Main

Use the main window to run and modify existing reports, create new reports, update the database or set various options.

Run Existing Reports

Select the report you want to run from the **Reports** list then click the **Run** button or select Run from the Report menu. You can also right-click on the report then select Run from the pop-up menu.

Modify Existing Reports

Select the report you want to run from the **Reports** list then click **Design** or select Design from the Report menu. You can also right-click on the report then select Design from the pop-up menu.

Create a New Report

Click the **New** button or select New from the Report menu.   If you want to create a new report that is similar to existing one, click Yes when asked if you want to base the new report on an existing one.

Tip: The *Template* report is a good starting point for new reports.   It includes a professionally-designed color scheme and standard report elements.

Updating the Database

Hit List creates reports based on data stored in its database, not directly from your log files.   Log files must be imported into the database before they can be analyzed.   Select Update Database from the Tools menu to update the database now.   You can also use Scheduled Reports to automatically update the database at specific times.

Setting Options

There are two different types of options that can be set.   The Options choice listed under the Tools menu lets you set various common options such as which web browser to use, how to update the database, where log files are located, etc.   The Database Manager choice listed under the Tools menu lets the system administrator choose various options that affect data already in the database such as Groups, Operating Systems, Browsers and Object Types (Professional Edition only).

Using a Different Reports Database or Importing Reports

Report definitions are stored in a database.   You can organize your reports into groups by using multiple databases. You can even share reports and report elements with other people by importing them from different report databases or sharing them across a LAN.

Select Open Report Database from the Reports menu to use a different collection of reports.   Hit List 2.0 can open Hit List 2.0 databases (HLRPT20P.MDB or HLRPT20S.MDB) and 1.0 report databases (MWUSER.MDB).

Select Import from the Reports menu to select a reports database containing entire reports or report elements that you would like to copy into the current reports database.   You may import from a Hit List 1.0 database or a Hit List 2.0 database.   See Import for details.

Tip: The *Template* report is a good starting point for new reports.   It includes a professionally-designed color scheme and standard report elements.

Report Scheduler

Use the Report Scheduler to automate running reports.   This is especially useful for automatically sending daily reports via email.

Hit List must either be running as a Microsoft Windows NT Service or left running with you still logged-in for Scheduled Reports to occur.

Important: Scheduled Reports will not run unless you turn **Disable Report Scheduler** (in the Tools menu on the main window) off.   This is provided so that scheduled reports do not run while you are using Hit List yourself.

Tip: If you want to update the database but not spend the time creating a complicated report, create a very simple report (like *Total Requests*) and ensure that **Update the database before running** in the Outline tab of the Design dialog is on.   Note that reports that do not ask for any calculations will not update the database, regardless of how this switch is set.

Tip: If you don't want the web browser to appear after your scheduled report is created, uncheck the **Launch web browser after a Scheduled Report** checkbox on the General tab of the Options dialog.

Create a New Scheduled Report

Click **New** then select a report and its schedule.   If you want more than one report to run at approximately the same time (they will be queued as necessary), use Shift and Ctrl to select multiple reports.

Tip: To run one report for each of several (or all) virtual servers, you can simply use the Run this report for each virtual server checkbox in the Design window for that report.

Modify an Existing Scheduled Report

Select an item from the **Scheduled Reports** list and click **Change**.

Delete an Existing Report

Select an item from the **Scheduled Reports** list and click **Delete**.   The report will no longer be scheduled to run. Deleting a Scheduled Report does not delete the report itself.

See Also:
Edit Schedule
Email

Edit Schedule

Reports can be scheduled to run at the same time on specific (or all) days or can use a completely custom schedule.

Daily Schedule

To establish a recurring daily schedule, enter the time you want the report to run in the **Time** box using a 24 hour clock (i.e. 11PM is 23:00:00) and select the day to run the report.

Tip: You can schedule a report for multiple days by creating several Scheduled Reports using the same report.   For example, you could create five Scheduled Reports to run a report Monday through Friday.

Custom Schedule

If the Daily Schedule system is not sufficient, you can create a Custom Schedule.   Select the date and time to begin then select the frequency using the **Repeat every** fields.   For example, to create a Scheduled Report that runs every two weeks starting today, enter today's date and the time to begin then enter 2 Weeks in the **Repeat every** fields.

Database Manager/URL Groups

URL Groups are collections of URLs that are analyzed as one unit, frequently simplifying reports and showing trends that might have been missed otherwise. For example, the Marketing Group could contain any URL that begins with *mktg, /Default.htm* or */downloads/ourdemo.zip*.   A report looking at Groups would count any of these as hit against the Marketing Group.   This feature is also commonly used to combine pages such as */* and */Index.html* that are really the same page but may appear either way in your logs.   It's also very useful for combining frame and no-frame versions of the same page (e.g. */Default.htm* and */frames-default.htm*)

Groups are managed from within the Database Manager.   To run the Database Manager, select Database Manager from the Tools menu of the main window.   Groups affect the database so you must have access to the database.

Any report element that normally uses URLs can be based on Groups instead by checking **Substitute group name for URL** in the <u>Definition</u> tab of the Properties window.

Groups can be based on complete URLs or on URL patterns.   A URL pattern contains a portion of a complete URL and the standard wildcard characters (* and ?).   For example,:

1.   A URL pattern of **mktg** includes any URL that contains *mktg* anywhere.   Case is not important.

2.   A URL pattern of */mktg/productinfo?.htm** includes any URL in the */mktg* directory that begins with *productinfo*, includes exactly one additional character and is an HTML page.   This would include */mktg/productinfo1.htm* and */mktg/productinfo2.html* but not */mktg/productinfo27.htm* or */NewStuff/mktg/productinfo3.htm*.

Tip: Groups based on wildcards take longer to calculate than groups based on complete URLs so you should not use wildcards needlessly.   For example, if you always end HTML pages with the *.HTM* extension, do not specify URLs with *.HTM** extensions.

Tip: Checking the **Assume each URL is member of no more than one group** switch in the <u>Options</u> tab of the Design window allows Hit List to combine URL groups (such as *Home Page*) with URLs that are not part of a group (like */graphics/mystuff.gif*).   This is the default behavior and normally produces the desired results.   However, it is possible to construct groups such that a URL is a member of more than one group.   For example, a URL may be part both of a *Marketing* group and a *Products* group.   If so, Hit List will normally just count it as a member of the first group in the Groups list.   However, if you want to see a report that shows it listed as both, you can uncheck the **Assume each URL is a member of no more than one group** switch.   Doing so may cause unexpected behavior, however, because double-counting of URLs is now possible.   Additionally, Hit List will not be able to combine 'normal' URLs and URL Groups so reports generated when this switch is off will only show URL Groups, suppressing URLs that are not members of any group.

If you change a URL Group, Hit List will recalculate all URL Groups the next time a report is run.

Creating a New Group

Click the **New** button next to the **Groups** list and enter then name of the new Group.

Modifying an Existing Group

Select the Group in the **Group** list and click the **Change** button next to the **Groups** list.   Enter the Group's new name.

Deleting an Existing Group

Select the Group in the **Group** list and click the **Delete** button next to the **Groups** list.   This removes the Group but does not delete any data from the database.

Adding a New URL or URL Pattern to an Existing Group

Select the Group that the new URL belongs to from the **Groups** list then click the **New** button next to the **URL within Group** list.   Enter the complete URL or URL pattern.

Modifying a URL or URL Pattern

Select the URL or pattern from the **URL within Group** list then click the **Change** button next to it.   Enter the new URL or URL pattern.

Deleting a URL or Pattern

Select the URL or pattern from the **URL within Group** list then click the **Delete**   button next to it.   This disassociates the URL from the currently selected Group but does not delete any data from the database.

See Also:
Object Types
Groups
Browsers
Operating Systems
Tools
Site Name Groups
App Argument Groups

Database Manager/Object Types

Object Types are like simple <u>Groups</u> in that they categorize URLs for easier analysis.   They are also used to filter-out   unwanted data to simplify reports.   For example, unchecking **Graphic** from the <u>Filter</u> tab in the Design window tells a report to ignore all graphic elements, usually removing a great deal of useless information.

There are eight Object Types: Application, Audio, Download, Graphic, Page, VB/Java, Video and Other.   URLs are categorized into Object Types by their extension (e.g. .htm and .html are Pages, .GIF and .JPG are Graphics, etc.). Every URL is a member of exactly one Object Type group.

Standard Object Types and related extensions are automatically created when the database is created.   You may add, delete and modify the extensions associated with each of these types but you cannot modify the Object Types themselves (unlike Groups).

Object Types are managed from within the Database Manager.   To run the Database Manager, select Database Manager from the Tools menu of the main window.   Object Types affect the database so you must have access to the database.

If you change a Type Extension, Hit List will recalculate all Object Types the next time a report is run.

Adding a New Type Extension

Select the Object Type to modify from the **Type** list then click **New**.   Enter the new extension.

Changing an Existing Type Extension

Select the Type Extension then click **Change**.   Enter the revised Type Extension.

Deleting a Type Extension

Select the Type Extension to disassociate from the selected Object Type and click **Delete**.

See Also:
<u>Filter</u>
<u>Groups</u>
<u>Browsers</u>
<u>Operating Systems</u>
<u>Tools</u>

Database Manager/Browsers

Although HTML is considered a standard, the major web browsers add their own enhancements, sometimes with conflicting syntax.   Additionally, some web browsers do not support commonly used features such as tables.   If your pages use features that your visitors cannot understand, you could be losing customers.   Likewise, if most of your visitors use a particular browser, you might add features to enhance their experience with your site.   Hit List lets you create reports that show which browsers your visitors are using.

Hit List identifies browsers by looking for particular words.   Since multiple browsers may contain some of the same words, the search priority is important.   Hit List looks for browsers listed towards the top of the **Browsers** list before looking for browsers at the bottom.   You can change browser priority with the **up** and **down** arrow buttons.

If you modify any browser information, Hit List will recalculate browser data the next time a report is run.

Browsers are managed from within the Database Manager.   To run the Database Manager, select Database Manager from the Tools menu of the main window.   Object Types affect the database so you must have access to the database.

Browser and Operating System information is not always added to server log files.   If the information is not in your log files, Hit List has no way of calculating it.   If you are using the Microsoft Internet Information Server, you can add Browser and Operating System information to your log files by installing the NewLog.DLL that comes with Hit List.   If you did not choose to install this during Setup, you can run Setup again to add it.   If you are using O'Reilly WebSite, the optional WinLog log format is required.   Once again, you can run Setup again and tell it to make the change.

Browser and Operating System information is determined by looking at the USER_AGENT string supplied by most browsers.

Adding a New Browser

New web browsers or new versions of existing browsers arrive every day.   Hit List can be instructed to look for the new browsers via the **New** button next to the **Browsers** list.   Click the **New** button then enter the name of the new browser.

Tip: Do not enter a new browser for minor revisions of existing browsers or your reports may become needlessly cluttered.   If the existing identifiers are insufficient, simply add a new identifier.

Modifying the Name of an Existing Browser

Select the browser from the **Browsers** list then click the **Change** button next to it.   Enter the new name of the browser.

Deleting an Existing Browser

Select the browser from the **Browsers** list then click the **Delete** button next to it.

Adding a New Identifier

Select the browser to associate the new identifier with then click the **New** button next to the identifier list.   Enter the new identifier.   *Note:* The use of this word or words anywhere within the USER_AGENT field causes a match.

Modifying an Existing Identifier

Select the identifier in question and click the **Change** button next to it.   Enter the new identifier.

Deleting an Existing Identifier

Select the identifier then click the **Delete** button next to it.

See Also:
Groups
Browsers
Object Types
Tools

Database Manager/Operating Systems

Hit List lets you create reports that show which operating systems your visitors are using.

Hit List identifies operating systems by looking for particular words. Since different operating systems may contain some of the same words, the search priority is important. Hit List looks for operating systems listed towards the top of the **Operating Systems** list before looking for operating systems at the bottom. You can change operating system priority with the **up** and **down** arrow buttons.

If you modify any operating system information, Hit List will recalculate operating system data the next time a report is run.

Operating Systems are managed from within the Database Manager. To run the Database Manager, select Database Manager from the Tools menu of the main window. Object Types affect the database so you must have access to the database.

Browser and Operating System information is not always added to server log files. If the information is not in your log files, Hit List has no way of calculating it. If you are using the Microsoft Internet Information Server, you can add Browser and Operating System information to your log files by installing the NewLog.DLL that comes with Hit List. If you did not choose to install this during Setup, you can run Setup again to add it. If you are using O'Reilly WebSite, the optional WinLog log format is required. Once again, you can run Setup again and tell it to make the change.

Browser and Operating System information is determined by looking at the USER_AGENT string supplied by most browsers.

Adding a New Operating System

If an operating system that is important to you is not already listed, you can add it by clicking the **New** button next to the **Operating System** list and entering it.

Tip: Keep your reports clear by not adding a new operating system unless you plan to use the information.

Modifying the Name of an Operating System

Select the operating system from the **Operating Systems** list then click the **Change** button next to it. Enter the new name of the operating system.

Deleting an Existing Operating System

Select the operating system from the **Operating Systems** list then click the **Delete** button next to it.

Adding a New Identifier

Select the operating system to associate the new Identifier with then click the **New** button next to the identifiers list. Enter the new identifier. Note: The use of this word or words anywhere within the USER_AGENT field causes a match.

Modifying an Existing Identifier

Select the identifier in question and click the **Change** button next to it. Enter the new identifier.

Deleting an Existing Identifier

Select the identifier then click the **Delete** button next to it.

See Also:
Groups

Browsers
Object Types
Tools

Database Manager/Tools

The Tools tab provides various ways for advanced users to manage the Hit List database. Some of these tools could potentially cause data loss so we recommend extreme care when using them.

Delete all records that match the criteria below

As the Hit List database grows over time, it may be useful to delete information from old requests to conserve disk space or speed report generation.

The *From* and *To* fields are used to specify the starting and ending points for the delete operation. Either (or both if you use the *Server name or IP* field) may be left blank. All requests made between and including the time period will be deleted. Any IP address that was seen only during the period will also be deleted. To delete all records from before a certain date, enter the date into the *To* field and leave the *From* field blank. Similarly, to delete all requests made after a particular date, enter the date in the *From* field and leave the *To* field blank. Dates should be entered in mm/dd/yy format.

Web sites that host multiple virtual servers can use the *Server name or IP* field to selectively delete records from a particular virtual server. Enter the IP or name (whichever your server logs) of the server in question. You may combine an entry in this field with a date range using the *To* and *From* fields.

Clear log file records

Hit List keeps track of which log files it has seen before. This lets it intelligently manage multiple log files. However, there may be times where you need to re-load data from a particular log file that Hit List has already processed. Hit List will not allow this unless you erase the log file information from its database, essentially causing it to forget that it has seen the file before.

You may either delete all log file records or delete just one. Use the *All* option very carefully. To delete any one record, choose the *Just from this log file option* and enter the full path name of the log file (i.e. c:\mylogs\access.log). Wildcards are not recognized.

Tip: Enter *MS IIS ODBC Log Table* in the file name field to clear the information about logs stored in an ODBC datasource.

Delete the entire Hit List database

If your database is stored in a file, this option will delete the Hit List database file from your computer. If your database is stored on an ODBC datasource, this option will drop all the Hit List tables. There is no way to undo this operation so be very careful.

Compact the Hit List database

Over time your Hit List database may grow somewhat larger than necessary. This option will remove empty space from your database, effectively shrinking it to the smallest possible size. This does not involve any data compression and cannot harm your data but you will need spare disk space for temporary processing. We suggest compacting periodically, perhaps once a month. This option is not available for databases stored in an ODBC datasource.

See Also:
Groups
Browsers
Object Types
Operating Systems

Options/General

The General tab determines how the Hit List application interacts with the Hit List engine, which web browser is used to view reports and whether Tip of the Day is shown when the application starts.

Web browser

Hit List uses your web browser to display reports.   If you're running Windows NT 3.51 you must enter the full path of your browser or use the **Browse** button to find it.   If you are running Windows 95 or Windows NT 4.0, it is not necessary to select a browser if you have one properly installed; the operating system will simply find it.   In that case, the **Web browser** entry will be ignored.

Default domain

Hit List can automatically turn cryptic URLs into friendly page titles.   However, because log files normally note URLs without the server name (i.e. */index.htm* instead of *www.mycompany.com/index.htm*), you must enter the default domain name – normally in the form of *www.yourcompany.com.*

Tip: If necessary, you can override this global settings for any report by entering a different name in the Default domain field in the Options tab of any report.

Tip: This setting is ignored if Run this report for each virtual server is checked on a report.   In that case, Hit List will determine the correct value itself.

Tip: When listing HTML pages, Hit List can display just the URL, just the page title or both (the default).

