log Entry:</p> <pre>193.191.137.205 - - [23/Jun/2003:10:15:39 -0400] "GET / HTTP/1.0" 200 35759 "-" "Mozilla/4.0 (compatible; MSIE 5.01; Windows NT 5.0)" "-"</pre> <pre>193.191.137.205 - - [23/Jun/2003:10:15:40 -0400] "GET / HTTP/1.0" 200 35759 "-" "Mozilla/4.0 (compatible; MSIE 5.01; Windows NT 5.0)" "-"</pre> <pre>193.191.137.205 - - [23/Jun/2003:10:15:40 -0400] "GET / HTTP/1.0" 200 35759 "-" "Mozilla/4.0 (compatible; MSIE 5.01; Windows NT 5.0)" "-"</pre> <pre>193.191.137.205 - - [23/Jun/2003:10:15:41 -0400] "GET / HTTP/1.0" 200 35759 "-" "Mozilla/4.0 (compatible; MSIE 5.01; Windows NT 5.0)" "-"</pre>
<p>Abandoned Page</p>	<p>Frequently, a user will follow a link only to hit the back button before the page has completely rendered. Depending on the location of the SiteCatalyst code, the page may or may not be counted (i.e., code near the top of the page is counted more often for this type of action than code near the bottom of the page). If the linked-to page did not completely load in the browser before the user clicks the back button, this will show up as a reload for the page in SiteCatalyst logging (in Pathing reports).</p>



CALL 1.877.722.7088
1.801.722.0139

www.omniture.com
info@omniture.com

550 East Timpanogos Circle
Orem, Utah 84097

