

Web Analytics

WHAT IS WEB ANALYTICS?

Web analytics is the process of analyzing the behavior of visitors at a web site. The use of web analytics helps organizations maximize the value of their Internet marketing efforts. By understanding visitor behavior, organizations can tailor their marketing efforts to attract, retain, and grow the value of customers.

This whitepaper discusses in detail the benefits of analyzing web traffic, methods for capturing web data, and the pros and cons of various technology solutions available for web analytics.

WHY ANALYZE WEB SITE TRAFFIC?

There are many reasons why your organization might want to analyze your web site's traffic. First and foremost, you will want to know whether your web site is attracting visitors and whether the investments that you are making in it, both in time and in money, are paying off.

At the next level, you will want to know more about trends on the web site. Marketing will want to understand where visitors came from: whether they were referred by a specific search engine, responded to an ad campaign, or used a particular keyword (either pay-per-click or organic). Content developers will want to know which pages and sections are most commonly viewed and how long visitors stay. Your webmaster and IT department will want to see how servers are performing and whether or not visitors are receiving error messages when they access various pages.

Additionally, you will want to learn as much as possible about individual visitors, including their names, companies, email addresses, telephone numbers, and geographic location so you can leverage this information for targeted content and marketing activities. You will also want to see what each of these individuals did on your web site, such as whether they filled out online forms, made purchases, or if they fall into any other visitor behavior segments that have been defined for your web site.

Finally, you will want to answer the specific, unique questions that you have about your web site traffic. Perhaps you want to track whether your visitors are taking advantage of a specific new web site promotion you are running. Maybe you have recently updated your checkout process and you want to see whether this is causing fewer visitors to drop out on making a purchase. With web analytics, you can answer all your tough web questions.

HOW CAN WEB SITE TRAFFIC INFORMATION BE COLLECTED?

There are several methods for collecting web data and accurately identifying visitors, including web server log files, page tags, cookies, and hybrid approaches.

Web Server Log Files

When you visit a web site, you actually connect with a specific web server on the Internet that serves you the files that make up that page (HTML files, image files, etc.). The web server creates a log of each file you request, when they were requested, your IP address, and your referring page.

With detailed logging, it doesn't take long for these files to grow very large, even for modest business-to-business web sites. It is impossible to extract any useful information by directly studying these records.

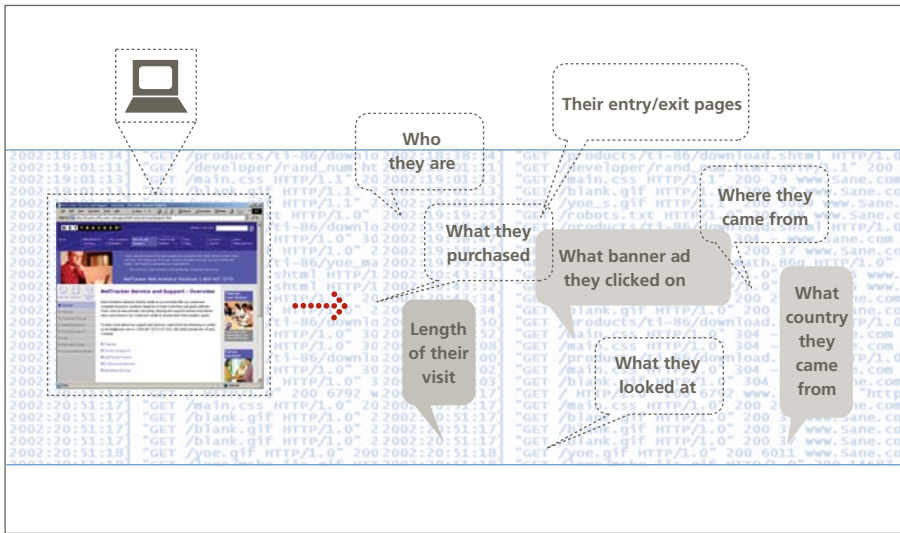


Figure 1: Log Files

This is where log file analysis solutions come in. These solutions take raw log file data and convert it to meaningful information about your web site traffic. Since these server logs can be configured to contain information on the page requested, the previous page visited (including referrers and keywords), date and time of the visit, IP address, and browser and platform types for your visitors, it is possible to use them to get a clear picture of your visitors and their actions on your site.

