




|

IBM Tivoli Web Site Analyzer - Why Web Analytics?



During the past decade we have seen the importance of the Internet as a mainstream communications and commerce channel. Nearly every company you can think of, large and small, has a Web site. However, eBusinesses are finding that simply being on the Web isn't enough. As eBusinesses grow, they need a detailed understanding of how their site is operating and how visitors are interacting with it generally speaking and from a detailed content perspective. The ultimate goal is for you to drive customers from browsers to buyers, and having a tool to help you evaluate this is critical for survival.

To understand why you need a transformational Web analytics tool for your eBusiness and how IBM Tivoli Web Site Analyzer can provide this for you, read on.

What is IBM Tivoli Web Site Analyzer?

IBM Tivoli Web Site Analyzer is an enterprise-level Web analytics tool that can transform random Web data into valuable e-business intelligence. It can provide a clear picture of the overall health and integrity of the e-business infrastructure supporting business outcome management.

By capturing, analyzing, storing and reporting on Web site usage, health, integrity and site content, IBM Tivoli Web Site Analyzer can provide meaningful metrics on visitor site interactions and the site's overall performance. You can leverage this insight to help optimize the site to increase customer loyalty and e-business effectiveness.

What are Web Analytics?

Web analytics are the monitoring and reporting of Web Site usage and behavior so that enterprises can better understand the complex interactions between Web visitor actions and what the Web site offers, as well as leverage that insight to help optimize the site to increase customer loyalty, sales and performance. Typically thought of for B2C sites, web analytics are equally applicable for B2B and B2E sites.

How effective is the Web as a channel for my ebusiness?

The best way to answer this is through the use of web analytics. The results can help determine how good your site is at attracting and retaining visitors. In a similar light, Web analytics show the most popular paths through your site, where visitors originated from, where they go, and what they do. To some looking at the overall user and visitor traffic and traffic trends is important. To others, it's how effective are my web campaigns and initiatives. They are trying to determine the level of resources and investment required to produce the desired outcome.

What web data does Tivoli Web Site Analyzer actually capture?

Tivoli Web Site Analyzer collects and reports on the user and visitor activity on a web site. This activity is commonly called the clickstream, as it is the record of viewers clicking on links on pages presented by the web browser.

A Web Browser and a Web Server communicate using the HTTP protocol. When a viewer clicks on a link, the browser sends a request to the server for the page pointed to by that link. Once it receives the page, the browser will then parse it and, if the page contains embedded images or other presentable objects, will send requests for these as well. For each request, the Web Server knows the following (input):

- The **address** of the requestor. This is the TCP/IP address of the workstation that our viewer is sitting at.
- The **id** of the requestor as reported by the identified/auth services. In reality, this information is quickly lost in transit, thus forcing most web sites to implement other ways of recording the identity of a viewer.

- The **timestamp** of the request. Usually some variation of [yyyy:mm:dd:hh:mm:ss dddd] where dddd is the offset from GMT time
- The **type** of request. The overwhelming majority of requests are either GETs, when a browser requests something from the server, and POSTs when a browser sends the server forms data
- The **resource** requested. The object requested by the browser, whether it be an HTML page or an image or a downloadable file. This is typically a URL.
- The **Return Code**. A code indicating whether the server was successful in fulfilling the request. The HTTP protocol has a well-defined set of possible codes, with 200 indicating that the server successfully processed the browser request
- The **amount of data** sent by the server in response of the request in bytes.
- The **Referrer**. This is the web page, which had the link that the viewer clicked on to generate the current request. This is typically a URL.
- The **User Agent**. This is the signature of the browser. It uniquely identifies the web browser used by the viewer.

A Key/Value pair is a string of the form "Key=Value". Two pieces of information in the request (URL string) come in this format:

- The **query string** of either the referrer or the resource. The referrer or resource URL might be "/servlet/search?query=abc&country=us". The query string is the part following the "?" character. In this case, it consists of two key/value pairs, "query=abc" and "country=us". They typically represent parameters being passed along to a program to be run on the server.
- The **cookie string**. "Cookies" are information exchanged privately between the web browser and the Web Server. Since the HTTP protocol is "stateless", i.e. each request is independent from the previous ones; cookies are used to store information that the web server wants to remember from one request to another from the same viewer.


