



WHITE PAPER



LOG FILE VS. ASP-MODEL ANALYTICS

A Critical Comparison of Leading Web Analytics Systems

February 22, 2006

Version 1.0



1 Overview

In today's online world, an in-depth understanding of web site customers and traffic trends is critical to e-business success. During the early days of the Internet, this site traffic monitoring was based on data extracted from web log files. Software tools were developed to help analyze and translate the virtually unintelligible data from these files into reports that contained useful information. Unfortunately, this approach to web traffic analysis has had many shortcomings from the beginning, which continue to grow more severe as Internet technology advances. E-businesses are now shifting toward Omniture's intelligent ASP analysis solution, SiteCatalyst, to avoid these weaknesses and to back their initiative decisions with quality, accurate information.

This document first explains how each solution gathers and presents web traffic information. It then compares the two side-by-side to help identify specific advantages and/or drawbacks.

1.1 Log File Analysis – The Software Model

Each time a web server receives a request for a site page from a visitor, the server records that data in a log file. As more requests are made from the server more data is stored, resulting in more files occupying space on the server. Software tools have been developed to translate log files into meaningful information, but they require time-consuming batch analysis that is typically performed on the site owner's system. These reports are static and limited to the web server's logged requests. Despite these and other problems, log file analysis is still widely used today.

1.1.1 The Log File Analysis Process

The steps outlined below describe the log file analysis process.

1. Visitors' browsers make requests to the site's web server to view specific site pages. The server receives these requests and serves the specified pages back to the visitors.
2. The web server creates a log file containing the visitors' requests.
3. The web site owner uses proprietary log file analysis software to extract meaningful information from the latest logs.
4. Reports are generated from the log file analysis software, which the site owner distributes to the appropriate users within the organization.

1.2 ASP Model Analysis – The Remotely Hosted Solution

An Application Service Provider (ASP) is a third-party entity that manages and distributes services and solutions to customers across a wide area network from a central data center. In essence, ASPs allow companies to outsource portions of their information technology needs to specialized service providers.

The Omniture ASP model is implemented through code that is placed on each page to be monitored, providing real-time browser-based analysis. Whenever a person views a page on the site, the tracking code sends information about the page and the visitor to SiteCatalyst. No log files are created or used for analysis. All information is gathered and stored immediately so you can see your audience's behavior as it happens, and from the visitor's perspective rather than the server's. When the site owner requests reports for the site, SiteCatalyst analyzes current site data in real time to obtain and deliver the information.

1.2.1 The ASP Model Analysis Process

The steps outlined below describe the ASP Model analysis process.

1. Web visitors' browsers make requests to the site's web server to view specific site pages. The server receives these requests and serves the specified pages back to the visitors.



2. When a page is displayed on the visitors' browsers, embedded SiteCatalyst code sends anonymous information about the visitor to SiteCatalyst, where it is immediately integrated into the site's SiteCatalyst database. Tracking is done through a series of steps:
3. The tracking code that has been included in the web page collects a stream of visitor information.
4. The collected information and any customer-defined variables are sent to the SiteCatalyst server via an image source request.
5. The SiteCatalyst server returns a 1 pixel by 1 pixel transparent GIF that is displayed on the client's web page. Hit information is usually aggregated into your reports within six seconds of the original hit request.
6. Authorized users log on to the SiteCatalyst site and request audience information. SiteCatalyst extracts the requested information from the database and serves it back to the requesting users.

2 Case by Case Comparison

SiteCatalyst is revolutionizing the web analytics industry by tracking web site traffic more accurately and efficiently than ever before. By comparing log file solutions to SiteCatalyst's ASP model, you can see for yourself why SiteCatalyst should be your product of choice. The following pages compare log file and ASP-model solutions by features within three basic categories: Features and Installation, Data Collection and Accuracy, and Reporting and Support. This comparison will help you better see why SiteCatalyst is more effective and efficient than log file software solutions.

