



Spiders and Bots

Help Documentation

Spiders and Bots

Traffic

Spiders

Spiders and robots are applications that crawl the Web indexing and retrieving content, usually for the benefit of search engines, information resources, and news organizations. For example, the primary search engine robots for Google, Bing, and Yahoo! are Googlebot, Slurp, and MSNbot.

This report displays information about the search engine spider activity (spider, page views, hits, and bandwidth) on your website. Site administrators and users can use this report to evaluate and enhance their SEO efforts.

Note: This report can be added as a report item to a custom report. For more information on report items, refer to [Understanding Report Items](#) .

For a definition of the terms listed on this page, please refer to the [Glossary](#) .

Spider Traffic Trend

This report displays all of the spider and/or bot activity (page views hits, and bandwidth) on your website for the selected date range. Note: You can specify how data is displayed (hourly, daily, weekly, etc.) by using the Group By report option.

Note: This report can be added as a report item to a custom report. For more information on report items, refer to [Understanding Report Items](#) .

For a definition of the terms listed on this page, please refer to the [Glossary](#) .

Recent Spider Details

This report displays information about your most recent website spiders (based on a 15 minute period). Information available pertains to the last hit logged for the spiders and information about their path through the site.

Note: This report can be added as a report item to a custom report. For more information on report items, refer to [Understanding Report Items](#) .

For a definition of the terms listed on this page, please refer to the [Glossary](#) .

Recent Spider Activity

This report displays information about your most recent website spiders (based on a 15 minute period). Information available pertains to the first hit for the spiders, the last logged hit for the spiders, and their hits, views and bandwidth usage.

Note: This report can be added as a report item to a custom report. For more information on report items, refer to [Understanding Report Items](#) .

For a definition of the terms listed on this page, please refer to the [Glossary](#) .

Files

Pages Not Indexed

This report displays a list of the webpages that were not indexed by spider and/or bots. Site administrators and users can use this report to determine which pages may not be available in search engine results pages. Pages can be excluded for a number of reasons. For example, they're excluded in robots.txt files, there are no internal links to the pages, etc.

Note: This report can be added as a report item to a custom report. For more information on report items, refer to [Understanding Report Items](#) .

For a definition of the terms listed on this page, please refer to the [Glossary](#) .

Pages Indexed Not Viewed

This report displays a list of the webpages that were indexed by spiders and/or bots but that were not viewed by the public. Site administrators and users can use this report to identify pages that may have been deleted recently or that may need to be optimized to increase page views.

Note: This report can be added as a report item to a custom report. For more information on report items, refer to [Understanding Report Items](#) .

For a definition of the terms listed on this page, please refer to the [Glossary](#) .

Spider Pages

This report displays a the top pages of your website crawled by spiders and/or bots. By selecting a specific spider in the report options pane, you can see the top pages indexed by that spider.

Each row returned in this report can be further analyzed by SmarterStats. By moving your mouse over a row, a down arrow appears at the end of the row. The options available when clicking on this arrow include: Data Mining, Open URL, Set Page Alias and View Details.

- Data Mining - Data mining allows site administrators and users to drill down into more detail about the item in the particular row. For example, viewing the referring URLs, bandwidth used, query strings and more.
- Open URL - Opens the item in a new window.
- Set Page Alias - Allows users to set a friendly name for sometimes complex URLs. For example, if a Page shows up as "/smartertools/products/smarterstats/pricing.aspx", users can instead attach the alias of "SmarterStats Pricing". That way, any report items that reference that page URL will display "SmarterStats Pricing" instead of the URL itself. This feature is especially useful when using dynamic pages, which can have cryptic or numerical variables such as ProductID=37.
- View Details - Displays a number of details about the specific item. For example, Traffic details, Performance results, any result codes returned (e.g., 404 errors) or any Site Tuning issues that are found.

Note: This report can be added as a report item to a custom report. For more information on report items, refer to Understanding Report Items .

For a definition of the terms listed on this page, please refer to the Glossary .

Spider Directories

This report displays the website directories (folders and paths) that received the most page views by spiders and/or bots.

Note: This report can be added as a report item to a custom report. For more information on report items, refer to Understanding Report Items .

For a definition of the terms listed on this page, please refer to the Glossary .

Spider Top Files

This report displays the website files (web pages, images, etc.) most frequently accessed by spiders and/or bots, regardless of the file extension.

Each row returned in this report can be further analyzed by SmarterStats. By moving your mouse over a row, a down arrow appears at the end of the row. The options available when clicking on this arrow include: Data Mining, Open URL, Set Page Alias and View Details.