Web Servers usually save the request information by logging it to a file. Alternatively, schemes have been devised to directly forward the information to measurements subsystems.

As it processes the request information, TWSA keeps the following measures of activity:

- The number of **hits**. A hit is a request to the web server. Simply summing up the hits gives us a coarse view of site activity.
- The number of **page views**. A page view is a hit on an actual page instead of, e.g. a request for a GIF image.
- The amount of **bytes transferred**.
- The number of **sessions**. A session is the activity of one viewer during one visit. A session is thus the collection of all the hits that occurred during a visit. If a web site implements a session cookie, grouping the hits is simply a matter of examining the cookie string of each hit to see which ones have the same session cookie. Otherwise, heuristic rules are used to determine what constitutes a session.
- The **session duration**. The elapsed time from the first hit to the next to last hit in a session. The time a viewer looks at the last hit is not known therefore, the last hit is not included in the calculation of session duration.

Why is multi-channel data collection important?

In order to get a broad picture of what is actually occurring on your web site(s), various types of data and data sources must be tapped. Some sources provide better information than others, depending on the question(s) that need to be answered. Examples would be:



Web logs – every web server produces a log. These logs come in three common formats - NCSA Combined Log Format, W3C Extended Log Format (MS-IIS 4.x, 5.x and 6.x), NCSA Separate (3-log Format: Access Agent Referrer)

Web browser – Capturing data directly from the web site by instrumenting specific pages with JavaScript. The data is captured every time the page is requested (accessed). It is favorable for high transaction web sites that do not want to process large web logs. It is also favorable for sites that need to capture near real-time data on web campaigns or capture form data on the fly.

Database tables – data from external databases can be used to supplement web log and page-level data. Typically demographic, order and commerce data are imported.

By analyzing these aspects of your ebusiness site(s) you can obtain a picture of what is actually happening.

What is web abandonment? Can this be determined? Tracked?

Abandonment refers to an event where a visitor leaves your site unexpectedly without going through a normal exit path. It can be indicative of many things such as, slow page download or poor site performance, poor page content, difficult site navigation and so on. Knowing where someone abandoned your site can help you pinpoint the cause and prevent further abandonments. However, it is difficult to measure abandonment. The concept of abandonment is foreign to web logs. They have no way of telling that a web page has been abandoned. One of the best ways to measure this is to instrument the web page(s) with JavaScript so that if the On Completion event never occurs (i.e. the page is never loaded) a record can be written to a data mart (like Tivoli Web Site Analyzer's) noting the event and specific data associated with that event.

What are the pros and cons between using batch web logs as input for analytic processing vs. dynamic page-level data?

It's really a function of the questions that need to be answered. If only web visitor traffic and traffic patterns are sought then batch web logs processing is sufficient. Web log processing also can provide statistics on referrals, new and repeat visitors, technical specifications (browser/platform), and network demographics (IP address, domains, sub domains). However, if you need to know how successful a campaign is in the first hours of it going live or how much traffic a new ad is generating or need to capture form information (typically registration & demographic data) dynamically then instrumented web pages generally is the preferred method. In addition, most of the data available from Web logs is also available though this method. It does however require that specific web pages be instrumented or that the templates used to create these web pages be instrumented.

Can I measure Referrals to my site?

Referrals are other web sites that send (redirect) visitors to your site. This is the role of most search engines. It is important to know which referrers provide the most business to you site. Typically, these sites charge location and/or advertising fees so you might as well know whether your money is being spent wisely.

What areas of my site are generating the most attention? Which are not?

Web sites can be fairly expensive to maintain so it is important to know what area of your site(s) are most traveled (visited) and which are not. By replacing page content, relocating or reordering web pages, strategically placing campaigns and ads, inserting proper messaging and confirmation notices you should be able to achieve a higher conversion ratio that can lead to increased profitability and brand exposure (as defined by the site analysts).



What is a browser-to-buyer-ratio? What does this mean?

Browser-to-buyer ratio refers to the number of visitors that come to your site vs. the number of visitors that actually make a purchase or download information. The site determines what a 'successful' visitor outcome is and uses this to determine a 'buyer'. This metric can be very helpful in determining whether visitors are just cruising your site or actually making a purchase. It can indicate how effective your site is at attracting and retaining visitors and whether desired visitor outcomes are being achieved. Combined with site path analysis, you can get a clearer view of successful visitor activity.

