

ClickTracks JavaScript Data Collector (JDC)

Web analytics buyers need to be aware of the way that data is fed into the analysis system, and the possible dependencies that are created between the website and the analysis engine. Different methods of collection have strengths and limitations related to accuracy, ease of maintenance, data privacy and ownership, scalability, and recurring costs.

All web analytics programs collect web visitor activity either by analyzing web log files or JavaScript page tags. Log files are generated by the web server for the site and are read directly by the web analytics program. JavaScript page tags (sometimes known as 'client-side tagging,' 'page tagging' or erroneously as 'cookies') require some lines of JavaScript to be inserted into each web page. When the page is loaded in the end user's browser, a request is sent to a data collection server, processed and made available for analysis.

When faced with the decision to use a hosted service or a log file-based web analytics solution, the buyer has been forced to compromise.

Log file based solutions come with compromises:

- Accuracy: While log file analysis keeps all data within the firewall, the technology is not as accurate as JavaScript tagging.
- Scalability: Log files can become too large for reasonable analysis while JavaScript page tagging allows high volume traffic.

The JDC is a server-based collection technology that permits customers to use our well-proven JavaScript code, and implement and host the collector on their own servers.

Hosted services also come with compromises:

- Data privacy/ownership: With hosted services, your data is stored on a 3rd party server, viewable (and sometimes used & owned) by the 3rd party, along with hundreds of other businesses' collected data.
- Recurring costs: Hosted data collection requires monthly fees that rise with visitor volume.

ClickTracks JDC provides the accuracy and scalability of JavaScript page tagging while simultaneously delivering the data privacy, cost and data-ownership benefits associated with log file-based systems. To achieve this unusual combination of benefits, ClickTracks sells a licensed, productized version of its own hosted data collection software. This gives businesses the option to use their own secure servers to implement the same software that ClickTracks uses to collect and serve web analytics to thousands of customers worldwide.

Accuracy

Log Files vs. JavaScript Tagging

Log files have been around since the beginning of the web, though they've only recently become interesting to marketing managers. Within early versions of web analytics programs, log files would typically yield data like bandwidth, server errors, peak usage, and lists of referring sites.

More complex data requires more sophisticated analysis. Modern log analyzers like ClickTracks can also determine:

1. **Visitor sessions:** These can be determined with statistically acceptable accuracy if the analysis software is able to strip graphics files, then joins distinct pages into a single visitor session, actively managing issues caused by dynamic IP addresses and session timeouts. ClickTracks uses well-established heuristic algorithms for determining visitor sessions.
2. **Session accuracy:** If a session cookie is available, session accuracy is improved. ClickTracks works with standard session cookies generated by JSP, ASP and PHP. A custom session cookie can be configured in ClickTracks Pro and ClickTracks JDC.
3. **Unique visitors:** ClickTracks Pro and ClickTracks JDC can calculate unique visitors if a unique persistent cookie is available, and the Pro/JDC server is configured (see 'cookie tracking' in the Pro/JDC Server). A database is built into Pro/JDC that can read the cookie and map back to an original campaign, even if several weeks have elapsed between click-through and purchase events. This is sometimes referred to as latent conversion tracking.
4. **Heuristics:** Tracking visitors across multiple sites/domains is usually done by falling back on heuristics when the cookie is dropped. Cookies are often not transferred when the user moves from domain to domain, so a fallback mechanism is needed.
5. **Robots and spiders:** Robots and spiders are automatically filtered and placed into appropriate reports. In ClickTracks, a robot is identified both through simple user agent checks, and also via detailed pattern recognition within the session.

Problems with Log Files

Much has been said about log files and their disadvantages. To summarize:

They aren't plug and play: Although ClickTracks can take advantage of both session cookies and persistent cookies, the fact remains that these must be present in the log file...and it's the responsibility of the website manager to set them. While there's nothing complicated about doing this, some companies simply can't gather IT resources needed.