Tip: If necessary, Hit List can resolve page titles by going through a proxy server.   See the Advanced tab of the Options dialog box for details.

Mode

Hit List can operate in two basic modes, **Remote** and **Direct**.   They generate identical reports, only differing in where the report is run.

Remote (Launch a web browser and run the report on the server)

In *Remote Mode*, the application assumes that your machine is not the web server itself but that a copy of Hit List has been installed on the server.   When you run a report, your copy of Hit List generates an HTML page that your web browser launches.   When you click OK on that page, your browser submits the report request to your server. The server runs its copy of Hit List, generates the report, and sends it back to your web browser.   This approach is extremely efficient because it means that very complex reports can be created by people who don't have direct access to the web server or who don't have the CPU or disk required to manage large databases.   For example, traveling marketing and sales people can use *Remote Mode* to get reports while they are on the road.   Or an Internet Service Provider can have each client generating their own reports without the ISP performing any work.   However, *Remote Mode* cannot provide any intermediate status because the web browser is simply waiting for the final report.

To use *Remote Mode*, you must enter the full URL of the Hit List report engine in the **Application URL** field.   This is normally in the form of *http://www.mycompany.com/Stats/MWHL20.DLL   (or EXE for WebSite and Netscape)*.

See Using Remote Reports for much more detail.

Direct (Run the report on this machine using my copies of the log file and database)

When Hit List is installed on a server, it normally runs in *Direct Mode*.   In *Direct Mode*, the Hit List application communicates directly with the Hit List engine, bypassing the need for an intermediate web browser.   This allows more efficient status reporting because the engine can send status messages to the main application as the report is generated.   *Direct Mode* is also used if you have arbitrary log files that you want to analyze.

Email

Hit List has the unique ability to send reports via email.   Combined with <u>Scheduled Reports</u>, this is especially powerful because you could have reports generated early in the morning and automatically sent to all interested parties.   To send via email, Hit List needs to know a valid email name and password.   This is necessary so it can logon as it sends the mail.   Passwords are stored in an encrypted format.

Email can be sent via either Microsoft Exchange (also known as MAPI) or via standard Internet POP/SMTP mail.   If you're using POP/SMTP, you may have to provide the name of your email server and the domain from which the mail will be sent.

Important: To send email via MAPI, your email client (usually Exchange or MS Mail) must already be running.

Tip: There is no need to tell Hit List your email name.   Your system administrator can establish a special email account for the application to use.

Email logon name

Enter the logon name for Hit List to use when sending email.   If you're sending via your company's LAN, the email name is probably in the form of *aliceb* or *Alice Baker*.   If you're sending via POP/SMTP, the email name may be in the form of *aliceb@yourcompany.com*.   Ask your system administrator if you are unsure.

Password

Enter the password associated with the email logon name.   The password will be shown as asterisks and stored in an encrypted format.   Be sure to use the correct upper/lower case combination if your system differentiates.

Server name (only necessary for POP/SMTP)

Enter the name of the mail server that will be sending the email.   This is normally in the form of *mail.mycompany.com*.

Email domain (only necessary for POP/SMTP)

Enter the domain from which the email should be sent.   This is normally something like *mycompany.com*.   **Note**: Entering the email domain may or may not be necessary, depending on your email system.   Ask your system administrator if you are unsure.

Show tips at startup

Check this box if you want to see Tip of the Day appear when you begin the application.   You can also set this in the Tip of the Day dialog box itself.

Launch web browser after a Scheduled Report

Your web browser is normally launched after a scheduled report is created.   However, you may find this inconvenient if you have several scheduled reports.   Uncheck this box to suppress the usual web browser launch.   Web browsers are never launched when Hit List is running as a Windows NT Service.

Options/Log File

For greater speed and size efficiency, Hit List processes your server logs into its own database. Reports are run against the database rather than the logs themselves.

Most web servers create an access log in the form of a text file stored on the server. Microsoft Internet Information Service also has the option of logging to an ODBC datasource such as Microsoft SQL Server. Hit List can read log information using either system.

It is not normally necessary to enter the location of log files if Hit List is running on the server because it will query your computer for the information it needs without your intervention It is only necessary to supply a file name if you want to force a particular log file into the database or you are using a server other than Microsoft Internet Information Server, Netscape (all servers) and O'Reilly WebSite (all servers). You also may need to do this if you're using Netscape and hosting virtual sites.

If you are not running Hit List on the web server, you'll need to tell it the location of your log file or files. To specify a particular log file, simply use the **Browse** button to choose it or type the full path name. If you want Hit List to find multiple log files, you can usually express this by entering a file name pattern instead of a specific name. For example, to find all Microsoft IIS log files, you could enter something like:

c:\MyLogs\in*.log

This would find all files that begin with *in* and end with *.log*. If new logs appear in that directory they will be automatically found and processed by Hit List (see the log cycling explanation below). For more information, see the **My server logs to a file** section below.

If you are using *Remote Mode*, the Hit List engine on the server will find the files. You should leave the file name field blank in these cases.

Log file cycling (if any) is handled transparently by Hit List; it is not necessary for you to do anything special. Hit List keeps track of all the log files on your server and remembers information about those it has seen before. Hit List automatically keeps duplicate and old log entries out of the database. If you need to add old data to the database, you can disable this feature by checking **Allow old data to enter the database** in the Updates tab of the Options dialog box.

Tip: If you're using Microsoft Internet Information Server and Hit List is running on the server, it is not necessary to tell the application if the logs are stored in a file or an ODBC datasource; Hit List will determine this by itself.

Tip: If you're hosting multiple virtual servers and using a web server that does not put any indication of which server is answering the request (Netscape, for example) into the logs, it is not a good idea to put all your logs into one database. Instead, use the Log File and Database override switches in each report to use specific logs and database combinations.

My server logs to a file

If your server logs to a file (as opposed to an ODBC datasource), you may optionally enter the name of the log file. File patterns are also permissible. For example, *c:\myserver\access*.log* finds all files in the *c:\myserver* folder that begin with *access* and have the *.log* extension.

To download logs from an FTP server, simply enter the URL of the file name or pattern. You can optionally add a username (such as Anonymous) and password after the URL itself. For example:

ftp://ftp.mycompany.com/MyLogDirectory/access* Anonymous SomePassword

You may enter multiple distinct file patterns by separating them with a | (pipe) character. For example, if your logs

come from two different sources such as one NT and one UNIX server, you enter something like *D:\NTServer\MyLogs.*|FTP://ftp.mycompany.com/MyLogs/Access* WebMaster FooBar*

If your log filenames contain an embedded date (e.g. *Data-Jun-14.log*), you should use the standard file pattern characters to specify them not a substitution field.   For example, use *Data-*.log*.

If you're using Hit List on a web server, you should not enter the log file name unless you want a specific file loaded into the database.

Tip: You can process log files stored on another machine by either mapping a drive letter to the appropriate network share or by using a UNC path name.   For example, if you want to process files stored on a machine named *WebServer* and from a share named *TheFiles*, you could enter something like *\\WebServer\TheFiles\System32\LogFiles\in*.log* in the file name field.   This would find all Microsoft IIS log files in the directory.   Be sure that you have read access to the network share.

Tip: Hit List processes multiple log files in chronological order based on the file time/date stamp.   File names are irrelevant.   If you are manually using FTP to get your logs from another computer, be certain to copy the oldest files first to guarantee correct processing (Hit List's built-in FTP download system will handle this automatically). Because Hit List normally keeps 'old' data from entering the database, coping the newest files before the oldest ones will cause Hit List to load the newer data first, excluding all requests from the older file.   If you mistakenly FTP your files in the wrong order, you can manually override the old data suppression with the Updates tab of the Options window.

My server logs to an ODBC datasource like Microsoft SQL Server

If you're using Microsoft Internet Information Server, log information could be going into a SQL Server database. In this case, you should select this option.

Tip: If you're using Microsoft IIS and you are unsure where the logs are being stored, choose the **My server logs to a file** option and leave the **Log filename or pattern** field blank; Hit List will find the log information by itself if your machine is the server or you are using *Remote Mode* to run reports on another machine.

Username

Enter the username for Hit List to log on with to access the ODBC data.   This does not have to be your name, simply a valid username for the ODBC datasource.

Password

Enter the password that matches the specified **Username**.   The password is shown by asterisks and is stored in an encrypted format.

Datasource

ODBC works by mapping database information to logical objects called *datasources*.   You must enter the name of the ODBC datasource that you'll be using.

Important: Datasources are defined on individual machines not by servers.   Therefore, it is necessary that your machine has 32-bit ODBC installed and an appropriate datasource established.   Datasources are created via the Control Panel and the ODBC32 applet.   Configuring ODBC is tricky and you should consult your system administrator if you are unsure how to do it.

Table

Enter the name of the database table where the log data is store.   By default, Microsoft Internet Information Server uses a table named *InetLogTable*.

Options/Database

For greater speed and size efficiency, Hit List processes your server log files into its own database.   Reports are run against the database rather than the logs themselves.

Hit List can store its database either in a file (which may or may not be local to this machine) or in an ODBC database such as Microsoft SQL Server.

Tip: In most situations, storing the database as a file provides easier use and better performance.

Tip: If you're hosting multiple virtual servers and using a web server that does not put any indication of which server is answering the request (Netscape, for example) into the logs, it is not a good idea to put all your logs into one database.   Instead, use the Log File and Database override switches in each report to use specific logs and database combinations.

My database is stored in a file

Select this option for Hit List to store its database in a database file.   You do not normally need to specify the name of the database; the application will determine it by itself.   The database will normally be stored in the same folder as the Hit List engine.   By default, that would be in *c:\program files\common files\marketwave* for Windows 95/Windows NT 4.0 and *c:\winnt\marketwave* for Windows NT 3.51.

You may enter the file location of the database to have Hit List use a different database for analysis or to move it to a location with more available space.

Tip: If you allow remote reports and move your database from its original location, be sure to use the Advanced tab to instruct remote reports to generate their reports from the new database location.

My database is in an ODBC database such as Microsoft SQL Server

This option is primarily intended for situations where the database is expected to grow to hundreds of megabytes or where several users are expected to simultaneously create reports.   To use it, you must have established an ODBC datasource that has the privilege to create and drop tables and read, update and delete data.   The first time a database update occurs, Hit List will automatically create the appropriate tables and indices.

Username

Enter the username for Hit List to log on with to access the ODBC data.   This does not have to be your name, simply a valid username for the ODBC datasource.

Password

Enter the password that matches the specified **Username**.   The password is shown by asterisks and is stored in encrypted format.

Datasource

ODBC works by mapping database information to logical objects called *datasources*.   You must enter the name of the ODBC datasource that you'll be using.

Important: Datasources are defined on individual machines not by servers.   Therefore, it is necessary that your machine has 32-bit ODBC installed and an appropriate datasource established.   Datasources are created via the Control Panel and the ODBC32 applet.   Configuring ODBC is tricky and you should consult your system administrator if you are unsure how to do it.

See Also:
Log File Options
Update Options

Options/Updates

The Updates tab determines how the Hit List engine imports data from your logs.

Calculate how long each page was viewed

A unique feature of Hit List allows it to show an estimation of how long visitors viewed your pages.   However, if you don't plan to ever use this, you can speed   database updates somewhat if you turn this off.

Store application arguments (queries)

If your site has a large number of complex query requests and those queries tend to be unique (a site that has a lot of searches, for example), you might be able to considerably reduce the size of your database by unchecking this box.   If unchecked, Hit List will not store the query argument anywhere in the database.   This option should only be used if you never intend to use this information.

Lookup site names (Reverse DNS lookup)

Servers usually record visitor information by storing their IP address not the name of their Internet site.   However, in addition to IP addresses being intelligible only by machines, Hit List requires the site name to run certain reports (*Country*, *Region*, *Organization*, etc.).   If your server is recording IP addresses, you can have Hit List find the matching site names (a so-called Reverse DNS lookup) as it imports data from your log files into the database. Even if you don't turn this option on, each report may be instructed to lookup site names by checking the **Lookup site names** box in the Options tab of the Design dialog.   Nonetheless, it is usually more efficient to check the option here so the lookups occur as the data is imported.

Tip: Before Hit List begins looking up IP addresses, it first attempts to confirm a working DNS by querying for *www.microsoft.com*.   If that fails, it tries *www.netscape.com*.   If both of these fail, Hit List aborts the lookup process.   This can occassionally cause a problem for intranet sites that do not allow external lookups or rely on a firewall and/or proxy servers.   If this applies to your site, you can 'fake' Hit List by adding *www.microsoft.com* to your HOSTS file with any IP address you like.   See your Windows documentation for more information.

Let old data enter the database

Hit List normally goes to great lengths to ensure that duplicate and outdated data does not enter the database. However, there may be cases where you find an old log file that you want to analyze.   Normally, Hit List would prohibit such a file from entering the database.   Check this option to force it to accept the old file.

Important: This option is very dangerous because it could easily allow duplicate log records to enter the database, making the information worthless.   We **very strongly** suggest that you only use this option when absolutely necessary and, when used, only with a specific log file name or pattern to import.   If this option is on while Hit List is doing its normal scan for all log files, duplicate information will enter the database.   There is no easy way to remove duplicate information.   Be certain to turn this option off immediately after loading the old data to ensure that old data is not later accidentally imported.

Assume one server, even if multiple virtual servers appear to exist

Hit List normally calculates Visits by assuming, among other things, that no single visit could span multiple virtual servers.   For large sites that use a rotating server, it might be necessary to check this option if your logs contain the IP of the answering server, making it appear as if virtual servers exist.   If your logs do not contain this information anyway, there is no need to use this switch although Visits might be calculated slightly faster when this is on.

Use cookies when calculating visits

Most logs do not contain cookies so Hit List does not normally use them to help determine distinct visits.   However, if your logs contain either persistent or session cookies, you can check this box to have Hit List use them when calculating visits.   Even with this switch on, however, Hit List will still calculate visits for requests that contain no cookie information; the additional information is used to help Hit List but is not required for accurate visit

calculation.

Maximum request-to-request delay per visit

Hit List estimates unique visits to the site even when visitors are only identified by an IP address. Internet Service Providers (ISPs) generally re-allocate IP addresses as users log on and off their system so IP address alone is not normally enough to determine visits.   Hit List estimates visits by looking at the time between requests for each IP address.   If a particular IP address does not request another URL for longer than a certain period of time, any new request made by that IP is assumed to be another visit.   Use this option to set the period of time (default is 15 minutes).

Update the database will all requests except those

If your web site handles a very large number of requests, you might consider using the options below to keep information that you'll never use out of the database.   This can both reduce your database size (sometimes considerably) and speed database updates.   These should be used sparingly, however, because you will never be able to run a report against information that isn't in the database.   We suggest using these filters only for high-volume sites.   For low or medium-volume sites, you should probably use standard report filters.

With these URLs

You can filter-out certain URLs by entering one or more URL patterns here.   Separate multiple URL patterns (all will be masked-out) with commas.   This option is frequently used to exclude graphics by entering something like **.GIF,*.JPG,*.JPEG*.   Case and blank spaces are irrelevant.   For speed reasons, do not enter more extensions than necessary as each URL must be checked against every member of the list.

This filter could potentially miss a few URLs if the URL does not actually end with these characters.   For example, an image map that appears in the log as *MyMap.GIF?200,300* would be missed by a **.GIF* filter because the URL does not end with *.GIF*.

You can also use a report filter to exclude particular types of objects for a specific report.

From the following IP addresses

You can keep requests from certain visitor IP addresses out of the database by entering them here.   Separate multiple IPs (all will be masked-out) with commas.   Wildcards are allowed.   This option is frequently used to exclude requests from within your own organization where the effects on overall statistics could be significant. Case and blank spaces between IPs are irrelevant but blank spaces between IP dot separators are significant (i.e. 127.0.0.0 is not the same as 127.   0.   0.   0).   For speed reasons, do not enter more IPs than necessary as each request IP must be checked against every member of the list.   If your server logs site names rather than IP addresses, be sure to enter site names instead.

Tip: Use a wildcard to filter-out all requests from your organization.   For example, if your company has IPs like *123.456.789.xxx*, you can exclude your entire organization with *123.456.789**

You can use a report filter to exclude particular visitor IP and/or site names for a specific report.

See Also:
Log File Options
Database Options

Report Design/Outline

Reports are collections that contain one or more *Report elements* plus optional Section dividers and explanatory Text. Each Report element defines a particular report table or graph (or both). All the Report elements (and, therefore, the entire report) share characteristics defined by the Format, Filter, Output and Options tabs.

The Toolbox

The key to efficient report creation is re-using elements instead of creating them again and again. To this end, Hit List uses a Toolbox that acts as a sort of a universal clipboard for Report elements. By default it contains more than 130 pre-defined elements but you can add new ones, remove unwanted ones or even modify the pre-defined elements.