- **Data Mining** - Data mining allows site administrators and users to drill down into more detail about the item in the particular row. For example, viewing the referring URLs, bandwidth used, query strings and more.
- **Open URL** - Opens the item in a new window.
- **Set Page Alias** - Allows users to set a friendly name for sometimes complex URLs. For example, if a Page shows up as "/smartertools/products/smarterstats/pricing.aspx", users can instead attach the alias of "SmarterStats Pricing". That way, any report items that reference that page URL will display "SmarterStats Pricing" instead of the URL itself. This feature is especially useful when using dynamic pages, which can have cryptic or numerical variables such as ProductID=37.
- **View Details** - Displays a number of details about the specific item. For example, Traffic details, Performance results, any result codes returned (e.g., 404 errors) or any Site Tuning issues that are found.

Note: This report can be added as a report item to a custom report. For more information on report items, refer to [Understanding Report Items](#) .

For a definition of the terms listed on this page, please refer to the [Glossary](#) .

[Server Responses](#)

Spider 404 - Page Not Found

This report displays a list of the files that a spider attempted to access that no longer exist on the server, possibly because they were deleted, they were never created or the URL was misspelled. While some 404 errors are common due to user error. Many 404 errors to a specific resource indicate a problem with your site. Typically, a 404 result to a spider will result in the page being removed from the spider's index and could affect your SEO efforts.

Note: This report can be added as a report item to a custom report. For more information on report items, refer to [Understanding Report Items](#) .

For a definition of the terms listed on this page, please refer to the [Glossary](#) .

Spider 5xx - Server Errors

This report displays a list of the files that resulted in an internal server error when trying to process the request. These errors tend to be with the server itself, not with the request, and usually occur when there is a problem on your web server preventing the data from being returned (usually a scripting

problem). While some 5xx errors occur due to server restarts, any 5xx error that occurs more than a few times should be investigated.

Spiders encountering 500-level errors will usually try loading the page again later. After several failed retries, the page will likely be removed from the spider's index.

Note: This report can be added as a report item to a custom report. For more information on report items, refer to [Understanding Report Items](#) .

For a definition of the terms listed on this page, please refer to the [Glossary](#) .

[Server Utilization](#)

Spider Page Bandwidth In

This report displays the incoming bandwidth for pages on the site accessed by spiders and bots. Website developers can use this report to identify the pages that will need the most optimization on incoming bandwidth. High utilization scores in this section are typically due to very large amounts of form data or viewstate, but can also be attributed to pages that allow file uploads. Note: If your site is based on ASP.Net, the pages in this page should be examined to see if viewstate can be reduced.

Regarding "Severity" - this is a calculated field that is based on the impact of the measured result (e.g., page load time) multiplied by the number of pages affected, then normalized across the entire result set. In general, Severity can be used as a way of making responsible decisions about the impact your site is having on the server and to avoid making unnecessary optimizations. For example, pages with high load times that are accessed infrequently may have lower severity than smaller pages that are hit much more often. In this case, there's no reason to optimize the higher load time page as it has less impact on the server.

Note: This report can be added as a report item to a custom report. For more information on report items, refer to [Understanding Report Items](#) .

For a definition of the terms listed on this page, please refer to the [Glossary](#) .

Spider Page Bandwidth Out

This report displays the outgoing bandwidth for pages on the site hit by spiders and bots. Website developers can use this report to identify the pages that will need the most optimization on outgoing bandwidth.

Highly accessed pages that are smaller than others may have a higher score because of the number of hits. In this way, the report shows you the key pages that will have the most impact if optimized. For

example, a page that is hit 10,000 times a day that is 500 bytes long causes much more load than a page that is hit 10 times that is 10,000 bytes long.

To optimize the page size, reduce the amount of HTML or content in the page or enable HTTP compression on the web server.

Regarding "Severity" - this is a calculated field that is based on the impact of the measured result (e.g., page load time) multiplied by the number of pages affected, then normalized across the entire result set. In general, Severity can be used as a way of making responsible decisions about the impact your site is having on the server and to avoid making unnecessary optimizations. For example, pages with high load times that are accessed infrequently may have lower severity than smaller pages that are hit much more often. In this case, there's no reason to optimize the higher load time page as it has less impact on the server.

Note: This report can be added as a report item to a custom report. For more information on report items, refer to [Understanding Report Items](#) .

For a definition of the terms listed on this page, please refer to the [Glossary](#) .

Spider Page Time Taken

This report displays the average time taken for specific pages to load in the given timeframe when hit by spiders and bots. Website developers can use this report to identify the pages that need the most optimization with regards to page loading.

To optimize the time taken on the page, optimize the code to reduce database calls, improve the page flow to reduce the work the server has to do, or implement caching.

Regarding "Severity" - this is a calculated field that is based on the impact of the measured result (e.g., page load time) multiplied by the number of pages affected, then normalized across the entire result set. In general, Severity can be used as a way of making responsible decisions about the impact your site is having on the server and to avoid making unnecessary optimizations. For example, pages with high load times that are accessed infrequently may have lower severity than smaller pages that are hit much more often. In this case, there's no reason to optimize the higher load time page as it has less impact on the server.

Note: This report can be added as a report item to a custom report. For more information on report items, refer to [Understanding Report Items](#) .

For a definition of the terms listed on this page, please refer to the [Glossary](#) .