2.1 Features and Installation

LOG FILES	SITECATALYST TECHNOLOGY
IT Resource Requirements	
<p>Log file analysis is typically performed using proprietary software running on the site owner's system. Additional software may also be required to view the resulting reports. Log files for a larger site can be quite sizeable – into the gigabyte range – for a single day. Companies using these solutions can face significant costs and operational burdens to:</p> <ul style="list-style-type: none"> > Pay the web hosting provider to generate log files (many providers do not normally provide log files) > Acquire the CPU and storage capacity to manage log files and run software > LICENSE, INSTALL AND UPGRADE ANALYSIS AND VIEWING SOFTWARE > Employ and train staff to manage log files and run analysis and viewing software 	<p>SITECATALYST REQUIRES NO ADDITIONAL HARDWARE, SOFTWARE OR IT STAFFING. OMNITURE HANDLES ALL DATA PROCESSING AND STORAGE. THE ONLY TOOLS THE COMPANY'S USERS REQUIRE ARE THE WEB BROWSERS THAT ARE ALREADY INSTALLED ON THEIR WORKSTATIONS.</p>
Implementation Requirements	
<p>Log file analysis software (and possibly viewing software) needs to be installed and configured on the site owner's system. Depending on the size of the web site and the location of the web server, the installation process can be very complex. IT staff and users also need to be trained to operate and use the software.</p>	<p>Site owners only need to insert a small snippet of SiteCatalyst code on each page to be tracked. Implementation typically can be completed in an hour or two, and with minimal impact on normal operations. Once this is done, people throughout the company can start accessing the information immediately.</p>
Maintenance Requirements	
<p>Vendors of log file-analysis software generally provide some level of support for their customers. However, to use a log file system effectively, the site owner must acquire and maintain the personnel needed to operate, maintain and</p>	<p>Aside from the simple task of adding SiteCatalyst code to the site's pages, any user familiar with a standard Web browser can easily perform virtually all SiteCatalyst tasks.</p>

LOG FILES	SITECATALYST TECHNOLOGY
-----------	-------------------------

troubleshoot the entire system.

Scalability

Log file analysis systems are typically offered at a few preset levels of capacity, constraining the site owner in terms of the traffic volume and the number of users that the system can handle. Upgrading the system to increase its capacity can be costly, and the site owner often ends up paying for too much capacity.

SiteCatalyst' service level can be easily scaled to the site owner's needs, freeing the owner from the extra cost of upgrades or unused capacity. This also enables the service to grow smoothly with the owner's business. SiteCatalyst' remote hosting provides the server and support infrastructure needed to keep up with and track high-traffic site and rapidly growing sites.

Security

Although most web site owners attempt to maintain high levels of security and availability for their information systems, the resources required to ensure near-perfect security and availability are beyond the physical and financial means of many site owners, adding significant risk to an in-house log file analysis system.

In order to deliver the most secure service, we choose only those hosting services that meet all of the following criteria:

- > Highest possible uptime through the use of substantial redundancy and parallel data connections
- > Electronic motion sensors and continuous video surveillance
- > Biometric access and exit sensors
- > On-site security officers and security breach alarms
- > Server operations monitoring
- > Gas-based fire suppression systems
- > Seismically-braced server racks (reliability/availability)
- > UPS backup generators (reliability/availability)
- > Redundant HVAC controlled environment (reliability/availability)

Reliability

When a site owner experiences a problem with system availability, other IS functions such as sales, accounting and MIS systems may compete with a log file-analysis system for limited resources. If the servers go down, log files do not track at all.

All information is collected, processed, delivered and stored through SiteCatalyst' reliable and safe infrastructure – giving the site owner extremely high levels of security and availability, and reducing contention when resources are limited.

2.2 Data Collection and Accuracy

LOG FILES	SITECATALYST TECHNOLOGY
-----------	-------------------------

User Identification

Log files track unique visitors by their unique IP

SiteCatalyst does not depend solely on IP

LOG FILES	SITECATALYST TECHNOLOGY
<p>addresses. However, this can be tremendously misleading. For example, many ISPs have a pool of Internet Protocol (IP) addresses that are dynamically assigned to individual people. In this situation, a single browser may use multiple IP addresses over time – even during a single visit to a site – meaning each IP address will not necessarily reflect a single and truly "unique" visitor. As a result, counts of unique visitors and measurements of how long people spend on a site and on individual pages may be grossly inaccurate.</p>	<p>addresses to identify individual visitors. It instead sets a cookie that tracks unique visitors accurately and acts as an identifier for each session and return visit. SiteCatalyst only uses the IP address for identification if the browsers' cookies are not enabled (usually about 2-4% of all site visitors).</p>

Page View Accuracy

<p>Web surfers following familiar paths often jump between pages very quickly without viewing page content. Log file analysis cannot detect this situation and counts each of these jumps as a page view. Because the visitor does not view the entire page, this potentially results in significant page view over counts. Though log file analysis can identify this situation by measuring time spent on a page, log files do not record page views from visitors that use the "back" button or the refresh command, because server requests are not made.</p>	<p>SiteCatalyst provides a unique alternative solution to this problem: The site owner inserts the SiteCatalyst code at the end of the HTML for a page, or following key content. If a visitor leaves the page too quickly, the SiteCatalyst code will not be loaded and SiteCatalyst will not record a page view. This technique makes it possible to obtain more realistic page-view counts without having to analyze the time spent on pages.</p>
---	--

Dynamic Page Tracking

<p>Web page content is often generated dynamically. A common example is a search results page, whose content is generated on the fly in response to the visitor's search. Log file analysis has difficulty identifying both this type of page and its dynamically generated content.</p>	<p>SiteCatalyst allows site owners to give dynamic pages unique identifiers (names) for the SiteCatalyst reports. By doing this, SiteCatalyst reports can provide accurate statistics for dynamic content.</p>
--	--