Once an element is added to a report, changing that element in the Toolbox does not affect any copies that have been added to reports.

Report elements are grouped by type of element: *Sections*, *Text*, *Tables*, *Graphs* and *Combined Tables and Graphs*. To see the list of each type, click the + symbol next to the element type name.

Sections

A Section is simply ordinary text that is formatted in a way that makes it easy to differentiate parts of the report. Select a Section by dragging it onto the **Report Outline** in the Design window.

Text

Text is just ordinary text that is frequently used to explain something, tell when the report was run, what dates it covers, etc. Add an existing Text element to a report by dragging it onto the **Report Outline** in the Design window. Text may include <u>substitution fields</u>.

Tables

Tables are text-based report elements that show the results generated from a standard report or crosstab. Most elements are tables. Add an existing Table element to a report by dragging it onto the **Report Outline** in the Design window.

Graphs

Graphs are visual report elements that show the results generated from a standard report or crosstab. Add an existing Graph to a report by dragging it onto the **Report Outline** in the Design window.

Combined Tables and Graphs

These elements contain both textual information (Tables) and graphs. Add an existing Combined element to a report by dragging it onto the **Report Outline** in the Design window.

Report Outline

Reports are designed by dragging elements from the Toolbox into the **Report outline**. Once in the Outline, you can use the **up** and **down** arrow buttons to change the sequence and use the **Properties** button to modify the definition of each element. Elements can also be copied and then modified, making it easy to create new elements that are very similar to existing ones.

Adding New Report Elements

If you want to create a new element that will be used in several reports, you should add it to the Toolbox. Click the **New** button on the Toolbox and select the type of element (the default is determined by which element is currently selected in the Toolbox). Complete the Properties dialog as required. Now drag the new element to the **Report Outline** or click **Add to Report**.

Tip: It's easier to start with an existing element then modify it than create a completely new one. Select the initial element, click **Copy** then click **Properties**.

Tip: If you're sure that you'll never need an element again, you can create one that exists only on the current report by using the **Copy** button on the Design window to clone an existing element. After it is copied, click **Properties** to modify it. If you later decide that you want to use it in other reports, select it with a right-click then choose **Add to Toolbox** from the pop-up menu.

Copying Existing Report Elements

Existing report elements, both on reports and in the Toolbox, can be copied by clicking the **Copy** button.

Modifying Existing Report Elements

Report elements are defined with the Properties dialog. Select the element in question then click the **Properties** button. See **Report Element Definition** for details.

Renaming Existing Report Elements

You can rename an existing report element either by clicking the **Rename** button or changing its name via the General tab of the Properties dialog box.

Deleting Report Elements

Select the element to delete then click the **Delete** button.

Tip: Once deleted, there is no way to restore an element. If you delete one of the built-in report elements from the Toolbox, there is no way to get it back other than re-creating it. However, if you've used the deleted element on a report, you could copy from the report back to the Toolbox.

Time period

Reports must cover a specific time period. Either select a choice from the list or use *Custom* to enter specific *From* and *To* dates.

Update the database before running this report

Check this box to ensure that the report is based on the most recent data available. However, be aware that updating the database could take a long time if the site has received a large number of requests since the last time it was updated.

Run this report for each virtual server (domain)

If you have multiple virtual servers, you can run this report for each of those servers simply by checking this box. Hit List will find all your virtual servers and repeat this report for each of them. Since Hit List normally generates reports as OUTPUT.HTM, you should be sure to set the HTML File Name field with a combination of fixed text and the @Domain@, @DomainIP@, @RootDomain@ or @ShortDomain@ substitution field. If you don't do this, your reports will overwrite each other. See the Output tab for more information.

Tip: You can combine this checkbox with the **Server name or IP** field in the Filter tab to run this report for specific virtual servers. This makes it easy to create report groups. Simply check this box then enter the desired servers in the field. Separate multiple virtual server names or IPs with a comma. Do not use wildcards.

Tip: This feature is especially useful when combined with Scheduled Reports.

See Also:
Updating the Database
Scheduling Reports

Report Design/Format

The Format tab of the Design window lets you make your reports more attractive by using colors, picking fonts and sizes, etc.

To select an item to format, either click on its picture in the preview area or select it from the **Item** list.   Not all design attributes are applicable to all items.

Formatting only applies to HTML pages; plain-text and comma separated text are not affected.

Bold

Make text darker.

Italic

Make text italic.

Size

Click the drop-down list to select the font size.

Note: HTML does not specify font size in points.

Font

Netscape Navigator 3.0 and Microsoft Internet Explorer (version 2.0 or better) allows HTML pages to display with particular fonts.   If you're using IE or Netscape 3.0, you can choose the name of the font to use for the item.

Tip: The font list includes all the fonts installed on your machine.   If you're creating reports that other people might read, you should consider which fonts they have installed.   If the requested font is not available, IE will display using the default font.   Similarly, although Netscape 2.x does not support fonts, it will display a page that uses them with the default font.

Custom color

Reports are normally generated using the default colors that the browser chooses.   For example, backgrounds are usually either white or gray and text is usually black.   Check this box if you want to select custom colors.

Back

If **Custom color** is checked, click the color button to visually select the item's background color.   Alternately, if you know the RGB value of the color, you can enter the Hex value into the box to the right of **Back**.

Text

If **Custom color** is checked, click the color button to visually select the item's text color.   Alternately, if you know the RGB value for the color you want, you can enter the Hex value into the box to the right of **Text**.

Alignment

Most elements may be aligned to the left, right or center.   Auto aligns the item in the most common way for that item.

Graph back color

Graphs may be displayed in any of 16 different background colors.   Select the color by clicking the list.

Show table borders

Tables may be shown with or without borders.   Check this box to display borders.

Tip: Combining borderless tables with cell colors can be very attractive;   however, tables are normally hard to read if they are borderless and use default colors.   If Netscape 2.x will be your primary browser, you should probably not use borderless tables because it does not support cell coloring.

Report Design/Filter

You can use the Filter tab of the Design window to simplify your reports by filtering out information that isn't important.   This will frequently make your reports easier to understand and might show trends that would have been obscured otherwise.

You can also override the report-level filter for any report element with the <u>Filter</u> tab of the Properties window for that report element.   This makes it easy to create reports that 'zoom-in' on multiple criteria.

Tip: Be careful when mixing report-level filters and element-level filters because element-level filtering overrides all report-level filters, effectively nullifying them.   If you have a report with elements that make use of element-level filtering (the built-in *Total Number of HTML Pages* element, for example), be sure to mirror any report-level filters on each element that overrides the report-level filtering.   For example, if you set a report-level **Server name or IP** filter to *123.456.789.123*, you must copy this setting into all elements that override the report-level filter.

Include these objects in the report

Hit List looks at each URL to determine its <u>Object Type</u>.   Objects may be categorized as Pages,   Graphics, Audio, Video, Applications, VB/Java applets, Downloads, Errors or Other.   Normally all of these objects will be used when running reports.   However, you can leave any of these out by unchecking the appropriate check box.

Object Types are determined by URL extensions.   For example, Pages normally include .HTM and .HTML URLs and Graphics include .GIF, .JPG and other common graphic extensions.   You can change how extensions are mapped to Object Types with the Object Types tab in the Database Manager.

Tip: You may find that defining your own <u>URL Groups</u> produces more useful information than relying on the built-in Object Type organization.   From the Groups tab of the Database Manager, you can map specific URLs and URL patterns to your own Groups.   For example, you could create a Group called *Marketing* that includes */products/info.htm*, */products/pricing.htm* and */downloads/demos/product3.zip*.   You could compare how well *Marketing* did relative to other Groups.   Unlike Object Types, URLs can be members of multiple Groups.

Tip: Certain types of reports, such as *Most Popular URLs*, work much better when you exclude Graphics. However, each Object Type excluded slows report generation so you shouldn't exclude objects that seldom or never appear anyway.

Tip: If you always want to exclude certain type of objects (graphics, for example), you could also filter them out before they enter the database by using the filter options in the <u>Updates</u> tab of the Options dialog box.   This is more efficient if you want to filter-out a high percentage of your total requests and will usually considerably reduce the database size and increase update speed.   However, it could slow database updates if you want to filter-out many types of objects or those objects are seldom requested from your server.   In either case, you're better off filtering here.

Include/Exclude information unless it comes from specific places/time

You can also simplify reports by limiting them to data that is related to particular times or places.   This is especially useful for large sites where any one person may only care about how her Group or Directory did.   Alternately, you can create reports that use all available information except when it came from particular places or times.   This is useful for masking-out requests from your own organization.

Select *Exclude*   to limit your report to the particular places or times noted below.   Select *Include* to base the report on all requests except the particular places and times listed below.

For all of the below, you may specify multiple places or times by separating values with commas.   Reports will include information that comes from any of those values (logical OR).

Leaving a field blank means ignore it; it will not be used to filter information.   Entering any information means to

include that value in the filter.   For example, entering *Demos/** in the **Directories** field restricts data to the */Demos/** directory and all subdirectories; requests from all other directories will be excluded (assuming *Exclude* is selected).   The filter will be even more restrictive if you enter information into any other field.

Tip: You can use wildcard characters (* and ?) to broaden the selection criteria in most of the fields below.   Use * to match multiple characters and use ? to match any one character.   For example, **foo** would find everything that includes the word *foo* anywhere while *foo?* would find only words that begin with *foo* and include exactly one more character.

Directories

Enter one or more directory or directory pattern.   For example, */Product1/* would limit the report to URLs that are in the */Product1/* directory.   URLs located in subdirectories of */Product1/* would not be included unless you specifically enumerate them or specify a directory pattern like /Product1/*.

Tip: Reports generate much faster when you don't use patterns.

URLs

Similar to **Directories**, enter one or more URL or URL pattern.   Only those specific URLs or URLs that match the pattern will be included in the report.

Web server name or IP

If your server responds to multiple IP addresses or names (for example, *www.mycompany.com* and *www2.mycompany.com*), you might want to restrict reports to relating to one server.   Enter the names or IP addresses to include.

Tip: For security reasons, reports created using ad-hoc forms are implicitly restricted to the same server name or IP address that generated the request.   This means that an ISP could allow each of her clients to use an ad-hoc report form to generate reports without any possibility of them seeing information about other clients.   You may override this and allow more flexibility; see the Advanced tab in the Options dialog for details.

Visitor IPs

You can filter by the IP addresses of your visitors by entering IP addresses here.   This can be used to either focus an entire report on the activity generated by one IP (if the *Exclude* option is used) or to filter-out requests from certain IPs such as your own organization (if the *Include* option is used).   Hit List will filter based on the IP address that was stored in your logs.   If your server does a reverse DNS lookup to turn IPs into site names, be sure to enter site names instead of IPs.

Tip: If you always want to exclude certain IPs (from your own organization, for example), you could also filter them out before they enter the database by using the filter options in the Updates tab of the Options dialog box. This is more efficient if you want to filter-out a high percentage of your total requests.   However, it could slow database updates if you want to filter-out a large number of IPs or those IPs seldom request information from your server.   In either case, you're better off filtering here.

Visitor site names

Filtering by Visitor site names is essentially the same as filtering by **Visitor IPs** except that the filter is based on the site names that Hit List has previously resolved from IP addresses.   Site names can be resolved when running the report with the Options tab of the Design window or when the database is updated by using the Updates tab of the Options dialog box.

Tip: If your server logs site names not IPs, filter by **Visitor IPs** not by **Visitor site names**.

User names

Enter the name or names of users to include in the report.   Logs only contain user name information when users

must enter user names/passwords to access certain URLs.   If you have no password protected areas, you will have no user names.

Tip: User names are not the same as email names.   Logs do not generally contain email name information.

Entry sources

Visitors generally enter your site by typing a URL directly, selecting a bookmark/favorite from their browser or following an ad or a link from another web site.   This filter lets you use this information to see how visitors who entered your site from different sources (i.e. *Entry sources*) vary in their use of your web site.   For example, you may find that you get more visitors coming from one site but the average visit length or the average number of requests per visit is higher from another site.   In this way, you can rate the quality as well as the quantity of your visitors.   This feature is especially useful when used at the element-level so as to allow one report to compare various coming from different sites (Yahoo vs. Excite vs. InfoSeek, etc.).   The built-in *How Our Advertising is Doing* report makes extensive use of this filter.

Sources are fully-qualified URL paths (e.g. *http://www.yahoo.com/Business_and_Economy/Marketing/*) so you should usually use wildcards to specify an Entry source.   For example, the *How Our Advertising is Doing* report uses a filter of **yahoo.com** to specify Yahoo.

Tip: Using filters that include a *www* prefix isn't a good idea because many sites (Yahoo, for example) use different servers but are all really from the same place (e.g. *www2.yahoo.com* and *search.yahoo.com*).

Entry pages

Similar to Entry sources, it's often useful to ask different groups of people to enter your site at different places. Often marketing departments use fictitious virtual roots to point to the same pages.   Since these virtual roots are recorded differently in your logs, it's possible to track such marketing experiments.   For example, you could send email to 500 potential customers asking them to visit your web site at *http://www.yourcompany.com/Test1* and email to another 500 asking them to go to */Test2*.   Both */Test1* and */Test2* are virtual directories that actually just point to your regular home page.   But, as they are recorded differently in your logs, you can filter on these Entry pages to compare the quality of the visits from each group much as you would use an Entry source filter.

Entry pages, like other URLs, are not usually fully-qualified so do not include the HTTP protocol prefix or the name of your web pages.   In the examples above, the entry pages would look like */Test1* and */Test2*.

URL groups

If you've used the Database Manager to define URL Groups, you might find it very useful to 'zoom in' on certain Groups by excluding information about other Groups.

Tip: If you know that each URL is a member of no more than one Group, check the **Assume each URL is a member of no more than group** checkbox in the Options tab of the Design window to speed report generation.

Ad groups

This filter lets you zoom-in to advertising at a very high level.   For example, if you've created five Ad groups (e.g. *Sports, Entertainment, Business, Family and News*), each of which contains several ads, you can constrict the report to one or more of these groups by entering the appropriate name or names.

Tip: You must have previously used the Ads tab of the Database Manager to tell Hit List how to recognize ads running on your site.

Ad names

The Ad names filter is similar to the **Ad groups** filter but more specific because it looks for individual ads not top-level Ad groups. For example, if you've created 25 specific ads, spanning several Ad groups, you can constrict the report to one or more of these ads by entering the appropriate name or names.

Tip: You must have previously used the <u>Ads</u> tab of the Database Manager to tell Hit List how to recognize ads running on your site.

Days of the Week

Similar to **Hours**, you might find that usage varies greatly when comparing different days (weekdays vs. weekends, for example)   Enter full names nor abbreviations of the days of the week to include.

Hours

You may want to limit the report to certain hours, especially when determining average statistics where very low usage periods such as 3AM would skew the results.   Enter hours using a 24 hour clock where 0 means midnight and 23 means 11PM.   Select a range of hours such as 9AM to 5PM by entering each individual hour.

Realms

If your site is using user name security, you may have established security realms.   You may enter the names of the realm or realms to include.

Parameter/Value Pairs

Hit List Professional includes extremely sophisticated application argument (query) and source (referrer) parsing that makes it easy to create reports for sites that are highly-database driven (for example, sites driven by Cold Fusion).   If you've used the <u>Parsing</u> tab of the Options dialog to configure parsing for your site, you can use this filter to restrict the resulting reports.

This filter normally applies to application argument parameters; however, when used during a report element based on source parameters, the filtering applies to the source parameters.   For example, the built-in *How Our Advertising is Doing* report uses this filter, at the <u>element-level</u>, to show which keywords your visitors are using to find your site.

Parameter/Value filters allow more complex expressions than the other filters.   You may specify parameters in a number of different ways, depending on what you want to do.   For example, if you're interested in a parameter named *City* and possible values are *Seattle*, *SanFran* and *Boston*:

City=Seattle

Would find all requests that contain the *City* parameter set to *Seattle*

City=SanFran,City=Seattle

Would select all requests regarding either Seattle or San Francisco.

City<>Seattle

Selects all requests that are not related to Seattle.

Specifying a parameter without a value causes Hit List to select all requests that have that parameter, regardless of the value.   Therefore,

City

Selects all requests that have the *City* parameter, regardless of the value (if any).

Tip: Set the **Include/Exclude** toggle to *Include* to negate the above.   For example, if *Include* is specified:

City=Seattle

Would find all requests that do not contain Seattle.   This would be the same as using *City <> Seattle* and the *Exclude* setting.

City=SanFran,City=Seattle

Would select all requests about any city other than Seattle and San Francisco.   Note that *City<>Seattle,City<>SanFran* and *Exclude* is not the same thing because the latter would be interpreted as City <> Seattle or City <> SanFran, which would actually select all records.