Of course, there are many different types of log files, each providing essentially the same information, just in different formats. That's because each web server vendor (Netscape, Microsoft, etc.) uses its own logging scheme.

Fortunately, most log file analyzers can be configured to analyze these different log file formats.

Page Tags

Page tagging consists of embedding a special file on each web page for the sole purpose of tracking the visitors who requested the page. A JavaScript "tag" is embedded in the source code for each page of your web site you wish to track. This tag is composed of an invisible, zero-pixel image hosted on a server which logs requests. When a page is viewed and a request is made to your page-tagging server for the image, the request contains a query string, which includes the page being requested, the referring page, the screen resolution of the visitor, and other information you may be tracking. This information can then be analyzed and used to create web traffic reports.

Additionally, you can use this approach to track things other than page views. "Event tags" can be placed on the web site to track on-page activities, like products removed from shopping carts before checkout, fields left blank on an online form, and how far into a particular page a visitor has scrolled. Page tagging can also gather information not found in log files, like screen resolution.

Hybrid-Page Tags and Log Files

Page tagging and log files each have advantages and disadvantages. Therefore, a particularly interesting and powerful option is to combine the benefits of both for a hybrid approach to data collection. For example, you may decide to have your main data collection in log files. This data can then be augmented by adding event/page tags, for example, to discover how far visitors scroll down a particular page. Other examples of this approach include the ability to tag pages that are hosted on other web sites where log files may not be available. All in all the result is more detailed and accurate data capture.

Cookies

A cookie is a small piece of information generated by the web server and stored on the client machine. Advanced web site analysis solutions offer the option to use cookies to get much more accurate visitor counts while also allowing analysis of repeat visitors. Here's how it works: When a web browser requests a page or page tag from a web server, it speaks hypertext transfer protocol (HTTP). Below is an example of an HTTP request:

```
GET /products/NetTracker/index.html HTTP/1.0
User-Agent: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)
```

In the sample above, the visitor requested the page /products/NetTracker/index.html using the HTTP 1.0 protocol. The visitor also informed the web server that they are using Microsoft Internet Explorer 6.0 as a web browser. After processing the request, the web server will send back a response containing the web page. An example response is below:

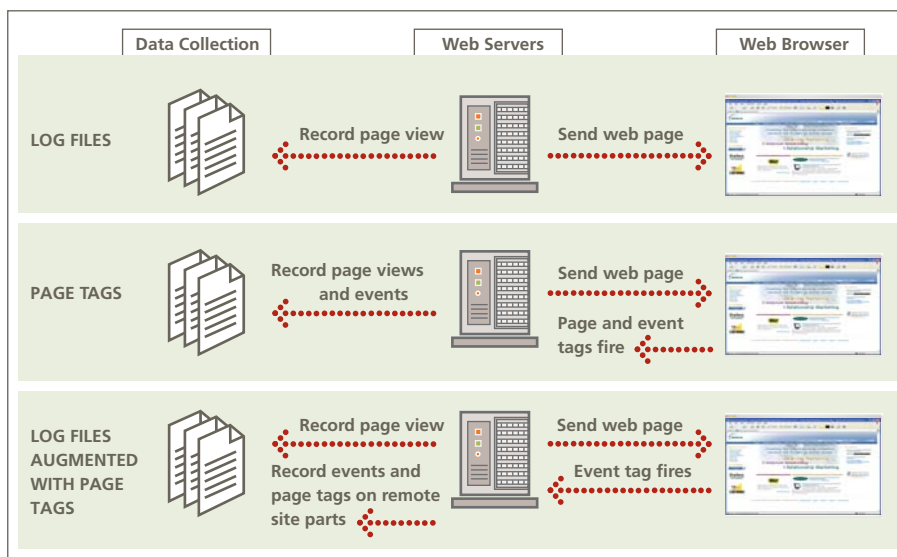
```
HTTP/1.0 200 OK
Date: Tue, 31 Aug 2006 13:31:25 GMT
Server: Microsoft-IIS/6.0
Set-Cookie: SaneID=208.144.251.1-936106299700; path=/; expires=Tue, 31-Aug-09 12:00:00 GMT
Connection: close
Content-Type: text/html

<HTML>
<HEAD>
<TITLE>NetTracker web Analytics and Usage Tracking Software</TITLE>
</HTML>
```