They aren't 100% accurate: Caching of pages by ISPs and proxies can distort the data and lead to inaccuracies—this is a major differentiator promoted by vendors selling only JavaScript solutions, which suffer from caching problems less. In general, the number of cached pages has declined as the cost of maintaining the cache hardware has outpaced the cost of the bandwidth saved. Nevertheless, because of page caching and proxies, log files are somewhat inaccurate.

JavaScript Tagging

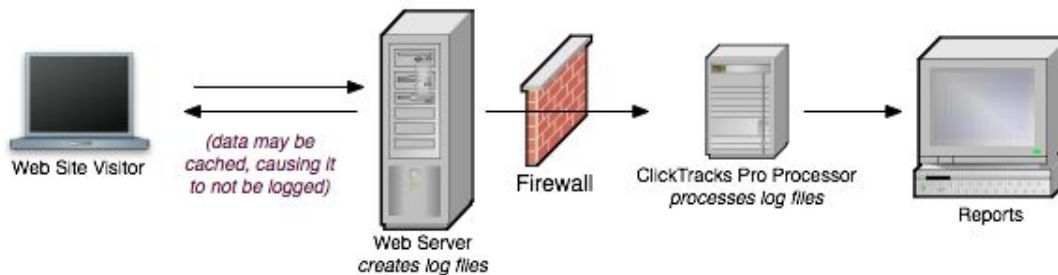
JavaScript Tagging (sometimes known as 'client-side tagging,' 'page tagging' or erroneously as 'cookies') requires code to be inserted into each page in order to be able to track it. When the page is loaded in the end user's browser, a request is sent to a server (often part of a third-party service) and the data is collected.

JavaScript gained popularity because of the ease with which one can generate reliable visitor cookie data. Since the script is able to set its own cookie, marketers that need good session and campaign data can get it, without needing the IT department to implement the necessary cookies for them.

JavaScript also excels when parsing data from the contents of the page even if the data is not available in the URL. Shopping cart total purchase value is a good example. And JavaScript-based tracking nicely sidesteps the problem of tracking across multiple domains, since the session cookie exists inside the domain where the data is gathered and not the domains of the site.

Problems with JavaScript

1. **It can't capture everything:** Some server activities, like redirects, PDF downloads, etc. are opaque.
2. **It doesn't offer technical stats:** Log file analysis is still needed for robot reports and technical stats like bandwidth/404s. You always end up needing both.
3. **It's trapped in third-party Neverland:** In almost all cases your data is trapped on a third-party service. As you grow and become frustrated with your present system, you must weigh up the problem inherent in switching.



ClickTracks Pro log file analyzer keeps all data behind the firewall

4. **It's not 100% accurate:** While more accurate than log files, JavaScript is still imperfect. For example, JavaScript errors, DNS failures and other glitches result in no data being recorded; in any of those cases, the log file would record the data without issue.
5. **It causes instability:** Pages become more unreliable as more JavaScript is added. The problem manifests as intermittent performance issues, rather than easily identified failure points.
6. **In some implementations it presents cookie issues:** The cookies issued might be 'third-party' in that they do not originate from the domain hosting the web pages. For session cookies this is acceptable, but persistent cookies require special handling through P3P and compact privacy headers.

ClickTracks JDC Combines Both JavaScript and Log Files for Peak Accuracy

ClickTracks is one of only a handful of vendors that supports both log file and JavaScript approaches to web analytics, and gives the customer total flexibility. Log files are simple, effective and inexpensive to process—and in truth, these reasons drive the majority of our customers to choose this approach.

Some customers also need JavaScript due to problems with tracking across multiple domains, or needing to parse ROI data from a complex shopping cart—or simply because they prefer the convenience.

In the situation where the customer needs the best of both worlds, ClickTracks supplies customers with the JDC (JavaScript Data Collector).