City<>Seattle

Selects all requests that relate to Seattle.   This is the same as *City=Seattle* with the *Exclude* setting.

City

Selects all requests that do not contain the City parameter.

You may combine multiple parameter/value pairs in one line.   For example, if *Exclude* is set and you enter

City=Seattle,Music=Alternative

all requests with either *City=Seattle* or *Music=Alternative* would be selected.

You may use wildcards to specify either parameter values or parameter names when no value is specified.   For example, *Template=/Cities/*, Template<>/Cities/** and *Temp** are allowed but *Templ*=/Cities/** and *Templ*=/Cities* are not.

Report Design/Output

Hit List lets you create reports in the form of HTML pages, plain ASCII text files, Microsoft Word documents and comma-separated text files (for importing into spreadsheets).   Each report must include at least one type of output but can include any two or even all three.

HTML page

Most reports are created as an HTML page for use in a web browser.   Although some of the options in the Format tab of the Design window apply to specific web browsers, reports are generally viewable by any web browser that supports tables.   This includes Netscape Navigator 1.2 or better and Microsoft Internet Explorer 2.0 or better.

You must check **HTML page** to generate graphs.

File name

You may specify the file name for the HTML file.   The default file name of *output.htm* will be used if you do not specify a name.   The file name is automatically generated by your server if you're using Remote Mode (reports generated on the server and sent to your browser) but you can override it.

The file name can be any valid path and name combination using the long file name system.   You may embed the date as part of the file name by enclosing a date format pattern with @ symbols.   For example, a file name of *c:\ data\report@mm-dd-yy@.htm* would expand to something like *c:\data\report12-31-96.htm*.   Remember not to use the slash character (/) as a date delimiter because the file system will interpret it as a path separator.   If the required directory does not exist, Hit List will create it if possible.

Tip: HTML files and their associated GIFs can be automatically posted to an FTP site by using a FTP URL instead of a normal file name.   Hit List will first create the required files on your computer then, when the report is done, send all the necessary files to your FTP server.   Be certain that the specified FTP directory exists or the files may end up in the wrong place.   Example:

ftp://YourServerName/YourDirectory/YourFile.HTML username password

You may also use the *@Domain@*, *@ShortDomain@, @RootDomain@* and *@DomainIP@* substitution fields in the file name or path.   This is almost essential when using the Run this report for each virtual domain option to keep the multiple output files from overwriting each other.   For example, you could store each client's report in its own directory with an file name such as *D:\MyServer\@Domain@\Report@yy-mm-dd@.htm*

Template

Hit List templates allow you to produce highly customized reports including your company logo, other graphics, or even sound and video.   A template is a standard HTML file with any text, graphics and tags you wish plus comment codes that instruct the Hit List engine to embed report elements in specific places.   When the report is generated, Hit List reads the template file and combines it with the specified report elements.

Report elements are inserted where the element Name or Display as field matches the comment.   For example, if a you want to embed a report element called *Total Requests*, include *<!-- Total Requests -->* in the HTML template. Spaces are significant (there is one space between the -- and the *T* and one between the *s* and the --).   Capitalization is irrelevant.   Report elements with the same display name will be embedded in the order that they are encountered in the HTML template.   Report elements not found in the template will be discarded.

Tip: You may add a header and/or footer to your report without naming each element by using the special *<!-- All -->* code.   The entire report, except the standard Hit List header/footer, will be substituted in its place.

Text file suitable for importing into a spreadsheet (CSV)

You may use Hit List to generate a series of comma separated text files that are easy to import into a spreadsheet.

This is useful if you need to do unusual analysis or create more sophisticated graphs than Hit List produces (overlay charts, for example).   Each report element will become its own file.

Base name

Each report element becomes its own file that includes a common base name plus the name that would normally be displayed on an HTML page.   For example, a base name of *c:\data\report* and a report element named *Total Requests* would produce *c:\data\reportTotal Requests.txt*.

Tip: Avoid using report elements with the same display name when using CSV output because the first element with that name will be overwritten by subsequent elements with the same display name.

Tip: You may use a substitution field to add today's date to any base name.

Tip: A CSV file will not be produced if the combination of the base name and the display name produces an invalid file name.   No warning will be given.

Text file

Reports may be output as plain ASCII text without any HTML codes or Microsoft Word documents.   This is useful for copying or sending reports to people who don't have a web browser.   ASCII text files will not include any graphs but Word documents will.

File name

The text file name is specified exactly as the HTML file name except that no default is provided.

Template

Like HTML files, text files may be combined with templates.   Text templates are defined exactly the same as HTML templates.

Tip: Be sure that all comment codes match report elements or the final report will include the unmatched comment codes.

Send as email message

Text output may be sent as an email message.   This is especially useful when combined with Scheduled Reports. In this way, an Internet Service Provider with multiple clients could send daily reports to each of them or an MIS department could route different daily reports to appropriate managers.

Reports will be sent by either Microsoft Exchange (MAPI) or by POP/SMTP mail as specified in the General tab of the Options dialog.

Important: To send email via MAPI, your email client (usually Exchange or MS Mail) must already be running.

To

Enter the recipient list as required by your email system. MAPI email names must be separated by semi-colons and POP/SMTP must be comma-separated.   You may use the same substitution fields that are allowed for filenames.

Tip: For reports that iterate over all your virtual domains, you can send email to each clients with the @RootDomain@ substitution field.   For example, *Webmaster@@RootDomain@* or *@RootDomain@@YourCompany.com*.

Subject

You should enter a subject for the email.   The subject may include the same date format and domain substitution fields that **File name** uses.

Report Design/Options

The Options tab sets assorted properties that affect the way the report is created.

Report description

Enter any text that you want to appear when this report is selected in the Main window. Unlike all other descriptions, this text will never appear on the report.

Tip: Use a Text report element to add explanatory text to the report if necessary.

Lookup site name (reverse DNS lookup)

Some web servers can resolve IP addresses into names when they log requests. However, doing so is usually not recommended because the lookup can increase critical response time. Nonetheless, converting IP addresses into human-understandable names not only increases readability but also makes it possible to determine which country a request comes from and they type of organization the user belongs to. Therefore, Hit List lets you resolves IP addresses into site names either when reports are run or when the data is <u>imported</u> into the Hit List database.

If IP addresses were not converted into site names when the data was imported, you can check this box to do it when the report is run. Once a site name is resolved from an IP, it will never be looked up again. If a name cannot be resolved after five attempts, Hit List will give up and never attempt to look it up again.

Tip: IP addresses sometimes change so it is recommended that you lookup site names when importing logs rather than waiting for a report to be generated.

Tip: Before Hit List begins looking up IP addresses, it first attempts to confirm a working DNS by querying for *www.microsoft.com*. If that fails, it tries *www.netscape.com*. If both of these fail, Hit List aborts the lookup process. This can occassionally cause a problem for intranet sites that do not allow external lookups or rely on a firewall and/or proxy servers. If this applies to your site, you can 'fake' Hit List by adding *www.microsoft.com* to your HOSTS file with any IP address you like. See your Windows documentation for more information.

Use site name and application argument groups, if defined

If you've used the Database Manager to define either <u>site name groups</u> (for example, turn all site names that end in *.AOL.COM* into one *AOL* group) and/or <u>application argument (query) groups</u>, Hit List will use these in place of the original values where applicable. This will ususally make your reports easier to understand. However, if you want to see exact values for these, you may turn this option off. Reports will also run somewhat faster with this off.

Shows descriptions on report

Report elements may contain a <u>Description</u> that explains the element. You can include these on the report by checking this box.

Assume each URL is a member of no more than one Group

If you are using <u>Groups</u> to categorize URLs, it is possible for a URL or URL pattern to be included in multiple Groups. Processing such a report may take a long time because the number of records analyzed could be dramatically increased. Using URL patterns will take an especially long time because Hit List must do much more work to match a pattern than a static URL. If you are certain (or want to pretend) that no URL is a member of more than one Group, you can check this box to dramatically speed report processing.

Delete old GIF files (more than 1 day old)

Each time Hit List generates a graph, a GIF graphics file is produced. Over time, these GIFs could take up a substantial amount of disk space. If this box is checked, Hit List will delete GIF files that are more than one day old. Only GIFs created by Hit List will be affected.

Tip: Be sure to turn this option off if you want to keep a report archive.

Default domain

Hit List turns the URLs in your log file into friendly page titles by combining the information in your logs (which do not mention which server they relate to) with the name of the default web server specified in the main Options dialog box selected from the Tools menu.   If you need to override this global setting for a specific report (a different virtual server, for example), simply enter the new server name here.   For example, if your global setting is *www.mycompany.com* but the logs accessed for this report relate to *www.yourcompany.com*, just enter *www.yourcompany.com* in this field.

Tip: This setting is ignored if Run this report for each virtual server is checked on a report.   In that case, Hit List will determine the correct value itself.

Tip: If necessary, Hit List can resolve page titles by going through a proxy server.   See the Advanced tab of the Options dialog box for details.

When listing HTML pages, show

When listing HTML pages (Most Popular Pages, etc.), Hit List can list the page as just its URL, as just its page title or with both the page title and the URL.   In cases where a title lookup is requested but cannot be accomplished (page has no title, page cannot be accessed from your machine, proxy not specified, etc.), Hit List will display just the URL.

Graphics folder

Hit List automatically determines where to place GIF files.   However, you might want to specify a particular folder. Enter the fully-qualified path of the folder or use the **Browse** button to select one.

Tip: You should usually leave this field blank.   For reports generated on a local machine, graphs will automatically go in the same directory as their parent HTML file; you don't need to use this feature to simply put your output in a particular directory.

Tip: The graphics folder will be relative to the machine that creates the graphs.   Therefore, if you are using *Remote Mode*, be sure to either use drive letter assignments known to the server or (recommended) use UNC path names such as *\\GraphicsServer\Myreports* instead of drive letters.

Report Element General Properties

All report elements must have a unique name. They may also include alternate text to use when displaying that element and a description that can appear on the report.

Name

Report elements must have names that are unique within the report (if modifying a report element in a report) or within all the elements of the Toolbox (if modifying a report element in the Toolbox). This name will normally appear in reports. A name may not include <u>substitution fields</u>.

Display as

If you would rather have something besides the **Name** appear in the report, you may enter that text here. This may include using substitution fields to add information like the date, the report range, the report sections and so on. The element **Name** will be used if **Display as** is left blank. **Display as** fields do not need to be unique.

Tip: If you have a graph and a table producing essentially the same results, you may want to add *(Graph)* to the front of the **Name** of the graph element. You can use the **Display as** field to title the graph without *(Graph)*.

Description

Report elements may include descriptive text that helps you remember what the element does or explains the element to someone reading the report.

Descriptions are added to the report if **Show descriptions on report** in the <u>Options</u> tab of the Design window is checked.

Report elements that produce a single number (*Total Requests*, for example) will never include the **Description** unless **Always display as a table** in the <u>Definition</u> tab of the Properties dialog is checked.

Report Element Definition

Report elements are the heart of Hit List because reports are essentially just collections of report elements. Each elements may be completely unique or based on a standard report but with unique grouping, sorting or other properties. Over time, you'll probably find that you can produce almost any report by varying these options.

Type

Report elements can produce variation on the built-in standard reports or can be completely unique by using the crosstab mechanism.

Standard reports

There are more than 20 standard reports covering the gamut from simply counting the number of requests to complex path analysis. Select a standard report from the drop-down list box.

Substitute group name for URL

Standard reports frequently include references to URLs. Any such report element can be run using Groups in place or URLs by checking this box. Groups must have been previously defined in the Groups tab of the Database Manager.

Display site names not URLs

When used in conjunction with a List of sources report, this switch determines whether the full source URL or just the name of the source site is shown.

Categorize

This option is only visible when using an advertising report (Ad impressions, Ad click-throughs or Ad impressions and click-throughs). It determines what level of detail is shown and how it is organized. The nine options let you group the report by most combinations of ad groups, ad names or source HTML pages. Often changing this setting gives you new insight into how the ads are working. For example, categorizing first by ad name then by source HTML page can show well each particular ad is doing when run on different HTML pages. On the other hand, categorizing first by source HTML page then ad name shows which ads are doing best on each page.

Crosstab

If you want information that is not already shown with a standard report, you can often get it by creating a crosstab of two columns in the Hit List database. A crosstab is a table that shows all the unique values in one column going down and all the unique values in the other column going across. The intersection of the two columns is a Calculation. For example, if you want to know how the choice of operating systems affects choice of web browsers, you can crosstab Browser by Operating System. You could check **Requests** in the Calculate section to show the total number of requests received for each combination or check **Visits** to show the total number of Visits for each combination.

A crosstab must have exactly one Calculation. Crosstabs cannot be sorted.

Show

Report results can be displayed as text (tables), as a graph or both.

Tip: Although you could use two different elements to show both a graph and a table of the same information, the report will be generated faster if you use one element and check both **Display results as text** and **Display results as a graph** because Hit List does not have to process the same information twice.

Display results as text

Check this box to show the results as text. The text will be a table if the results include more than one number. If

the result is a single number, Hit List will normally display it without a table, making it possible to create a series of simple summary report lines.

Always display as a table

Check this box if you want single-number results to be shown in a table.

Tip: This is useful if you want to include single-result information between tables.   Without forcing a table, such a result might get visually 'lost'.

Display results as a graph

Hit List can show results using any one of 10 built-in 2D and 3D graphs.   Click the type of graph you want to produce.   Single-result information cannot be graphed.   By default, graphs are based on a maximum of seven data points to ensure readability.   Use the *Limit results* area of the Options tab to show more or fewer points.

Tip: If graphs do not appear when using Remote Mode, check with your system administrator to ensure that the application directory (usually *Scripts*, *cgi-bin* or *cgi-win*) allows Read access as well as Execute access.

Calculate

Report elements can generally include either one calculation or several.   Crosstabs can only include one calculation.

Tip: Although you could use multiple report elements to show different calculations, it's generally more space-efficient and easier to understand if you use one element with multiple calculations.

Total requests

Total requests, frequently referred to as *hits*, is the most commonly quoted web usage statistic.   Unless you're using the Filter tab in the Design window to exclude certain Object Types, a request could be anything including HTML pages and all graphic elements that pages request.   The way that pages are authored, perhaps with many small graphics versus with one large one, can greatly affect the number of total requests without indicating anything about the popularity of the site.   For this reason, we suggest using the **Visit** calculation as the primary measure of popularity instead of **Total requests**.

Avg requests

If the report element is grouped by something other than *Totals* (*Weekly*, *Hourly*, etc.) you can display the average number of requests received per unit of time.   If **Group by** is set to *Totals*, **Avg requests** is, by definition, the same as **Total requests**.

If **Group by** is set to anything other than *Totals*, the value that **Avg requests** calculates is also determined by how **Time consolidation** in the Options tab of the Properties dialog is set.   If **Time consolidation** is set to *All* (the default), **Avg requests** will calculate one number, the number of requests typically received during the **Group by** time period..   If **Time consolidation** is set to *Similar times*, the results will be a table showing the average number of requests in each of the time periods produced by **Group by**.   For example, if **Group by** is *Hourly*, **Average requests** will produce a different average for hour in the day.   This is frequently a very useful statistic because it shows usage as a function of time.   Finally, if **Time consolidation** is set to *No consolidation*, **Avg requests** will produce a different averages for each *distinct* time period implied by **Group by**.   In the case of **Group by** being set to *Hourly*, this will produce one average for every individual hour that the server has been running, perhaps thousands of rows.   For this reason, setting **Time consolidation** to *No consolidation* is not recommended unless you've used **Time period** to specify a very brief report time period.

Percent of requests

Check this to see what percent of the total number of requests each row represents.   If **Group by** is set to anything other than *Totals*, the value will be calculated as a percent of each time group not the total amount.   For example, if the report element produces four rows, each with the same number of requests, **Percent of requests** would show 25% for each row if you're grouping by *Totals*.   If, however, those four rows are part of two distinct time groups

(two different weeks, for example), each row would represent 50% of the total for each week.

Standard reports that always produce groups, such as *Jumps within site*, *Path to pages* and *Previous pages*, are also totaled by group.

Visits

A *visit* is a collection of requests from a single IP address where each request came within a certain time of the previous request (usually 15 minutes).   In this way, Hit List estimates how many discrete visits the site had.   Visits generally produce more useful information than **Total requests** although both are useful.   You can set the time limit with the Updates tab of the Options dialog box.

Tip: Visits take longer to calculate than **Total requests**.   Although Visits is generally more useful, you may want to use reserve it for places where it really matters if report generation time is important.

Avg visits

Average visits is the same as **Avg requests** except based on visits not requests.   See **Avg requests** (above) for more information.