The response is divided into two sections separated by a blank line: the HTTP headers, which provide meta-information about the document, and the document itself in HTML. The bold line in the HTTP headers above issues a cookie to this browser, thus tagging it as a unique visitor. When the browser requests another page or page tag from the same server, it will send the cookie back to the web server, thus identifying itself:

```
GET /products/NetTracker/index.html HTTP/1.0
User-Agent: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)
Cookie: SaneID=208.144.251.1-936106299700
```

Advanced web site analysis solutions will capture this cookie information and use it to accurately calculate unique visitors. Also, you will notice that in our first example, the browser did not send a cookie to the web server while for subsequent requests it did. By analyzing requests where the browser did not have a cookie for the first request, software can determine when a visitor comes to the site for the first time. This enables new visitors to be compared to repeat visitors.



The example to the left illustrates the procedure when a web site visitor requests a page. A page tag residing on this server will also receive the cookie information with each subsequent request. If the page tag resides on a separate server, a web server plug-in can be employed to set and receive cookies for each file being accessed from the server, including page tags.

Figure 2: How Web Analytics Data is Collected

THE PROS AND CONS OF DATA COLLECTION

Table 1 below summarizes the pros and cons of the various data collection methods discussed in this whitepaper: log files, page tags, and cookies.

TABLE 1: Data Collection	PROS	CONS
Log files	<ul style="list-style-type: none"> · Automatically generated/minimal set-up · Ability to archive and leverage historical data · Detailed data is collected related to errors and the behavior of robots and spiders that crawl a site 	<ul style="list-style-type: none"> · By default, extraneous data is often captured (e.g. images and pages that do not need to be analyzed) · Cached pages are not captured in log file data · Tracks a form in its entirety—not the individual elements
Page tags	<ul style="list-style-type: none"> · Script must be embedded on web site pages, providing greater control on what is and isn't tracked · Cached pages are addressed in tracking · Event-based tagging (e.g., tracking individual items within a form) 	<ul style="list-style-type: none"> · JavaScript must be enabled on browsers for tags to fire and capture information · Tracking is from point of implementation forward—no historical data · Unable to track images or files downloaded
Hybrid – page tags and log files	<ul style="list-style-type: none"> · Ability to test and compare approaches to optimize data capture · Enables analysis of site sections 	<ul style="list-style-type: none"> · Lots of data to review and analyze · Visitors can control via their web browser whether or not a cookie is used to track their activity on a specific site
Cookies	<ul style="list-style-type: none"> · Approach to tracking repeat activity for a specific visitor · May be used regardless of the data collection method (log files, page tags, or hybrid) 	<ul style="list-style-type: none"> · Additional security blocking third party cookies (i.e., cookies that are accepted on behalf of a company by a third party)

HOW DOES WEB SITE TRAFFIC MEASUREMENT SOFTWARE WORK?

Web site tracking typically consists of four stages: defining goals, gathering data, sessionization, and reporting.

Defining Goals

On a web site more can be measured than what is useful to the business. Therefore, the first step of web site traffic measurement is to define the goals of a site in terms of key performance indicators. Leveraging this focused approach, what data to gather is easily determined and prioritized, and the reports needed are identified.

Gathering Data

The data gathering stage differs from solution to solution, depending on how the data is collected. For log file analysis solutions, this stage consists of parsing the log files into a common format at scheduled intervals. For page tagging solutions, this stage consists of monitoring which tags are sent across the network. Another method of collecting data is through server plug-ins, modules written to extend the web server's functionality which take the data in real time, before it is written to the log file, and pass it along to the analysis stage.