What is a session? Sessionization?

Sessionization is the calculation of the number of visits to a web site. The phrase number of visits and number of sessions are typically interchangeable. Tivoli Web Site Analyzer uses a sessionization algorithm to determine the number of user sessions. This algorithm takes into consideration several pieces of information when sessionizing each hit, such as IP address, referral address, user agent (browser/platform), time stamp, etc., as well as a user customizable time out. Each piece of information is prioritized. A sessionization algorithm must make a 'best guess' as to where to 'sessionize' a hit. Tivoli Web Site Analyzer does not expose the step-by-step logic of its algorithm to the customer. Hit sessionization algorithms have been researched in the industry and several are available. In fact, due to the variety of algorithms and the heuristic nature of the sessionization problem, sessionization standards have been established in the industry. Tivoli Web Site Analyzer uses its own sessionization algorithm, which takes into account many variables for better sessionization.


It should also be noted that if the web server logs are in 3 log format (access, agent, referral), the records from each log must first be synchronized before sessionization can be done. The process of synchronization is itself not entirely accurate due to the fact that often a web server's logs will contain some invalid records, or one or more of the logs will be missing corresponding records. The synchronization process can, in effect, ultimately resolve down to a "best guess" algorithm itself.

When sessionizing a record into an existing session, Tivoli Web Site Analyzer uses a variety of criteria such as cookie values, which is its primary, preferred source. IP address, referral address, user agent (browser/platform), time stamp, etc., are also used to determine when a new visit has occurred. For example:

"-" as a referral. The "-" indicates no referral. The visitor may have come from a browser bookmark, which would indicate the start of a new session. In this example, the customer may have returned to the site rather than continuing an existing visit. The user may also have typed the URL directly into the browser entry field. Tivoli Web Site Analyzer interprets no referral as the start of a new session.

External referral - The visitor came from an external web site. An external referrer is a referral URL that is not part of the web site you are analyzing. With respect to Tivoli Web Site Analyzer, an external referral is a referral that does not match one of the "Site Host Names" configured in the Tivoli Web Site Analyzer usage analysis dialog. When a visitor enters your site from another, this is interpreted as a new session. Note that this means even if two log records have the same external referral site, they will be considered as two visits.

Session cookie or user cookie changes - Cookie assignments are under the control of the web site implementation. Web site administrators often assign cookie values based on sessions or users. Therefore Tivoli Web Site Analyzer interprets a new session or user cookie value as the beginning of a new session. Within Tivoli Web Site Analyzer, for each log file one can identify which cookie to use for visitor analysis and whether it is a user or session cookie.



User agent changes - The user agent identifies the HTTP client. Usually, the user agent specifies a browser and an operating system. Tivoli Web Site Analyzer interprets a new visit to have begun when this value changes.

Different IP address - When the IP address of the requesting HTTP client changes, Tivoli Web Site Analyzer assumes that a new session has begun.

Time on a page exceeds time specified in the usage analysis - Default is 30 minutes. Any page that is displayed longer than this time period is considered the exit page and end of the visit. This time value can be configurable by the web administrator in the Tivoli Web Site Analyzer usage analysis dialog.

All of these questions are addressed by the functionality and capabilities of IBM Tivoli Web Site Analyzer v4.2.

For more information go to : <http://www.ibm.com/software/tivoli/products/web-site-analyzer/>



© Copyright IBM Corporation 2003

IBM Corporation
Software Group
Route 100
Somers, NY 10589
U.S.A.

Printed in the United States of America
03-03
All Rights Reserved

IBM, the e-business logo, the IBM logo, Tivoli and Tivoli Enterprise Console are trademarks or registered trademarks of the International Business Machines Corporation in the United States, other countries or both.

Other company, product and service names may be the trademarks or service marks of others.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

The Tivoli home page on the Internet can be found at ibm.com/software/Tivoli/

The IBM home page on the Internet can be found at www.ibm.com

**THE INFORMATION CONTAINED IN THIS DOCUMENT IS PROVIDED "AS IS"
WITHOUT WARRANTY OF ANY KIND, INCLUDING THE IMPLIED WARRANTIES OF
MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE.**