Avg time viewed

Use this to see an estimation of how long each HTML page was viewed.   This is only an estimation because it does not include the viewing time of the final page seen in a visit and cannot take web browser caching into account. With most web browsers, clicking the Back button does not send any message to the server; it simply re-loads the page from a disk cache.   Since your server wasn't notified of the request, it didn't appear in the log and, therefore, will not appear in the Hit List database.

This calculation is not available if **Calculate how long each page was viewed** in the Updates tab of the Options dialog is turned off.

Bytes

This calculation shows the number of bytes sent to web browser from the server.   Most web browsers and servers support a feature where the server is told to only send the information if it is newer than the data already cached by the browser.   In that event, the number of bytes sent is 0.   Therefore, you cannot assume that each request for a particular URL will produce the same number of bytes.

Avg delay

The Microsoft Internet Information Server and O'Reilly WebSite (using the WinLog format) log how many seconds elapsed between when they received a request and when the data began flowing to the browser.   This can be extremely useful for pinpointing server bottlenecks.   If you're using either of these servers, you can calculate this statistic by checking the box.   If you're using another server, calculating this will always produce *(None)*.

Group by

This determines how information is organized relative to time.   The **Time consolidation** setting in the Options tab of the Properties dialog also affects time organization.   If **Time consolidation** is set to *All* (the default), **Group by** has no effect except how average statistics are calculated.   When you want to see distinct groups for different time periods, you should be sure that **Time consolidation** is set to *Similar times* or *No consolidation*.

Totals

Information is combined without regard to time.   Results will be based on analyzing all the data in the **Time period** as a whole.   This is most useful for general statistics such as *Total Requests*, *Most Popular URLs*, etc.

Monthly

The month that each request occurred is used to organize the information.   Unless **Time consolidation** is set to *All*,

each month in the **Time period** range will become its own group.

Weekly

The week that each request occurred is used to organize the information.   Unless **Time consolidation** is set to *All*, each week in the **Time period** range will become its own group.   If **Time consolidation** is *Similar times*, there will be a maximum of 52 distinct groups even if the **Time period** range is greater than one year.   Setting **Time consolidation** to *No consolidation* produces a group for each week in the period.

Day of month

The day of the month that each request occurred is used to organize the information.   Unless **Time consolidation** is set to *All*, each day in the **Time period** range will become its own group.   If **Time consolidation** is *Similar times*, there will be a maximum of 31 distinct groups even if the **Time period** range is greater than one month.   Setting **Time consolidation** to *No consolidation* produces a group for each day in the period.

Day of week

Day of week is a little different than the other **Group by** settings.   If **Time consolidation** is set to *All*, *Day of week* produces exactly the same results as *Day of month*.   If **Time consolidation** is *Similar times*, there will be up to seven distinct groups even if the **Time period** range is greater than one week.   Setting **Time consolidation** to *No consolidation* produces a group for each day in the period, exactly the same as *Day of month* would.

Hourly

The hour that each request occurred (using a 24 hour clock where 0 is midnight and 23 is 11PM) is used to organize the information.   Unless **Time consolidation** is set to *All*, each hour in the **Time period** range will become its own group.   If **Time consolidation** is *Similar times*, there will be a maximum of 24 distinct groups even if the **Time period** range is greater than one day.   Setting **Time consolidation** to *No consolidation* produces a group for each hour in the period, potentially creating thousands of rows.   Hourly grouping is most useful when analyzing a very brief period of time.

Domain

Grouping by Domain makes it easy to organize data on a virtual server basis.   Of course, this is only relevant if you have virtual servers and this information is in your logs.   **Time consolidation** does not affect **Domain** grouping; **Domain** grouping is always calculated as if **Time consolidation** were set to *No consolidation*.

Visit

Visit is a very special type of grouping that does not work like the other groupings.   It is designed simply for determining average statistics per visit.   For example, you can determine the average number of requests per visit or the average time viewed per visit with it.   **Time consolidation** does not affect **Visit** grouping; **Visit** grouping is always calculated as if **Time consolidation** were set to *All*.   There is no way to produce a report element that shows data on a visit by visit basis.

Sort by

If the report element produces a list of items, you can choose the way that the list is sorted.

Alphabetical/Chronological

Result lists will be sorted such that items coming first in the alphabet come before those that come later and earlier time periods come before later ones.   If **Limit results** in the Options tab of the Properties dialog is set to *Bottom* instead of the default *Top*, the sort order will be reversed.

Requests

Result lists will be sorted such that items with the most requests will be listed first. If **Limit results** in the Options tab of the Properties dialog is set to *Bottom* instead of the default *Top*, the sort order will be reversed.   Results may be sorted by *Requests* even if no request-based statistic is shown.

Visits

Result lists will be sorted such that items with the most visits will be listed first. If **Limit results** in the Options tab of the Properties dialog is set to *Bottom* instead of the default *Top*, the sort order will be reversed.   Results may be sorted by *Visits* even if neither **Visits** nor **Avg Visits** is calculated.

Tip: Sorting by *Visits* takes much more time than sorting by another other field.   If you're concerned about report generation time, refrain from sorting by *Visits* unless necessary.   It can also be confusing for those reading the report to see report elements that show both **Total requests** and **Visits** but is sorted by *Visits* because the **Total requests** calculation is always shown before **Visits**.

Bytes

Result lists will be sorted such that items with the most bytes sent will be listed first. If **Limit results** in the Options tab of the Properties dialog is set to *Bottom* instead of the default *Top*, the sort order will be reversed.   Results may be sorted by *Bytes* even if **Bytes** is not calculated.

Delay

Result lists will be sorted such that items with the longest delays will be listed first. If **Limit results** in the Options tab of the Properties dialog is set to *Bottom* instead of the default *Top*, the sort order will be reversed.   Results may be sorted by *Delay* even if **Delay** is not calculated.

Time viewed

Result lists will be sorted such that items viewed the longest will be listed first. If **Limit results** in the Options tab of the Properties dialog is set to *Bottom* instead of the default *Top*, the sort order will be reversed.   Results may be sorted by *Time viewed* even if **Avg time viewed** is not calculated.

Tip: Only HTML pages have time viewed information.   For best results, you should consider using the Filter tab of the Design window to mask out all objects other than Pages if you want to sort by *Time viewed*.

Impressions (Ad impressions and Ad impressions and click-through reports only)

Results will be sorted such that ads or ad groups with the most impressions will be listed first.

Click-Throughs (Ad click-throughs and Ad impressions and click-through reports only)

Results will be sorted such that ads or ad groups with the highest number of click-throughs will be listed first.

Click-Through Rates (Ad impressions and click-through reports only)

Results will be sorted such that ads or ad groups with the highest click-through rates (defined as click-throughs divided by impressions) will be listed first.

Tip: Note that it is possible, however unlikely, for ads to have a click-through rate of more than 100% due to browser caching.

See Also:
Standard Reports
Update Options
Time consolidation
Filters

Report Element Options

The Options tab sets advanced properties that affect the way the report element is generated.

Limit results

For report elements that produce lists (*Most Popular URLs*, *Most Common Exit Pages*, etc.), you may want to limit the number of list items so the report is easier to understand.   You may limit the number of items shown by setting a combination of the three options below.

When **Group by** is set to something other than *Totals*, **Limit results** applies to the number of items listed in each group not to the total number of items.

Important: Limits are dependent on how **Sort by** is set in the <u>Definition</u> tab of the Properties window.

Top vs. Bottom

Generally you want to list the items that have the most requests, the most visits or are listed first alphabetically.   This is the *Top* setting.   However, if you want to see the least popular items, select *Bottom*.

Number of items

Set the number of items by entering a number in the box to the right of the **Top vs. Bottom** list box.   Leave it blank to produce an unlimited number of list items for tables and a default of seven items for graphs.

Tip: Graphs are easiest to understand when the number of items is relatively low, usually under 10.   Therefore, if **Number of items** is left blank (hence producing an unlimited number of items), graphs based on those results will automatically be restricted to the first seven items (seven for Days of the week).   If you need more graph points (24 should be used when graphing Hours, for example), be sure to set the **Number of items**.

Items vs. Percent

You may limit the results to an absolute number of items (the *Items* setting) or to a percent of the number of items in the group (*Percent*).

Minimum requests required to show

Result lists may also be limited by excluding objects that failed to have a minimum number of requests.

Tip: Combining **Limit results** with **Minimum requests required to show** is allowed but could easily produce a list with no items.

Ignore requests from this Source site

Reports that show REFERRER information such as the *List of Sources* standard report or a crosstab involving *Source_Site* or *Source_URL* may give more useful information if you filter-out any REFERRER strings that come from your own site. Enter the name of your site (e.g. *www.mycompany.com*) in this field to filter-out such requests.   Wildcards are allowed.   Separate multiple site names with a comma to exclude more than one source site   (for example, *www.mycompany.com, www.yourcompany.com*).   This option is only available for the reports listed above.

Tip: Web servers sometimes log the REFERRER site name as an IP address rather than a name.   To ensure a complete filtering, enter both your site name and IP address.   For example: *www.mycompany.com,123.456.789. 1*.

Time consolidation

Time consolidation is an extremely important part of a report element's definition.   Generally speaking, it controls how 'tightly' information is organized by <u>Group by</u> and, therefore, how many result items are produced.

All

When **Time consolidation** is set to *All*, the result set contains the fewest number of items possible, usually just one row for simple reports. **Group by** is ignored except as it relates to how **Avg requests**, **Avg visits**, and **Avg delay** are calculated.

Similar times

This is the most useful setting when you want to see information grouped by time (*Weekly*, *Daily*, etc.)   Here, data is aggregated such that similar times (same day of the month, same week of the year, same month of the year, etc.) are pooled.   This is especially useful when looking at average statistics.

No consolidation

With *No consolidation*, each distinct time period implied by **Group by** produces a group.   For example, if **Group by** is set to *Hourly*, *No consolidation* produces a different group for each and every hour that the server recorded a request.   In the case of a server running for six months, that could be as many as 24 times 180 or 4,320 distinct groups.   Clearly this option is most useful for scrutinizing a relatively brief overall time period.

Tip: Combine **No consolidation** and **Limit results** to list only the most recent time periods.   Be sure that **Limit results** is set to show the *Bottom* (not the default *Top*) items and **Sort by** is set to *Alphabetical/Chronological*.

Standard Reports

The following is a list of standard reports that can be created by selecting them from the Standard reports drop down list in the Definition tab of the element Properties dialog box.   See also Crosstabs.

Totals

Shows statistics based on all the data in the selected Time period.

URLs

Shows statistics based on each URL.

Pages

Similar to URLs above but statistics are only based on HTML pages, Applications and Downloads; graphics and multimedia elements are always excluded even if the report filter includes them.   If the report filter excludes Pages, Applications or Downloads, they will not be included in this report.

Domains (virtual servers)

Shows statistics based on each domain (virtual server).

Directories

Shows statistics based on each directory (folder).

Object types

Shows statistics based on each Object Type.   (Professional Editon only) Object Types may modified with the Object Types tab of the Database Manager.

Groups (Professional Edition Only)

Shows statistics based on each Group that has been previously defined with the URL Groups tab of the Database Manager.

Browsers

Shows statistics based on each web browser encountered in the data.   Requires that log information contains USER_AGENT information.   If this information is not in the logs and you're using the Microsoft Information Server or O'Reilly WebSite, you can run Setup again to install the necessary components.

Operating systems

Shows statistics based on each operating system encountered in the data.   Requires that log information contains USER_AGENT information.   If this information is not in the logs and you're using the Microsoft Information Server or O'Reilly WebSite, you can run Setup again to install the necessary components.

Organizations

Shows statistics based on the types of organization that visitors belong to (US commercial, US non-profit, US educational, etc.).   These types are derived from site names so IP addresses must be resolved into site names before this can be used.   Site names can be resolved when running the report with the Options tab of the Design window or when the database is updated by using the Updates tab of the Options dialog box.

Countries

Shows statistics based on the countries that visitors come from.   These are derived from site names so IP addresses must be resolved into site names before this can be used.   Site names can be resolved when running the report with the Options tab of the Design window or when the database is updated by using the Updates tab of the Options

dialog box.

Regions

Shows statistics based on the region of the world that visitors come from.   These are derived from site names so IP addresses must be resolved into site names before this can be used.   Site names can be resolved when running the report with the Options tab of the Design window or when the database is updated by using the Updates tab of the Options dialog box.

Number of visitors

Shows the total number of visitors (based on unique IP addresses) that have visited the site.

Number of visitors based on cookies

Shows the total number of visitors based on a combination of unique IP addresses and cookies.   This only applies to sites that use persistent cookies.

Number of new visitors

Shows the number of visitors (based on unique IP addresses) that visited the site for the first time during the time period.

Number of users

Shows the total number of authenticated users that have visited the site.   This is only applicable where users are required to log into the site with a username/password combination.   This is not the same as an email name.

Jumps within this site

This is a complex report that shows how visitors jumped around the site.   For every URL, it shows all the pages jumped to from there.   You can use this to determine which links are most (and least) commonly taken from key pages.   This report is only available if the logs contain REFERRER information. If this information is not in the logs and you're using the Microsoft Information Server or O'Reilly WebSite, you can run Setup again to install the necessary components.

Jumps to other sites

If you're using either REDIRECT.EXE (provided with Hit List) or Redir.dll, this report shows you where redirections occurred from and where they went.

List of visiting site names

Shows a list of all the unique site names that have visited your web site.   Site names can only be displayed if you have previously used the Lookup site names option to resolve them from IP addresses.   Site names can be resolved when running the report with the Options tab of the Design window or when the database is updated by using the Updates tab of the Options dialog box.

List of visiting IPs

Shows a list of all the unique IP addresses that have visited your web site.

List of new visiting site names

Shows a list of all the unique site names that visited your web site for the first time during the time period.   Site names can only be displayed if you have previously used the Lookup site names option to resolve them from IP addresses.   Site names can be resolved when running the report with the Options tab of the Design window or when the database is updated by using the Updates tab of the Options dialog box.

List of new visiting IPs

Shows a list of all the unique IP addresses that have visited your web site for the first time during the time period.

List of sources (Referrers)

Shows a list of the sites or URLs that referred visited to this site. This report is only available if the logs contain REFERRER information. If this information is not in the logs and you're using the Microsoft Information Server or O'Reilly WebSite, you can run Setup again to install the necessary components.

Tip: This report may produce the most useful results when you filter-out any referrers from your own site.   See the *Ignore requests from this Source site* field in the Options tab of the Design window.

List of users

Shows a list of all authenticated users who have visited the site. This is only applicable where users are required to log into the site with a username/password combination.   This is not the same as an email name.

Number of unique cookies

Shows the total number of unique cookies.

List of cookies

Shows a list of the unique cookies observed.

Entry pages

Shows a list of the first pages that visitors saw when they entered the site.

Exit pages

Shows a list of the pages from where visitors left the site. This report is only available if the logs contain REFERRER information. If this information is not in the logs and you're using the Microsoft Information Server or O'Reilly WebSite, you can run Setup again to install the necessary components.

Path to pages

This is a very useful report for determining how visitors moved through the site.   For each Page, it shows the path (up to five jumps deep) that visitors took to get there.   This cannot take browser caching into account.   **Group by** is ignored when running this report.

Previous pages

This is similar to *Path to pages* except that it only goes one jump back.   That is, for each Page, it shows the previous page that visitors were looking at.   This can be useful in determining how users found certain pages. **Group by** is ignored when running this report.

Number of single page visits

A single page visit occurs when a visitor enters the site then leaves without seeing any other pages.   A high number of single page visits may indicate that a revised home page could entice visitors to see more pages before they leave. This report is only available if the logs contain REFERRER information. If this information is not in the logs and you're using the Microsoft Information Server or O'Reilly WebSite, you can run Setup again to install the necessary components.

List of single page visits

Shows a list of the pages that visitors saw during their single-page visit.   The home page (/, */Index.htm* or */Default.htm* is usually the most common, of course).   You might look at these pages to see if they can be enhanced to convince visitors to see other pages. This report is only available if the logs contain REFERRER information. If this information is not in the logs and you're using the Microsoft Information Server or O'Reilly WebSite, you can run Setup again to install the necessary components.

Application arguments (queries)

Shows a list of the distinct application arguments (queries) that were found.   (Professional Edtion Only) If Application Argument Groups have been defined, the list will normally be composed of a combination of Application Argument Groups and queries that have no matching group.   If the <u>Use site name and application argument groups</u> switch is off, only queries will be displayed; groups will be suppressed.

Application argument (queries) parameters (Professional Edition Only)

If Application Argument Parsing has been properly defined with the Parsing tab of the Options dialog box, this report will list each query parameter and its associated values (if applicable).   You can use a <u>Filter</u> to limit the report to specific parameters.