The JDC is a server-based collection technology that lets the customer use our well-proven JavaScript code, implementing and hosting the data collector on their own servers. Data from the

JDC can be freely mixed with log file data—with some sub-domains tracked by log files and some by JDC—all combined into a single view that accurately tracks users as they moves from one domain to another.

Scalability

When Log Files Get Too Big

Limits on scalability most often occur when very large log files are analyzed for high volume sites over long periods of time. On a 3.0 Mhz desktop PC with 1GB ram, the ClickTracks log file analysis processes raw server logs into more manageable files at a rate of 2000 lines per second (two to three times faster than our competitors). On sites with six million page views per day, the processing of the daily raw logs can take as much as four hours unless special hardware is applied to the process. Add the time to transfer raw logs from one machine to the other, and it becomes clear why many customers prefer the JavaScript collection method.

ClickTracks JavaScript collection systems eliminate the need for logs and generate files ready for analysis that are 1-2% of the size of a comparable raw log file, eliminating the time to transfer large files as well as the processing time. This applies to both ClickTracks Hosted service and the ClickTracks JDC product.

Hundreds of Millions of Visitors

ClickTracks JDC can record a virtually limitless amount of data. Even at its minimal requirements, a single low-end server, 2Ghz Pentium, with 1GB ram, routinely collects **4000 page views per minute** peak. This translates to approximately 30 million page views per month. Add more processing power and memory and the scalability increases significantly.

For higher volume websites, ClickTracks JDC is load-balanced across multiple servers for unlimited scalability. Traffic volumes exceeding **hundreds of millions of visitors per month are routine** work for a set of load balanced servers running JDC.

Privacy and Ownership

Whether or not you use web analytics, your website probably has a privacy statement with a sentence similar to the following:

"We will only share your information with third-parties outside OUR_COMPANY_NAME with your express permission or as required by the applicable law."

Or

"OUR_COMPANYNAME will never sell or share personal information with third-parties..."

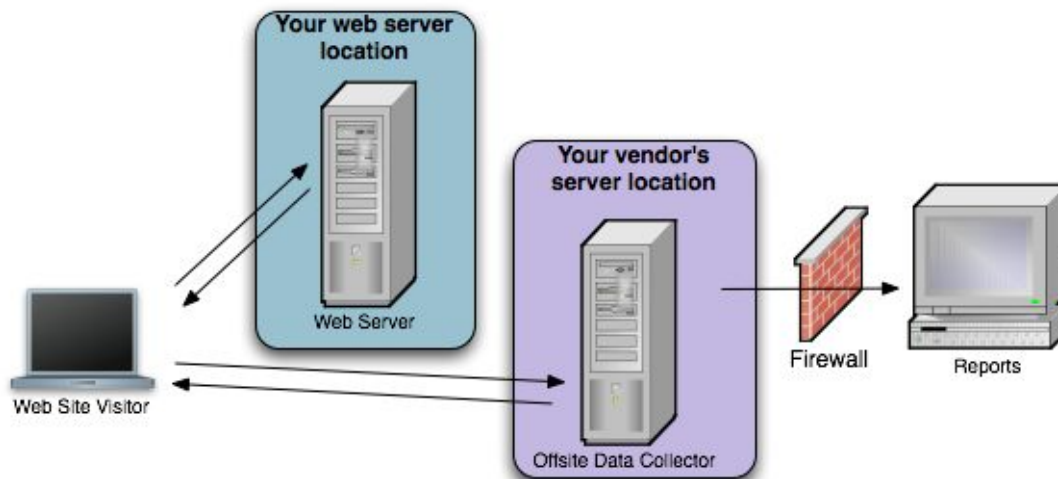
With third-party hosted JavaScript data collection, your visitors' activity is sent to and collected on a server that belongs to a third-party—perhaps even ClickTracks. According to most licensing agreements (ClickTracks excluded), the collected data belongs to the third-party—not you. This has not been a problem for web analytics vendors—until recently.