Frame Tracking

<p>Frames make use of multiple, independently controlled sections within a Web site. This effect is achieved by building each section as a separate HTML file with one "master" HTML file that identifies each one. When a visitor requests a Web page that uses frames, the address requested is actually that of the "master" file that defines the frames. One HTML file is then returned for each visual section. Typical log files count one request for each page returned, including the "master"</p>	<p>SiteCatalyst is included on just one page's HTML code (usually a frame page containing important content, rather than mere navigation), thus circumventing the log file analysis problem. One complete page view always results in one count.</p>
--	--

LOG FILES	SITECATALYST TECHNOLOGY
-----------	-------------------------

page, resulting in multiple counts for one complete page view.

“Spider” Tracking

A spider is a program that visits and “reads” web site pages and other information in order to create entries for a search engine index. Log files cannot decipher between spiders and actual visitors because they both request pages in the same way. Log file analysis tools can be programmed to detect some spiders, but since thousands of spiders appear every day, these tools cannot detect every one. Spider hits often result in a substantial page view over count, drastically skewing site traffic statistics.

SiteCatalyst code generates a transparent image that spiders do not load. As a result, SiteCatalyst automatically excludes spiders because they are not actual visitors.

Cache Server Limitations

Many ISPs maintain proxy servers that store millions of pages copied from the web. When visitors request these saved pages, the ISPs deliver them from the proxy instead of the web server to increase the speed with which they are delivered and to bypass server downtime. Log file analysis has absolutely no way of seeing hits from proxy servers, resulting in significant undercounting of page views. For example, if a page on your site had ten million and one visitors, the first page view would be recorded while the other ten million would be served from the proxy server. All of those visits would be logged as just one page view.

SiteCatalyst uses “Cache-Busting” technology to detect all displayed pages, regardless of the source. Because all data is gathered directly from browsers and is analyzed by SiteCatalyst, every hit and page view is tracked accurately, regardless of caching.

2.3 Reporting and Support

LOG FILES	SITECATALYST TECHNOLOGY
-----------	-------------------------

Resource Requirements/Performance

Building log files requires a significant amount of CPU time and data processing on the web server. In addition, log file analysis done on the web server can consume a great deal of the server’s CPU time. This consumption of system resources can degrade site performance on a busy Web server.

SiteCatalyst requires no client resources. Since all data processing is performed by SiteCatalyst, the Web server’s entire capacity can be used to operate the site at peak performance, freeing up existing servers and bandwidth and avoiding upgrades and additional server purchases.

Data Collection and Reporting

Though small log files can be analyzed fairly

SiteCatalyst collects, analyzes and delivers

quickly, large log files are more common and require off-hour batch-process analysis, resulting in delays of a day or more.

information in real time, enabling site owners to observe traffic and visitor activity on the site as it happens.

Report Access

Web servers only create log files from the web page that is located on that specific server. Site owners may need to pull log files from several servers to get aggregate page view counts. If they don't pull all the log files from all servers, they will not get an accurate count. Although log file analysis reports can be made available throughout an organization on a network, intranet or web page, this information is typically static and at least one day old.

SiteCatalyst tracks every page containing the tracking code and delivers reports to a web browser in real time. No additional work is required for the site owner, except to log in and view the reports. And, with SiteCatalyst, authorized users throughout the company can access up-to-the-minute information anytime, over virtually any Internet connection.

Customer Service

Each SiteCatalyst customer receives his or her own Account Support Manager, daily account monitoring, report and analysis training and a dedicated enterprise server.

2.4 Summary

LOG FILES	SITECATALYST TECHNOLOGY
Examines site visitors' activities from the distant perspective of the Web server.	Examines site visitors' activities from the close perspective of visitors' browsers.
Limited by the information contained in visitors' Web server requests.	Provides additional information available from visitors' browsers.
Requires time-consuming batch analysis to extract useful information.	Provides information on demand in real time.
Typically performed on the site owner's system using proprietary software tools.	All-in-one service requiring no IT resources.
Delivers information through static reports distributed by the site owner.	Makes information available anytime, through almost any Internet connection.

2.5 Conclusion

Your e-business success depends on an accurate knowledge of your web site customers and traffic trends. You need a web analytics solution that will deliver the best information possible and with minimal impact on your business and staff. The side-by-side comparisons in this document reveal SiteCatalyst's clear advantage over traditional log file systems. Comprehensive real-time reporting, industry-leading reliability and incredible ease-of-use provide the best information with minimal effort. SiteCatalyst truly is the clear choice for people who want the most accurate, reliable web traffic analysis data.



CALL 1.877.722.7088
1.801.722.0139

www.omniture.com
info@omniture.com

550 East Timpanogos Circle
Orem, Utah 84097