Advertising impressions (Professional Edition Only)

Providing you've previously used the Ads tab of the Database Manager to tell Hit List how to recognize ads running on your site, this report will show you how many times each ad has been shown (*impressions*).   Used in conjunction with the **Categorize** setting, this will organize the report by ad group, ad name or even by the HTML page that presented each ad.

You can use a <u>Filter</u> to limit the report to specific ads or ad groups.

Advertising click-throughs (Professional Edition Only)

Similar to the above, this report shows you how many times each ad has been clicked-on by a visitor (*click-through*). Used in conjunction with the **Categorize** setting, this will organize the report by ad group, ad name or even by the HTML page that presented each ad.

You can use a <u>Filter</u> to limit the report to specific ads or ad groups.

Advertising impressions and click-throughs (Professional Edition Only)

This report is essentially a combination of the two reports above in that it shows both the number of impressions and the number of click-throughs for each ad in one concise table or graph.

You can use a <u>Filter</u> to limit the report to specific ads or ad groups.

Source (referrer) parameters (Professional Edition Only)

If Source Parsing has been properly defined with the Parsing tab of the Options dialog box, this report will list each source parameter and its associated values (if applicable).   This is normally used to show the criteria people entered into search engines to find your site.   You can use a <u>Filter</u> to limit the report to specific parameters.

Substitution Fields

Hit List includes several substitution fields (i.e. *macros*) that can be used to embed information in reports.   You may use these in <u>Display as</u> and <u>Description</u> fields and in the <u>Subject</u> line of an email.   They are expanded when the report is run.

You may use *@Anything@* within <u>File name</u> fields to enable time-stamped reports.   The *@Domain@*, *@ShortDomain@*, *@RootDomain@* and *@DomainIP@* fields can also be used for similar purposes.   No other substitution field will be accepted in a file name.

@ReportTitle@

Prints the report title.

@ElementName@

Prints the name of the current report element.

@Domain@

Prints the full name of the domain from which the report is running if a server name or IP filter was used or the report is generated via an Ad-Hoc form (e.g. *www.mycompany.com*). This is frequently used in directory naming for reports that iterate over all domains.   Note that @Domain@ will be blank if no filter information was specified because Hit List has no way to verify that all requests were made on one domain.

@RootDomain@

Same as above but strips off a leading *www.* This is frequently used in directory naming and email (e.g. *webmaster@@RootDomain@*) for reports that iterate over all domains.

@ShortDomain@

Same as @Domain@ but strips off a leading *www.* and trailing *.com* and *.net* sections.   This is frequently used in directory naming for reports that iterate over all domains.

@DomainIP@

Prints the IP of the domain from which the report is running.   See @Domain@ for limitations.

@TimePeriod@

Prints the name of the time period of the report (e.g. *This Week*).

@TimePeriodFrom@

Prints the starting date of the report, based on *@TimePeriod@*.

@TimePeriodTo@

Prints the ending date of the report, based on *@TimePeriod@*.

@FirstDate@ and @FirstDateAndTime@

Prints the first actual date (and time) encountered while running the report.   These differ from *@TimePeriodFrom@* in that these are based on actual days/times not on the hypothetical first date that *@TimePeriodFrom@* uses.

@LastDate@ and @LastDateAndTime@

Prints the last actual date [and time] encountered while running the report. These differ from *@TimePeriodTo@* in that these are based on actual days/times not on the hypothetical last date that *@TimePeriodTo@* uses.

@SectionNames@

Prints a hyperlinked (HTML only) list of all the <u>Sections</u> in the report.

@Anything@

Anything else surrounded by *@* causes the text within to be used as a format mask for the current date using standard Microsoft Excel time-formatting codes.   For example, *@mm/dd/yy@* will print the current date such as *12/31/96*.

Tip: Unlike most substitution fields, this can also be used when specifying file names.   However, be sure to replace any date delimiting slashes (/) with a - or some other character because file names cannot include the   / character.   For example, you could specify a file name as *WebReport@yy-mm-dd@.htm*.

Tip: If your log files have a date somehow embedded in them (e.g. *data-Jun-14.log*), do **not** use a substitution field to specify the file name in the <u>File Name</u> field of the Log Files tab.   If you were to enter something like *data-@mmm@-@dd@.log*, Hit List would look for one specific log with today's date.   Instead, use standard file pattern matching.   In this case, that would mean something like *data-*.log*.

REDIRECT.EXE

Redirect.exe is a small WIN-CGI program that allows you to add jumps to other sites in a way that will be noted in the log and, therefore, detected by Hit List.   This is extremely useful if you want or need to tell other sites how many requests you brought to them.   The <u>Jumps to other sites</u> report analyzes these jumps.

Setup puts a copy of REDIRECT.EXE and REDIRECT.DLL in the Hit List application folder.   See your server documentation for information about using WIN-CGI programs.

Tip: Because of the way Microsoft IIS handles Win-CGI programs, IIS users will need to add both REDIRECT.EXE and REDIRECT.DLL to the execution location (usually \Scripts).   Your form or application should call REDIRECT.DLL not REDIRECT.EXE.   The .DLL will then call the .EXE providing they're both located in the same directory.

Usage (WebSite or Netscape):
REDIRECT.EXE?[URL]

For example:
<A HREF="http://CGI-WIN/REDIRECT.EXE?http://www.SomeOtherSite.Com/">Go to Some Other Site</A>

Usage (Microsoft IIS):
REDIRECT.DLL?[URL]

For example:
<A HREF="http://Scripts/REDIRECT.DLL?http://www.SomeOtherSite.Com/">Go to Some Other Site</A>

Getting Your Log Files

Although Hit List is optimized for Windows NT/95 web servers, it works very well with the standard log files produced by the popular UNIX servers like Apache, CERN and NCSA so long as you are using them with either the Common Log Format or the Extended Common Log format.   Simply enter the name of the file into the *File name* field in the Log Files tab of the Options dialog box and Hit List will do the rest.

Tip: Some servers can produce separate files with Access, Error and Access-Client information.   There is insufficient linkage information in these files for Hit List to correlate them.   Therefore, if you want to analyze User_Agent information (like Browser, Operating System, etc.), you should configure your server to output one Extended Common Log format file instead.   See your server documentation for details.

Important: If you need to use FTP to get your log files, you should use the built-in FTP download system to ensure correct processing.   See the *File name* field in the Log Files tab of the Options dialog box for details.   If you manually use FTP to copy files from a server to your local machine for analysis, be very careful to copy them in chronological order where the oldest file is copied first and the newest one is copied last.   Hit List uses the file's time stamp to determine the sequence to load multiple files (when pattern matching is used) to ensure that old data doesn't inadvertently enter the database.   The file name itself is irrelevant.   If the time stamps are not in the order that the logs were produced, Hit List may refuse to load actually newer data because it had already loaded data from what was incorrectly identified as a newer log.   If you mistakenly copy files out of order, you can override the normal data integrity check by using the *Let old data enter the database* checkbox in the Updates tab of the Options dialog box.   However, be very careful when turning this option on; it is very easy to mistakenly add duplicate information.   We strongly suggest that you only use this option when absolutely necessary and immediately turn it off when done.

If your web site is located at an Internet Service Provider, you may have to make arrangements to get copies of your log files for Hit List to analyze.   There are a few different ways to accomplish this:

1.   If your ISP is running a Windows 95/NT based web server like Microsoft IIS, O'Reilly WebSite or an Netscape NT version, he can purchase Hit List and install it on his server.   If he purchases the Professional Edition of Hit List, all of his customers could use the Remote Reports features of Hit List to run their own reports at their own offices without any work on his part.   This would be easiest for both of you.

2.   If your ISP is running UNIX or some other operating system or does not want to install Hit List, he will have to provide you with copies of your log files.   In many cases, the web server creates one large log file for all the virtual servers hosted.   If this is the case, your ISP will have to use a small program to split the information by virtual servers and send the relevant subset of the file to you.   Your ISP is probably already doing this for at least one other client.   You and your ISP will then have to make arrangements about how you get the files.   In most cases, your ISP will ask that you use an FTP program (any web browser will usually do) to download the files from his site.

Once you have your files, use the Log Files tab of the Options dialog box to specify the file location of the log file you want to analyze (you can also specify a file pattern to load multiple files at once).   After the file has been loaded into the Hit List database, you may discard or archive the log file itself; Hit List will use the database, not your logs, to run reports.

Import

The import dialog lets you copy entire reports and individual report elements into the current reports database. Choose <u>Import</u> from the Reports menu of the main window to select the database to import from.

By default, Import will show you a list of all reports and report elements in the new database.   You may find it more useful to check **Only show reports and report elements that don't exist in the current report database** to filter-out items that appear to exist in both databases.   Note that Hit List doesn't actually know if two reports or report elements are really identical; it simply looks for identical names.

Import specific reports or report elements by selecting them then clicking the **Import Selected** button.   Import all the items currently shown by clicking **Import All**.

Imported items that already exist in the current database will have their names appended with a number to ensure that no two objects have the same name.   This could cause <u>Scheduled Reports</u> to fail until the object's name is changed or the Scheduled Report is modified.

Report Design/Log File

Hit List normally uses a global log file setting to specify where logs come from.   You may override this setting and pick a particular log file or set of log files to use for each report.   This is normally used in situations where your log files do not have enough information for Hit List to differentiate between virtual servers (generic Common Log and Extended Common Log formats, for example) so you need to put each log file into its own database.   This tab is usually used in conjunction with <u>overriding the global database</u>.   The Database tab can be used to filter your logs by server name/IP or URL pattern such that a single log can be split into different databases.

This setting has no effect unless this report is set to <u>update the database before it runs</u>.

To select a particular log file or files for this report, simply check the **Override** box and complete the log file fields exactly as you would for the <u>global log file</u> setting.

Report Design/Database

Hit List normally uses a global database setting to specify where it stores its data after importing your logs.   You may override this setting and pick a particular database to use for each report.   This is normally used in situations where your log files do not have enough information for Hit List to differentiate between virtual servers (generic Common Log and Extended Common Log formats, for example) so you need to put each log file into its own database.   This tab is usually used in conjunction with <u>overriding the global log file</u>.

To select a particular database for this report, simply check the **Override** box and complete the fields exactly as you would for the <u>global database</u> setting.

You use the **Only add requests that match the following** area to filter incoming logs into different databases. This is only necessary if you want to split a single log file into different databases for differentiation or performance reasons.   You may filter based on either server names/IP addresses (Microsoft IIS and O'Reilly WebSite logs) or URL patterns (most other servers).   For example, you can split IIS logs into multiple databases by using multiple reports, each with a different IP setting.   Or you could filter Netscape logs by using multiple reports, each with a different URL pattern (*/Server1/** vs. */Server2/**, for example).

Tip: Even if your logs have sufficient information to differentiate between virtual servers, you still may want to split them to keep databases smaller, increasing overall performance.   You can freely combine databases with multiple servers (for low-traffic servers) with databases that represent just one server (high-traffic) to achieve a good mix or performance and simplicity.

Tip: You can combine report-by-report filtering with <u>global filtering</u> to ensure that requests from certain machines (yours, for example), never enter any database.

Options/Advanced

Warning!   These options generally affect Hit List in a significant way and could potentially degrade performance or create security holes if used incorrectly.   They are intended for advanced users only.

Name or IP of proxy server

In some environments, web pages cannot be accessed directly but must be requested from another machine (a so-called *proxy*).   If this applies to your site, you must tell Hit List the name or IP address of your proxy server.   If necessary, you may also add a port number or Hit List will assume the standard HTTP port of 80.   Enter either the name of the proxy machine (e.g. *BigProxyServer*) or its IP number (e.g. *123.456.789.012*).   If you must also add a port number, separate it from the name or IP with a colon (e.g. *BigProxyServer:84*).

Tip: While this does not directly affect DNS lookups, Hit List will bypass its usual test for an active DNS if a proxy server is specified.

See Using Hit List on Your Intranet for more information.

Allow ad-hoc reports to specify domain and database.

Ad-hoc reports (Remote Reports generated from HITLIST.HTM or similar) are normally automatically filtered to apply only to the virtual server that answered the POST request.   Combined with simple username/password validation, this allows ISPs to offer clients simple remote report capabilities without any worry that they could potentially look at other clients' data.   Unfortunately this also means that remote reports could be running on your primary server during peak usage periods.   And it makes it impossible to offer clients remote reports unless your primary web server is running Microsoft Windows NT or Windows 95.   This switch solves both problems by allowing you to use any Windows NT or 95 web server to generate remote reports.

If this box is checked, the POSTed HTML form can specify a server name or IP address filter similar to the way you could from the GUI.   See the comments in HITLIST.HTM for details.   Likewise, the HTML form could also specify a particular database from which to derive the report.   This is especially useful with servers that don't log virtual server name or IP information (Netscape, for example) so you use multiple databases.

Tip: Since this option makes it possible for someone to look at someone else's data, you might consider taking the precaution of using HIDDEN fields on the HTML form to contain this information.   This also makes it easier to most clients to use.

See Using Remote Reports for details about using ad-hoc reports.

Under Windows NT, run as an Application or Application and a Service

Microsoft Windows NT has a concept called a *Service* that is an application that is always running in the background, usually invisibly, regardless of which user (if any) is logged into the system.   This makes them easy to "set and forget".   Most web servers run as a Service.   This option does not apply to Microsoft Windows 95.

If you're using Scheduled Reports, you may find it convenient to run Hit List as a Service rather than (or in addition to) being an application because it ensures that such reports always run, even if no one is currently logged in.   To run Hit List as Service, simply select *Application and a Service* in the list.   Hit List will run as an invisible Service whenever you manually instruct it to start using the Services applet in the Control Panel.   If you want Hit List to automatically startup when NT boots (strongly suggested for Scheduled Reports), check the **Start when the system boots** checkbox.   Note that checking the Start option does not start the Hit List service; it only ensures that it will start the next time the operating system boots.   Use the Services applet to start it immediately.

Hit List shares all the same settings whether it runs as a Service or an application.   Therefore, you set options on the Service simply by running the Hit List application.   Options and Scheduled Reports changed with the application are automatically picked-up by the Service.

Web browsers are never launched when Hit List runs a Service because the browser would be invisible.

By default, Windows NT Services do not have access to your network.   If this is required because Hit List will be retrieving logs via the network, be sure to use the Services applet in the Control Panel to manually change the MarketwaveHitList service to log-on as a specific user with appropriate network privileges.

Email cannot be sent via Microsoft MAPI when running as Service unless you create an account for the Hit List Service and use the Services applet to configure Hit List to use it.   POP/SMTP mail has no such limitation.

Tip: Be sure to turn **Disabled Scheduled Reports** off or the Service will not run any Scheduled Reports.

Priority

If you're running Hit List on your primary server and you either want to allow remote reports or otherwise run it during peak periods, you should probably lower Hit List's **Priority** setting to *Low*.   This allows other applications (such as your web server) to remain responsive even while Hit List is running.   Setting to low only slows Hit List during the specific times that another application requests attention.   Therefore, in most situations, it does not significantly slow Hit List.

Template Database

Hit List normally fills each newly created database with certain default Site Name Groups, URL Groups, Browsers, etc.   You may tell Hit List to use defaults of your choosing instead by entering the name of an existing Hit List database file (usually a MWD file).   When creating a new database, Hit List will copy the URL Groups, Site Name Groups, App Argument Groups, Ads, Object Types, Browsers and Operating Systems settings from this file instead of the using the built-in defaults.   This can save a tremendous amount of time if you've customized these lists and use multiple databases.   You can also use the **Import** button in the Database Manager to selectively copy this information between two Hit List databases.

Default database when running an ad-hoc report

Since ad-hoc reports cannot normally specify the database to get their data from (see **Allow ad-hoc reports to specify domain and database** above), they must have a default database to use.   Hit List will normally use the database that it created the very first time it was run on your machine.   If you later move the database or otherwise wish to have remote reports use another database, you can specify the new file in this field.   If you move your database but don't change this setting, the first remote report to run will force Hit List to re-load all your logs and re-create the original database.   You will need to manually modify some Registry settings if you're storing your database on an ODBC datasource.   Contact support@marketwave.com for more information.   See Using Remote Reports for details about using ad-hoc reports.

Options/Parsing

Hit List Professional has the unique ability to parse complex query strings (application arguments) and referrers (sources) into meaningful pairs of Parameters and Values.   This makes it very easy to run reports for sites that are highly-database driven (Cold Fusion sites, for example).   With application argument parsing Hit List can take a query such as

?Template=Cities.tpl&City=Seattle&State=WA

and parse it into

Template=Cities.tpl
City=Seattle
State=WA

Given these parsed values, Hit List can create a report that shows all the various values for each parameter.   The <u>Application Argument Parameters</u> standard report, for example, would produce something like:

```
Parameter     Value         Requests
=========     =========     ========
Template      Cities.tpl    4500
              States.tpl    4000
              Maps.tpl      3500

City          Seattle       1000
     Boston        500
     LA            350

State         Washington    1250
              Oregon        1000
     MA            750
              California    500
```

etc.