"Sharing" and "use" of visitor data now have unique definitions in light of who collects it. ClickTracks and a few of our competitors have no business benefit or purpose in looking at or using visitor data. However, WebSideStory blurs the lines by looking at aggregate visitor behavior, ostensibly to gain press coverage by [reporting](#) Web browser market share and even more disturbing, [specific conversion statistics](#) garnered from its customers' sites. Recent advertiser entrants into web analytics, i.e., Google and Kanoodle, have a clear business benefit and purpose in looking at and using your visitors' behavior data, and have no policy against doing so.

Beyond the legal ramifications of use of these systems, website *visitors* are becoming increasingly wary of websites [companies] that send their browsing activity to a third-party (see [users complain about iTunes use of Omniture](#), [Kanoodle pays people to use first-party cookies](#), and coverage of a Firefox plugin that blocks Google Analytics).

Tough Decision

The website owner faces a difficult decision: Give up privacy for more accurate, scalable JavaScript collection? Or, live with the less-scalable log file analysis while maintaining visitor privacy? This either-or choice is not necessary with ClickTracks JDC. ClickTracks JDC provides the benefits of JavaScript data collection combined with privacy because the data never leaves your own co-located servers.



A traditional hosted service model puts your data in 3rd party control... and sometimes, 3rd party use

Recurring Fees

ClickTracks JDC represents the 'buy' side of the traditional 'rent or buy' decision.

Traditionally, the ability to collect and process tagged-based visitor behavior data has been a specialized task, requiring proprietary software and servers with high bandwidth. Businesses using a subscription-based data collection system pay third-party vendors to maintain servers, collect, store and process click-stream data, and serve analytics to various client viewers.

ClickTracks is one of several vendors that offer hosted web analytics services for a monthly subscription fee.

By releasing a productized version of its own hosted data collection software, ClickTracks gives businesses the option to pay a one-time license fee, thus eliminating recurring costs and letting businesses collect page-tagged web analytics data on their own co-located servers.

Assuming the average price of a JDC implementation is \$16,000, the following chart represents time to return on investment based on price alone:

Current Monthly Subscription Costs	Months to ROI for JDC
\$ 200	80
\$ 500	32
\$ 1,000	16
\$ 2,500	6
\$ 5,000	3
\$ 7,500	2
\$ 10,000	2
\$ 15,000	1.1
\$ 20,000	0.8

ROI becomes a significant factor when monthly payments on hosted web analytics exceed \$500

Other factors that affect ROI will include:

- Cost of an additional co-located server
- Revenue/profit changes (typically increase) based on improved web analytics
- Cost of implementation and ongoing maintenance

Maintainability

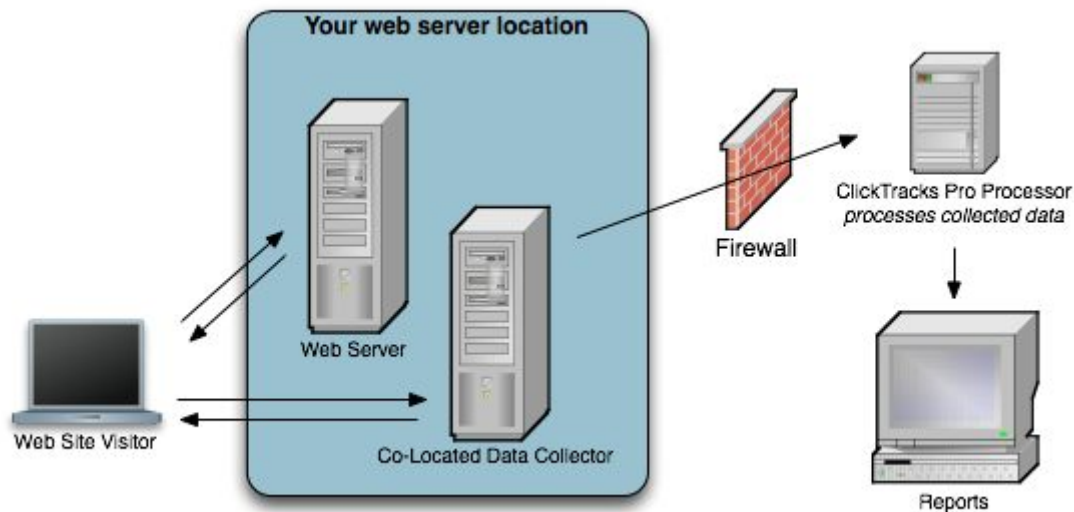
JavaScript data collection within ClickTracks is distinctly different from other services such as Omniture, WebSideStory (Hitbox, HBX) and Coremetrics. ClickTracks JDC is designed to be minimally intrusive and require no ongoing development effort for the marketing or IT teams. The architecture permits both ad hoc analysis of web analytics data, and more fundamentally, ad hoc *changes* to the reporting configuration, *without IT intervention*. In addition, The ClickTracks approach of simplifying JavaScript data collection is unique in the following ways:

- Deploy the same JavaScript on every page. Site-wide template include of 5-15 lines
- No use of custom JavaScript variables except in extreme circumstances
- Campaigns can be added at any time without changes to JavaScript
- Visitor segmentation analysis requires no additional programming
- Architecture that separates data collection from analysis
- Privacy of individual visitor data is maintained

Further technical and architectural differences are covered in a paper titled "[ClickTracks' Approach to JavaScript Data Collection](#)."

JDC Architecture

The JDC is software that permits web analytics data collection via JavaScript tagging instead of, or in addition to, log file analysis. The JDC software runs on your server hardware, typically located in the same co-location as your web server hardware. Much like a web server, the JDC is constantly alive, looking to accept requests (messages) from tagged pages. From these requests, the JDC generates a highly efficient representation of visitor behavior on the measured website that is approximately 1-2% of the size of a traditional log file.



ClickTracks JDC architecture keeps you in control of all data

It is important to note that ClickTracks efficiently captures all necessary data for future analysis. No additional JavaScript programming or return trips to the web pages are necessary for future custom queries and ad hoc analysis. Further technical and architectural differences are covered in a paper titled "[ClickTracks' Approach to JavaScript Data Collection.](#)"

The **ClickTracks Pro Processor** is a software program that runs on a local Windows PC, usually inside your firewall. The sensitive data and reports are thus secured, while the live data collection runs unattended at the data center. The Pro Processor collects and processes web analytics data, performs site archiving, automatically coordinates multiple data sets, and serves reports via the web to various clients including a Windows-based client viewer, web browsers and e-mail.

Reports are viewed from the Pro Processor using a standard web browser, or using the purpose build client application which permits analysts to run their own reports using a simple and efficient wizard based query system. Most importantly, the reports are ad hoc and require no changes to the JavaScript deployed.

A Typical Data Collection and Analysis Scenario

1. To begin data collection, each page to be tracked must include a JavaScript tag. This is often implemented through a simple site-wide include.
2. When visitors reach a tagged web page, the tag captures visitor activity and then sends it over the Internet to the JDC server.
3. The JDC interprets the information from the JavaScript tags and creates the raw click-stream data that will be the basis for future analysis.

4. ClickTracks Pro Processor regularly polls the JDC server and reads the latest data for local processing and report generation. The result is ready-to-view web analytics reports.
5. Marketing and executive staff use web browsers or client viewers to view web analytics reports.

Integrating with a Data Warehouse/OLAP

The JDC produces click-stream data in the form of flat files, organized in a data structure that is exposed through an XML file. The click-stream data can be fed into data warehouse and OLAP tools such as Brio and Hyperion. This creates a single, unified repository for all online and offline aspects of the business, creating consistency in data collection.