Sessionization

Once the data is gathered from log files, page tags, or server plug-ins, it is sessionized. During the data sessionization stage, complex heuristics are applied to the raw data to categorize it into page views, visits, unique visitors, and business events. This analysis may be based on the IP address of the visitor or, for more accurate calculations, based on cookie data or registered user names. Once the data is sessionized, it is typically placed in a database to facilitate reporting.

Reporting

The final stage of web site tracking is generating reports from the analyzed data. During this stage, the analyzed visit and page view data is read from the database and tabulated into many different reports, each answering a specific question. For example, a report on what people have viewed on a web site might consist of the top ten pages viewed, how many times each page has been viewed, and how long, on average, visitors spent looking at each page. Graphs and charts are also generated during this stage.

WEB SITE TRAFFIC MEASUREMENT SOLUTIONS

There are various solutions companies may use to measure web site traffic, including in-house software, shareware, commercial software, or service providers.

In-house Software

In the early days of the Internet, companies had to write their own software to analyze web site traffic. Even today, some companies still write their own software to meet specific requirements.

This software, often written in a text manipulation language called PERL, tends to be costly and time consuming to produce and support. And it does not have the general functionality of commercial software, nor does it conform to industry standards.

Shareware

Shareware is free or very inexpensive software. Since shareware solutions tend to have limited features, companies tend to need a more complete or sophisticated solution as their reporting needs become more detailed.

Commercial Software (On premise)

With its abundance of features, commercial web site traffic analysis software is often the next logical step for online marketers after shareware. Most commercial software products offer advanced user interfaces that include graphs, calendars, and visual reports, such as funnel and path reports. Most have online help and provide technical support via phone and email. There are two categories of commercial software: desktop and server.

Desktop solutions are relatively inexpensive and can be run on a personal computer. These solutions almost always utilize web server log files for analysis. Users must download their log files from their web server to their desktop to analyze them. So, depending on the size of your log files and the speed of your Internet connection, it may take a while to download and process your log files to create reports.

Server software solutions require more sophisticated hardware, specifically, a web server. The advantage is that you do not have to download log files or page tag data, and you are able to process web traffic data much faster. In addition, since the web server is most often connected to the Internet, you can make your web site traffic reports available to anyone within your organization with access to the Internet.

Service Providers (Hosted)

A web analytics service allows organizations who may not have the software, hardware, or IT staff necessary to implement commercial software in-house to obtain web site analysis with minimal upfront costs.

The disadvantage of a web analytics service is that you do not have the same level of control that you do if your company is maintaining the setup. Also, depending on the service vendor and the web data collection methodology used, it may be difficult or impossible to integrate the web site data with data from other sources. Some web analytics service vendors offer more functionality than others; however, there are often additional fees for this functionality.

When selecting a web analytics service provider, it is important to make sure that the vendor has a reputation for data security. You also want to ensure that the vendor is collecting and storing your data in such a way that, if you ever sever your relationship with the vendor, you will still have access to your historical data.

THE PROS AND CONS OF HOSTED VERSUS ON PREMISE SOFTWARE

As discussed in the section above, there are various solutions companies may use to measure web site traffic. These solutions may be used internally (on premise) or externally (hosted). Table 2 below summarizes the pros and cons of each of these approaches.

TABLE 2: Hosted vs. On premise Solutions	PROS	CONS
Hosted	<ul style="list-style-type: none"> · Leverage a powerful infrastructure and web analytics expertise · Capabilities are deployed quickly and easily · Focus on using, not creating, detailed Web site analysis reports · Shortened software upgrade cycles · Start-up costs are minimal · No hardware has to be purchased 	<ul style="list-style-type: none"> · Many offerings do not provide the flexibility to move to an on premise solution without starting over · Additional complexity when integrating web site data with data from other sources
On premise	<ul style="list-style-type: none"> · Web analytics expertise is in-house · Complete control over what is analyzed when and how · Custom reports may be created and distributed at any time · Potentially sensitive web traffic data is kept internally 	<ul style="list-style-type: none"> · Start-up costs · Hardware may need to be purchased · IT involvement may add delays in tracking and analysis