You can use sophisticated <u>Parameter/Value filtering</u> to limit your reports to specific parameters.

With source parsing, Hit List can also show you what people entered into search engines to find your site.

Query arguments such as

?Seattle+Music+Live

would be parsed into parameters with no values.   That is, the above would look like the following in a report

```
Parameter     Value         Requests
=========     =========     ========
Seattle       (None)        3750
Music         (None)        2500
Live          (None)        1000
```

Query and source arguments that contain multiple values in the same parameter such as

?Search=Seattle+Music+Live

are logically recorded as distinct queries.   For example:

```
Search=Seattle
Search=Music
Search=Live
```

Hit List parses queries into pairs when the data is loaded from your logs into the database.   Therefore, it is essential that you enable application argument parsing and correctly configure it for your logs before updating the database. You will need to delete and reload your database if it already contains query data that has not been parsed.

Parse application arguments into Parameter/Value pairs

Query parsing may be enabled and disabled with this switch.   By default, it is off because most logs do not contain interesting query data and the parsing may take some time.   Moreover, parsing queries before the separator characters have been properly defined will produce incorrect results.

Parse sources (referrers) into Parameter/Value pairs

Parsing of referrer strings (such as *http://www.excite.com?search=logs+analysis*) may be enabled and disabled with this switch.   By default, it is off because most logs do not contain interesting referrer queries the parsing may take some time.   Moreover, parsing sources before the separator characters have been properly defined will produce incorrect results.

Parameter/Value separator character(s)

Enter the character or characters that separate the parameter name from its value (e.g. *MyParam=MyValue*).   Often this is = or sometimes ∴.   You may enter multiple characters if your queries use different characters for this purpose. If so, separate each possible divider with the pipe (|) character.

Parameter/Parameter separator character(s)

Enter the character or characters that separate parameters from each other (e.g. *MyParam1=Foo&MyParam2=Bar*). This is usally *&*.   You may enter multiple characters if your queries use different characters for this purpose.   If so, separate each possible divider with the pipe (|) character.

Value/Value separator character(s)

Enter the character or characters that separate different words that might be found in the same search key (e.g. *MyParam=Value1+Value2+Value3*).   The + character is commonly used for this.   You may enter multiple characters if your queries use different characters for this purpose.   If so, separate each possible divider with the pipe (|) character.

In addition to query parsing, Hit List can also create <u>application argument groups</u> that can make cursory analysis very easy.

Database Manager/App Argument Groups

Application Argument Groups are collections of application arguments (queries) that are analyzed as one unit, frequently simplifying reports and showing trends that might have been missed otherwise.   For example, you could create a *Seattle* group that contains all queries that begin with */Template=Seattle.tpl* and a *San Francisco* group that contains all queries that begin with */Template=Sanfran.tpl*.

Groups are managed from within the Database Manager.   To run the Database Manager, select Database Manager from the Tools menu of the main window.   Groups affect the database so you must have access to the database.

The *Most Common Application Arguments* report element will list the most commonly accessed queries.   If no application argument groups have been defined, it will simply show all the unique queries.

Tip: By default, Hit List will display an application argument group in place of a complete query if the query is a member of a group.   If you'd prefer to list only actual queries (not groups), uncheck the <u>Use site name and application argument groups, if defined</u> check in the Options tab of the Design window.

Groups can be based on complete queries or, more often, on patterns.   A pattern contains a portion of a complete query and the standard wildcard characters (* and ?).   For example,:

1.   A pattern of **seattle** includes any query that contains *Seattle* anywhere.   Case is not important.

2.   A pattern of */Template=Seattle** includes any query that begins with */Template=Seattle* but would not include something that ends with Seattle.

If you change an Application Argument Group, Hit List will recalculate all Groups the next time a report is run.

Creating a New Group

Click the **New** button next to the **Groups** list and enter then name of the new Group.

Modifying an Existing Group

Select the Group in the **Group** list and click the **Change** button next to the **Groups** list.   Enter the Group's new name.

Deleting an Existing Group

Select the Group in the **Group** list and click the **Delete** button next to the **Groups** list.   This removes the Group but does not delete any data from the database.

Adding a New Pattern to an Existing Group

Select the Group that the new pattern belongs to from the **Groups** list then click the **New** button next to the **Argument within Group** list.   Enter the complete query   or query pattern.

Modifying a Pattern

Select the pattern from the **Argument within Group** list then click the **Change** button next to it.   Enter the new pattern.

Deleting a Pattern

Select the pattern from the **Application within Group** list then click the **Delete**   button next to it.   This disassociates the pattern from the currently selected Group but does not delete any data from the database.

See Also:
<u>Object Types</u>

Database Manager/Site Name Groups

Site Name Groups are collections of site names (names resolved from IP addresses) and/or IP addresses that are analyzed as one unit, frequently simplifying reports and showing trends that might have been missed otherwise. This is most commonly used to organize requests from different machines within the same organization. By default, Hit List creates groups for the common online service providers such as AOL, Compuserve, MSN, etc. For intranet sites, this feature can also be used to create groups for internal company departments or regions.

Groups are managed from within the Database Manager. To run the Database Manager, select Database Manager from the Tools menu of the main window. Groups affect the database so you must have access to the database.

Tip: By default, Hit List will display a site name group in place of a complete site name if the site is a member of a group. If you'd prefer to list only actual site names, uncheck the Use site name and application argument groups, if defined check in the Options tab of the Design window.

Groups can be based on complete site names or, more often, on patterns. A pattern contains a portion of a complete site name and the standard wildcard characters (* and ?). For example, a pattern of *.aol.com includes any site that ends with .AOL.COM as do all AOL servers in the United States. Case is not important.

If you change a Site Name Group, Hit List will recalculate all Groups the next time a report is run.

Creating a New Group

Click the **New** button next to the **Groups** list and enter then name of the new Group.

Modifying an Existing Group

Select the Group in the **Group** list and click the **Change** button next to the **Groups** list. Enter the Group's new name.

Deleting an Existing Group

Select the Group in the **Group** list and click the **Delete** button next to the **Groups** list. This removes the Group but does not delete any data from the database.

Adding a New Pattern to an Existing Group

Select the Group that the new pattern belongs to from the **Groups** list then click the **New** button next to the **Argument within Group** list. Enter the complete site name, IP address, site name pattern (e.g. *aol.com) or IP pattern (e.g. 123.456.789.*)

Modifying a Pattern

Select the pattern from the **Argument within Group** list then click the **Change** button next to it. Enter the new pattern.

Deleting a Pattern

Select the pattern from the **Application within Group** list then click the **Delete** button next to it. This disassociates the pattern from the currently selected Group but does not delete any data from the database.

See Also:
Object Types
Groups
Browsers
Operating Systems
Tools
Application Argument Groups

<u>URL Groups</u>

Using Virtual Servers

Hit List has many powerful features that make working with multiple virtual servers very easy.   For many sites, Hit List will just automatically "do the right thing" without any manual intervention.   This section is mainly designed to describe issues that may arise in some situations.

Step 1. Configuring Your Server Logs

This varies depending on your server.

Microsoft IIS

Use the *Standard format* and your logs will contain all the necessary information.   Do not use the *NCSA format* option.   This is configured from the IIS Service Manager's Logging tab.

O'Reilly WebSite

All of the choices in the Logging tab provide virtual site information because even the *Common* and *Combined* formats actually use a modified log format that Hit List understands.   However, we suggest using the *Windows* option because it provides the most information.   The *Windows* format is logically similar to the *Combined* option in that it provides valuable User-Agent and Referrer information but also provides the amount of time between when a client requests a URL and when the server actually delivers it to him.   Hit List can report this "delay" information, which is useful in finding and resolving server bottlenecks.   If you choose not to use the *Windows* format, we very strongly suggest that you use the *Combined* rather than the *Common* format because the *Common* format does not provide User-Agent and Referrer information, making it impossible for Hit List to show which browsers your visitors used and perform complex path analysis.

Netscape, Apache and NCSA

Optimally using these servers with Hit List requires some manual configuration changes.   See Using Netscape, Apache and NCSA topic for more information.

Other

Most other servers are limited to using the old Common log format, which does not provide any virtual site information.   You can still use Hit List with these servers but you will not be able to generate some types of reports. See below for details.

Step 2.   Configuring Hit List

If your logs have virtual site information (IIS and WebSite), you do not need to do anything special to have Hit List recognize this.   By default, Hit List will read your log files and differentiate between servers but still store all the data in one database file.   This makes it very easy to run reports that combine server-wide information with data about each virtual server (the *Virtual Server Overall Report*, for example).   It also makes it extremely simple to create one report runs for each of your virtual servers.   For example, if you're an ISP with 20 clients and you want each to get a weekly report, you can create the desired report and check the Run this report for each virtual server checkbox.   A different report will be generated for each virtual server.   As you add more clients, each new clients will also get the report generated for him with no additional work on your part.   You can also create report groups by combining this checkbox with the Server name or IP filter to create a specific list of virtual servers to run the report over.   In this way, you can automatically create different reports for different types of clients (Marketing-oriented clients versus technically-oriented clients, for example).   You can also run reports for specific virtual servers simply by leaving the **Run this report…** checkbox off and entering the name (for WebSite) or IP address (for IIS) of the virtual server in the **Server name or IP** filter field.

You can still create reports for each virtual server even if your logs do not contain virtual server information.   In this case, you'll want to use report-level overrides for the log file and database settings to tie reports to specific logs and databases.   In other words, you'll end up with a different database for each virtual server.   Given that, you can run reports for each of your clients by creating a single "base" report and using the Copy command to duplicate it

for each client.   You should give each report its own unique log file and database information.   Since each virtual server will be stored in its own database, there is no way to run one report that shows summary information about all virtual servers.

Loading Old Logs

Hit List normally goes to great lengths to ensure that only "new" data enters the Hit List database. This makes it possible for it to automatically work with cycled log files. It also means that you might be surprised if you try to load an old log and find that the information didn't enter the database (the report would be essentially empty).

If need to analyze logs that were created before the last record entered your current database, you have a couple of options.

1. You can simply put this new log in its own database. This is best when you want to look at the old data in isolation. The easiest way to do this is to create a new report and use the <u>Log File</u> and <u>Database</u> tabs within that report to override the global settings.

2. If you want to add the data to your overall database, you have to override Hit List's normal 'old' suppression by turning on the <u>Allow old data to enter the database</u> using the Updates tab of the Options dialog box. When you do this, Hit List will add all records it finds to your database. This could easily lead to duplicate records if you aren't very careful. Therefore, to ensure that only a particular log enters the database, you should specify that file by either changing your <u>global log file</u> setting or using a report that specifies its <u>own log file</u> setting. Be certain to turn the **Allow old data** off after you are done processing the specific log.

If you've already tried to load old data into the database and it didn't seem to work, you might considering simply deleting the database file and processing your logs again, this time making sure to load the older logs as well as the newest ones. You can do this by either finding the database file yourself (usually HITLIST.MWD) or simply using the <u>Delete the entire Hit List database</u> option in the Tools tab of the Database Manager.

Getting Your Files via FTP

If your logs are stored on a UNIX machine or at your Internet Service Providers location, you might need to retrieve them via the Internet standard file transfer protocol (FTP).   Hit List supports this directly by letting you specify a log file name or pattern as a standard URL. In the log files tab of either the Options dialog or the Design dialog, you can enter a file name such as:

ftp://ftp.MyISP.com/MySite/access.* Anonymous Password@me.com

This would instruct Hit List to attempt to connect to *ftp.MyISP.com*, switch to the *MySite* directory, and find all files that begin with *access*.   Note that UNIX is normally case-sensitive so *Access* and *access* are not the same.   The connection could fail if you don't provide an appropriate username and password.

In the example above, Hit List might find more than one file name that begins with *access*.   If so, it will apply the same file cycling behavior as it would for local files.   This means that it will only download files that it thinks it might need to process, saving you a great deal of time and making it normally unnecessary to change the file name every time your log if cycled.

If you need to retrieve multiple files that cannot be expressed with simple pattern matching, you can specify multiple file specifications on the same line by separating them with the pipe (|) character.   For example:

ftp://ftp.MyISP.com/MyFirstSite/MyLog.* Anonymous Password@me.com|
ftp://ftp.MyISP.com/MySecondSite/YourLog.* MyUserName MyPassword

Of course, if these files represent different virtual servers and they do not contain enough information for Hit List to differentiate between them, you probably shouldn't add them to the same database anyway.   To keep them distinct, use report overrides for the log file and database settings and enter each of the above file names into its own report.

Non-US Date Formats

Hit List can work with logs that contain any date format recognized by Windows.   The only caveat is that the log file itself and the machine processing the logs must use the same regional settings.   That is, if your logs contain US-formatted dates, you will have to set your processing machine to US dates as well.   Similarly, if your logs are in German, for example, the processing machine will have to be using German date formats as well.

If you aren't sure which date format your logs use, open the log with Notepad, WordPad, Microsoft Word or any other text editor.

Note: O'Reilly WinLog is always in mm/dd/yy format, regardless of your date settings.   Hit List may have trouble reading logs with mm/dd/yy formats on machines that aren't set to that format.   You should either change the date format of the log processing machine to US format or use a different log format (Combined, for example).

Using Crosstabs

Crosstabs show correlations between information.   For example, if you want to see how web browser usage varies depending on the country, you can crosstab (short for Cross Tabulate) the Browser field with the Country field.   The resulting table will have a list of countries going down and a list of browsers going across.   In the middle, you can pick any calculation to apply but *Requests* is most commonly used.   You should avoid using *Visits* for speed reasons.

Some of the built-in reports use crosstabs.   For example, the *Technical Analysis* report uses a crosstab to show the correlation between HTTP Response code and Week so you can look at errors on a week by week basis.   The *Marketing* report uses the Browser vs. Country crosstab described above.

See Report Element Properties for usage information.

Moving Your Database

Hit List is really two distinct pieces, a graphical user interface which most people see and a report "engine" that actually does the work.   This is part of what makes Remote Reports possible.   Since the Hit List engine is potentially shared by both the GUI and the Remote Report application, it is stored with other "common" files in either your *\program files\common files\Marketwave* or *\winnt\Marketwave* directories.   The engine is what normally interacts with your database so, by default, the engine stores the database in a "known" location – its own directory.   Because the database can grow quite large (up to 1 gigabyte), it's not unusual to want to move the database to a new location.   This can be done easily.

1.   Physically move the Hit List database (usually *HITLIST.MWD*) from the default directory (see above) to its new location.

2.   Use the <u>Database</u> tab in the Options dialog box to point to the new location.

3.   (Professional Edition Only) If you allow Remote Reports, be sure to also update this information in the Advanced tab of the Options dialog box.   Remote Reports have the flexibility to default to a different database than "local" reports so be sure to keep them in synch unless you need this capability.   If you don't update the database location in the Advanced tab, any incoming Remote Report request will cause Hit List to regenerate the database in the original location.

Sharing Reports

Hit List stores your report definitions in a small database named HLRPT20S.MDB or HLRPT20P.MDB (the S is for Standard; P is for Pro).   This is normally in the same folder as the main Hit List application.   If you want to trade report definitions or new report elements with other people, you can do it in two ways.

1.   You can simply open their HLRPT20x.MDB file by selecting <u>Open Report Database</u> from the Reports menu. Any changes you make will be made to that database so be careful, especially when specifying file paths because paths are always relative to the machine running them.   For this reason, using UNC path names is a good idea in a LAN environment.   A UNC pathname doesn't use drive letters, rather it's similar to an Internet URL in that is has machine names.   For example, a UNC file would look like *\\MyMachine\MySharedDirectory\Dir1\File1.TXT*. That same file could also look like *G:\Dir1\File1.TXT* if the *G:* drive letter were mapped to *\\ MyMachinMySharedDirectory*.   The UNC version would work from any machine in your LAN; the mapped version would only work in cases were *G:* was already mapped to *MyMachine*.

2.   You can import specific reports or report elements from one database to your database.   Begin with having the "target" database open in Hit List.   Then select Import from the Reports menu and choose the database that has the new items you want.   You will be presented with a list of both entire reports and individual report elements.   You can import any combination of these.   See Import for more information.

This capability is very powerful in that it makes it possible for people to work cooperatively and allows you to extend Hit List be downloading new reports and report elements from other people.   Watch the *Tips and Tricks* section of the Marketwave web site for some of these.

Performance Tips

There are a number of ways to increase the speed that Hit List generates reports.   These are listed in no particular order.