The ClickTracks JDC provides documented, direct access to all data gathered from online visitor activity, including URLs, referrers, cookie values, and more. Advanced users can augment the reporting of ClickTracks with their own tools, integrating offline data. The data is an exact match to what is reported in ClickTracks since the source is the same.

Technical Requirements

ClickTracks JDC does not require specialized resources; in fact, it runs on software and hardware commonly found in most business settings.

Basic technical requirements (JDC)

- Co-located server with 2GHz P4 , 80GB hard drive, 1GB RAM
- Unix or Linux OS
- Apache/IIS web server running Mod_Perl and PHP

Basic technical requirements (Pro Processor)

- Windows NT/2000 /XP Service with admin UI
- Pentium III 1.67GHz
- 1GB RAM
- 500Mb free disk space

JavaScript Tags

The JavaScript tag is typically between 5-15 lines and is the same for every page on the website. The script is smaller if it uses your existing first-party cookies (cookies already set by the website), and slightly larger if it must set a new first-party cookie for you. At its largest, the script can grow to 15 lines if it is used to track site exit links.

The User's Perspective

All architectural and implementation details are designed to be transparent to the end user. To access ClickTracks reporting and analysis, the end user simply opens a client viewer and begins analysis of up-to-date visitor data.

Analysis of a website or the relevant online marketing may be performed through a Windows client viewer, a web browser or via scheduled e-mail reports.

White Paper



Web analytics features available through use of the JDC match or exceed those of the popular ClickTracks Pro product used in the broader web analytics market. These features include:

- Navigation Report
 - Overlay View
 - Keywords per page
 - Robot Simulation
 - AB/Split Views
 - Archived Site Split Views
- Keyword Report
 - Keyword Ranking
 - Revenue by Keyword
- Site Overview
 - Internal Search
 - Geographic Analysis
- What's Changed Report
- Funnel Report
- Campaign Report
 - Click Fraud Report
- Report export to PDF/Excel
- Full Revenue Analysis
- On-the-fly visitor segmentation analysis

Note: ClickTracks' Robot report requires log files in order to detect and report search engine robot activity. For the same reason, the click fraud report, when used with JavaScript, may be less able to detect fraud robots.



Web analytics features available through use of the JDC match or exceed those of the popular ClickTracks Pro product.

Conclusion

ClickTracks JDC provides the accuracy and scalability associated with JavaScript page tagging, as well as the data privacy, cost and ownership benefits associated with log file-based systems. Businesses now have the option to use their own secure servers to implement the same software that ClickTracks uses to collect and serve web analytics, keep visitor data private and eliminate recurring costs.

Company Information

Based in Santa Cruz, California, ClickTracks is the maker of award-winning web analytics software. The ClickTracks family of software was designed out of sheer frustration with existing web metrics offerings. Spending hour upon hour trying to understand web site visitors was certainly a motivating experience: there had to be a better way.

ClickTracks was voted 'Best Web Site Analysis Tool' in ClickZ's 2003, 2004 and 2005 Marketing Excellence Awards, and ClickTracks earned a "Positive" rating in Gartner's 2006 MarketScope for Web Analytics. The company received a Computerworld Innovative Technology award in the web site management category and ClickTracks' software has been reviewed and featured on CNET, ZDNet, Builder.com, TechRepublic, Internet.com, CRM Magazine, Media Magazine, MarketingSherpa and About.com. The ClickTracks line of software includes ClickTracks Analyzer, ClickTracks Optimizer, ClickTracks Professional, ClickTracks Hosted, ClickTracks Agency Edition, ClickTracks ISP and ClickTracks JDC. For more information, please visit www.clicktracks.com or send e-mail to info@clicktracks.com.

Contact:

ClickTracks Analytics Inc.
501 Mission Street
Santa Cruz, CA 95060

Website: www.clicktracks.com

Toll-Free: (877) 773-2249
Outside the US: (831) 621-6380
Fax: (831) 480-5728
UK Office: +44 (0)207 669 4296

Sales Inquiries: sales@clicktracks.com