Ultimately the decision as to whether or not a company should host or use an on premise solution depends on their specific business strategy, technical infrastructure, resources, and expertise. After evaluating these criteria, companies are better able to make a decision. If a hosted solution is selected, it is important to make sure that the solution is compatible with an on premise solution, so that if your needs change at any point, you can seamlessly migrate your setup in-house without losing any customized key performance indicators, reports, and data you may have.

INTEGRATING WEB ANALYTICS WITH OTHER OFFLINE DATA

Today many organizations are beginning to integrate their web marketing activities with offline marketing activities to create a full picture of how various customer touches are influencing sales, customer loyalty, and the overall customer experience. Linking web data to the rest of the organization can be a daunting task, but it empowers you to act on the insight you gain. How much of your “offline” business is influenced by the web and your pay-per-click advertising? How many of the callers to your service line tried unsuccessfully to serve themselves on your web site before calling you? What is the impact of your catalog or direct marketing on web traffic and conversion?

Capturing your web data with a web analytics application is the first step in this process. The next step is integrating with a complete enterprise marketing management (EMM) solution so you can turn this web insight into immediate action—a follow up email to an individual that abandoned a brokerage account application, or a real-time call to an individual inquiring about a new product, furthering your cross-channel marketing success. With EMM solutions, your web site and internet marketing become an integrated part of the customer dialog and allow you to effectively measure your cross-channel marketing success.

SUMMARY

Analyzing the behavior of visitors to a web site is a key step in optimizing web site and Internet marketing efforts—and one component of a cross-channel marketing strategy. There are many approaches to capturing web analytics—from page tags to log files to cookies—and then there is the ability to do this using on premise or hosted software. Determining what solution is best for your organization depends on a number of factors, including resources, skill sets, and the questions you want to answer.

ABOUT UNICA

Unica Corporation (NASDAQ: UNCA) is a leading global provider of Enterprise Marketing Management (EMM) software. We focus exclusively on the needs of marketing organizations. Our Affinium® software addresses the principal functions of EMM, including marketing and customer analytics, demand generation, and marketing resource management.

Affinium provides a complete platform to transform the entire marketing process, including planning, budgeting, project management, execution and measurement for brand, relationship, and internet marketing. Over 500 companies worldwide rely on Unica's solutions to manage the complexities and processes of marketing and facilitate the operations of a customer-centric business, including ABN AMRO, Bank of Montreal, Best Buy, Capital One, Choice Hotels, Cinergy, Debitel, E*TRADE, Lands' End, Medco, Monster, Orbitz, Vodafone, and Wells Fargo.

Founded in 1992, Unica® is headquartered in Waltham, Massachusetts, with additional locations in the United States, the United Kingdom, France, Belgium, Germany, India, and Singapore. For more information, visit www.unica.com.

Unica Corporation
Worldwide Headquarters
Reservoir Place North
170 Tracer Lane
Waltham, MA 02451-1379
USA

T +1.781.839.8000
F +1.781.890.0012

www.unica.com
unica@unica.com

© Unica Corporation, 2006. All rights reserved.

Unica, the Unica logo, and Affinium are registered trademarks of Unica Corporation with the U.S. Patent and Trademark Office. Affinium Campaign, Affinium Campaign Collaborate, Affinium Campaign eMessage, Affinium Campaign Interact, Affinium Campaign Optimize, Affinium Detect, Affinium Leads, Affinium Leads Referral, Affinium Leads Contact, Affinium Model, Affinium NetInsight, Affinium Plan, Affinium Mart, Universal Dynamic Interconnect, Interconnect Modules, Marketic, Powering Marketing Success, Customer Value Maximization, MAP, One-to-One Optimization, and Right-time Marketing are trademarks of Unica Corporation. All other trademarks are the property of their respective owners.

WP-WA-0605-PDF April 2006