1.   Make your reports simpler by deleting report elements that you don't really need or run slight variations of the same report on different schedules such that you get the simpler report more frequently than you get the complex one.

2.   Use the <u>Report Scheduler</u> to automate report generation such that reports are run when you aren't in the office.   Although this doesn't really speed reports, it has a similar apparent effect.

3.   Don't use <u>Visit</u> calculations more than necessary.   Because of the way visits are calculated, Hit List takes 2-3 times longer to determine them than it does to calculate *Requests* or anything else.   Of course, don't Sort by *Visits* either.

4.   Don't do <u>Reverse DNS Lookups</u> more than necessary.   You can control these on a report-by-report basis with the Options tab in the Design window for each report.

5.   Keep your database small.   You can do this in a number of ways including:

A.   Don't add data that you'll never use.   If you don't care about GIFs, keep them from ever entering the database with the <u>Updates</u> tab of the Options dialog.

B.   Purge old data periodically.   If you really only care about data on a monthly basis, don't let your database contain more than a month of data at any one time.   Simply backup your database then use the <u>Tools</u> tab of the Database Manager to delete the old information.

C.   Split virtual servers into different databases.   Even if your logs contain sufficient information for Hit List to differentiate, it might make more sense to use the database import filtering available in the <u>Database</u> tab of the Design window to split them into different physical files.   While this would preclude getting overall statistics for your server, it speeds database updates and report generation because it keeps the database smaller and eliminates the need for any <u>Server name or IP</u> filtering.

6.   Don't run certain standard reports more than necessary.   Reports that include complex path tracing (*Jumps within the Site*, *Previous Pages*, *Path Thru Site*, etc.) can take a very, very long time to generate.   Additionally, as this information tends to be fairly static over time, if you need these reports, run them as infrequently as possible, not as part of a daily or weekly report.

7.   Compact your database.   As Hit List adds logs to the database, the database tends to grow very rapidly. Generally speaking, using the <u>Compact</u> option in the Database Manager's Tools tab can shrink your database by as much as 40%.   While this doesn't directly impact speed, it may increase it considerably in practice because it reduces the amount of RAM required for efficient processing.   See *Add more RAM*, below.

8.   (Professional Edition Only) If you use URL Groups, be sure that the <u>Assume that each URL is a member of no more than one Group</u> switch is on.   This lets Hit List calculate groups much, much more efficiently.

9.   Add more RAM to your machine.   Hit List is a database application.   Therefore, it spends much of its time hitting your hard disk if your database is larger than the amount of available physical RAM.   For example, if your database is 400MB and you run Hit List on a 32MB machine, Hit List has to swap memory twice as often as when run on a 64MB machine.   This accounts for much of the processing time.   Given the low price of RAM, we suggest adding as much RAM as possible, especially if you have databases over 100MB.

10.   Get a faster machine.   Although we don't like asking people to buy new machines, the fact is that Hit List does a tremendous number of very complex calculations.   As with any similar database program, the CPU speed

dramatically impacts processing time.   In generally, we never suggest running Hit List on anything less than a Pentium 90 and strongly suggest a P133 or better (with lots of RAM) for large sites.   In fact, given the prices of Pentium Pro machines, if you're buying a new machine, you might consider a Pentium Pro 200.

What to Do If You Get "Not enough information…"

Hit List will display the "Not enough information…" message for any report element that didn't find any matching records.   Depending on the report element, this can be caused by a number of different situations.

1.   If your entire report is filled with 0's and "Not enough information" it probably means that the Hit List database is either empty or you requested a time period for which there is no information.   Find the database file (search for HITLIST.MWD).   If the file is larger than about 400K, the database isn't empty so the cause is probably the Time Period setting.   If you know that you imported information for that time period, it could also result from Filter settings set such that no data matches or from attempting to load data that was older than data previously in the database.   If the later is the case, you should learn more about Loading Old Data.

2.   *Most Common Visitors* and *Most Common Visiting Countries/Regions/Organizations* will be empty if you haven't used the Lookup site names option to turn the visitor IP addresses into site names.   For example, the *Complete Analysis* report has a *Most Common Visitors* element that will be blank unless your server has already done the lookup or you've used this option (This is option is normally disabled on Windows NT 3.51 systems the first time you run Hit List due to a serious bug in NT 3.51).   If you're running Windows 95 or Windows NT 4, simply open the report for design then check the Lookup box.   If you're running Windows NT 3.51, we don't suggest using this option because a bug in Windows NT could cause Hit List to hang.   Instead, install Hit List on machine running Windows 95 or NT 4 and set it to use the database file stored on your Windows NT 3.51 machine. You can use the "secondary" machine to run a report that does DNS lookups.   Since these lookups are stored in the database on your Windows NT 3.51 machine, your "primary" machine can now run reports that include site names. Alternately, if you don't mind looking at IP addresses, you can switch the report element to use the IP variation of the Visitors standard report element.

3.   Most path-oriented report elements (*Previous Pages*, *Jumps within the Site*, *Path Thru Site*, etc.) will be empty if your logs do not contain so-called referrer information.   You can add this to future Microsoft IIS logs simply by running Setup on your IIS web server (you do not need to run Hit List on that machine, however).   O'Reilly WebSite logs can contain this information if you switch to either *Combined Common Log* or *Windows Log* (preferred) logging format.   Either run the Hit List Setup on your WebSite server or see the Logging tab in WebSite. See Using Netscape, Apache and NCSA if you're using any of these servers.

4.   Similar to the above, *Browser* and *Operating System* report elements will be blank if your logs do not contain USER_AGENT information.   Logs that lack referrer usually lack this as well and vice-versa.   Therefore, see above for more information.

5.   Graphs will not be displayed (no warning or error message) if there is only one data item to graph.   That is, a graph of *Most Popular Browsers* that only shows one browser will be suppressed.   If you want to see this data, consider making the element either into just a table or a table and a graph.

Using Netscape, Apache and NCSA

Hit List can easily work with logs created by these servers.   However, to get as much information from them as possible, you might want to modify their log formats slightly.   Essentially, all of these create logs in the Common Log Format specified by the original NCSA web servers.   Unfortunately this format does not contain User_Agent (for Browser and Operating System information), Referrer (for all complex path analysis reports) or server identification information (for differentiating between virtual servers).   These can all be customized, to some extent, to include this information.

Generally speaking, the Hit List issues when using all of these servers are the same; only the exact customization methods are different.   Therefore, the discussion of how to use Netscape with Hit List generally applies to Apache and NCSA as well.   If you are using one of these other servers, you should read the Netscape section first to understand the general issues then jump to the section that explains the exact customization method for your server.

See the <u>Using cookies</u> topic for additional information if your site uses cookies.

Netscape

Customizing Netscape servers depends on the particular server involved.   If you're using the older Commerce or Communication servers, you will have to get an NSAPI to add the additional information to your logs.   If you're using Enterprise or Fast Track, you can add User_Agent and Referrer information simply by changing a setting. You can also insert server identification information by getting a little tricky.

Adding User-Agent/Referrer to Enterprise and Fast Track

1.   Find all copies of your OBJ.CONF files for your server.   There will probably be two of these, one in a directory like *c:\netscape\<YourServerName>\httpd-<YourServerName>\config* and one in *c:\netscape\<YourServerName>\ admserv*

2.   Copy them to backup files.

3.   Open the original files in Notepad.   Be certain that Word Wrap is off.

4   Find the line, towards the top, that contains:

Init fn=flex-init access=

and replace whatever is currently there (up to but not including any additional *Init fn* strings) with

Init fn=flex-init access="PutCompletePathAndFileNameForYourLogsHere" format.access="%Ses->client.ip% - %Req->vars.auth-user% [%SYSDATE%] \"%Req->reqpb.clf-request%\" %Req->srvhdrs.clf-status% %Req->srvhdrs.content-length% \"%Req->headers.referer%\" \"%Req->headers.user-agent%\""

Be sure to not add any line breaks and to enter this exactly as shown except replacing the *PutComplete...* text with an actual file path and name to your log file.   Netscape servers also demand that you use the UNIX/Internet standard forward-slash (/) character to separate directories not the DOS standard back-slash (\) character. Therefore, a log file name of *c:\Netscape\MyServer\Logs\access* would be entered into the file as *c:/Netscape/MyServer/Logs/acces*s even on Windows 95 and NT systems.

Working with virtual servers in Netscape

If you also want to add server identification information (necessary for all of Hit List's automated domain processing), simply replace the - after the *client.ip%* phrase with the IP of the virtual server in question.   As Netscape demands a new instance of the server (and, therefore, a new instance of *obj.conf*) for each virtual server, so long as you vary the identification information in each *obj.conf* file, Hit List will be able to differentiate between servers.   For example, if your first virtual server is at IP 123.456.789.1 and your second is at 123.456.789.2, your *obj.conf* files would look like:

For the first server:

Init fn=flex-init access="PutCompletePathAndFileNameForServer1LogHere" format.access="%Ses->client.ip% 123.456.789.1 %Req->vars.auth-user% [%SYSDATE%] \"%Req->reqpb.clf-request%\" %Req->srvhdrs.clf-status % %Req->srvhdrs.content-length% \"%Req->headers.referer%\" \"%Req->headers.user-agent%\""

For the second:

Init fn=flex-init access="PutCompletePathAndFileNameForServer2LogHere" format.access="%Ses->client.ip% 123.456.789.2 %Req->vars.auth-user% [%SYSDATE%] \"%Req->reqpb.clf-request%\" %Req->srvhdrs.clf-status % %Req->srvhdrs.content-length% \"%Req->headers.referer%\" \"%Req->headers.user-agent%\""

Alternately, if you don't want to add this information to your logs, you should not load all your virtual sites into the same database.   Instead, you should use report-level overrides for both the <u>log file</u> and the <u>database</u> specification to ensure that each virtual site is treated separately.   Since each virtual server will be stored in its own database, Hit List's automatic <u>Run this report for every virtual server</u> feature cannot work.   Rather, to replicate a report for each virtual server, you should create the report you want for one server and use the Copy command to clone it for the other servers.   Modify the log file and database specifications for each virtual server.

Apache

Working with Apache is effectively the same as working with Netscape except that the log format customization method is different.   See the Apache site for detailed information (*http://www.apache.org/docs/mod_log_config,html*).

Adding User-Agent/Referrer to Apache

1. Find your copy of the mod_log_config.c file.

2. Modify the LogFormat directive to match the following string:

%h %l %u %t \"%r\" %s %b \"%{User-agent}i\" \"%{Referer}i\"

3. Compile and link this module into Apache.

Working with virtual servers in Apache

See *Working with virtual servers in Netscape* as all the same issues apply except that you would replace the *%l* field with the virtual server IP.

NCSA

Again, you can view using NCSA as the same as working with Netscape except that the log format customization method is different.

Adding User-Agent/Referrer to NCSA

1. Find your copy of *httpd.conf*

2. Set the *LogOptions* directive to *Combined* with the following line:

LogOptions Combined

3. Save the file and restart your server.

Working with virtual servers in NCSA

It does not seem possible to add virtual server identification to NCSA servers.   Therefore you should not load all

your virtual sites into the same database.   Instead, you should use report-level overrides for both the <u>log file</u> and the <u>database</u> specification to ensure that each virtual site is treated separately.   Since each virtual server will be stored in its own database, Hit List's automatic <u>Run this report for every virtual server</u> feature cannot work.   Rather, to replicate a report for each virtual server, you should create the report you want for one server and use the Copy command to clone it for the other servers.   Modify the log file and database specifications for each virtual server.

Using Remote Reports

If you are running your web server with Microsoft IIS, O'Reilly WebSite or any NT version of Netscape, Hit List has the unique ability to allow reports to be generated from anywhere in the world via the Internet. This allows clients, traveling managers, webmasters, or sales people to get new reports whenever they want.

Step 1.   Enable Remote Reports

For security reasons, Remote Reports are not enabled by default. You should run Setup again to install the necessary components if you haven't already done so. Setup will install a small connector application to link your web server to the Hit List engine. This connector, a small Win-CGI application, is only invoked when someone is running a remote report; it does not interfere with your web server in any way. Setup will also give you the option of installing one or more copies of HITLIST.HTM. See *Step 2* below for details about using this HTML form.

Step 2.   Decide Which Type(s) of Remote Reports to Allow

Simple Reports

Remote users can generate simple reports just by running any web browser, filling out the HITLIST.HTM form and clicking OK. They do not need to buy any additional software or do anything complicated. This is a good solution for ISPs who want to allow their clients to run their own rudimentary reports, usually in addition to fancy reports that the ISP would routinely generate. To use Simple Reports (generally called *Ad-Hoc* reports), you only need to enable Remote Reports (see above) and put the HITLIST.HTM form in a directory that your clients can access. You can customize this form to give your remote users as much or as little flexibility as you like. Simple Reports are not as flexible as Full Reports (below) and do not have fancy formatting.

Full Reports

You can also generate reports that are as sophisticated as any report that could be generated by someone sitting at the server itself. To do this, you need to purchase and install Hit List on the client machine. When Hit List is installed on a client machine, it will ask if it should run reports against local copies of the logs or use the server to generate them (Remote). If you tell it to use the server, it will ask for the URL to the Hit List application. For Microsoft IIS users, this would normally be in the form of *http://www.YourCompanyName.com/Scripts/RunHL20.DLL*. If you're running Hit List on an O'Reilly WebSite or Netscape server, the URL would look more like *http://www.YourCompanyName.com/CGI-WIN/RunHL20.EXE*.

You can also manually configure Hit List to run remotely at any time. Select <u>Options</u> from the Tools menu and look at the Mode section. Choose the *Launch a browser and run the report on the server* option. Fill in the **Application URL** with the appropriate URL (see above). Of course, you can switch between Local and Remote at any time.

When using Remote Reports, Hit List will generate a very complex "script" for the server to execute. That script is encoded as an HTML file. Hit List will launch your web browser with that script and, when you click OK on the form, it will submit the request to your web server. When your server is done generating the report, it will be sent back to the browser. Unfortunately, there is no way for your browser to be updated with status information as the report runs.

Step 3.   Security Issues

Simple Reports

Since Simple Reports are designed to be easy to use, anyone with access to the appropriate HTM form can generate a report. In some cases, that's a bit of a security concern. Therefore, we strongly recommend that you only allow authenticated users (username/password security) to access the forms. Since you might have several of these forms on your system, each should probably have its own unique username and password.

For security and simplicity reasons, Simple Reports are automatically filtered such that they only retrieve data that relates to the server that answered the request.   That is, if someone uses HITLIST.HTM from *www.VirtualServer1.com*, all data is automatically limited to *www.VirtualServer1.com*.   In this way, there is no way that a remote user, using the simple ad-hoc reports, can see data from any other virtual site you host.

This above system has a problem, however, in that it also means that your main web server must be used to generate the remote reports.   In a high-traffic site, you may want to use a secondary server to handle ad-hoc reports.   With the default Hit List configuration, that won't work because the incoming ad-hoc report request would be filtered by the IP address of the secondary server, usually providing the user with an empty report since the virtual server he's interested in probably had a different IP address than the secondary server running the report.   You can override this automatic filtering, however, if you are willing to allow the ad-hoc reports to specify the server IP or name by checking the Allow ad-hoc reports to specify domain and database from the Advanced tab of the Options dialog. Although this opens up a security hole, you can mitigate it somewhat by using HIDDEN fields on the HTML form to contain the appropriate domain and/or database information.. See the comments in the HITLIST.HTM form for details of how to specify virtual server and/or database information.

Full Reports

Unlike Simple Reports, Full Reports do not get automatically filtered by server IP.   This is necessary so that you can run the same reports remotely as you could from the server itself.   This also means that anyone with a copy of Hit List could potentially run reports on your server.   To guard against this, you should probably 1) Not put *RunHL20.DLL/EXE* in the default location.   Pick something obscure.   2) Put username/password protection on it.

Step 4.   Performance Issues

Obviously running Remote Reports will impact your web server's performance.   You can minimize this by having Hit List run as a low-priority task by selecting the *Low* option from the Priority area of the Advanced tab in Options. In this way, all other applications, including your web server, will have a higher priority than Hit List so they will remain responsive even when Hit List is running.   However, if nothing else is happening on your server, Hit List will run just as fast as when its using the *Normal* priority setting.

When running a Simple Report, there is no way that the user can force a database update unless no database exists. If there is no database or the database specified as the default for remote reports doesn't exist, Hit List has no choice but to create it.

Due to the bug in Microsoft Windows NT 3.51 regarding reverse DNS lookups, Hit List will never do lookups for ad-hoc reports when run on an NT 3.51 system even if the report requests it.   This automatic negation does not apply to Full Reports.

If you're using Simple Reports, you can ensure better report performance by removing options that are especially slow to calculate.   For example, *Visits* take much longer to calculate than *Requests* so you can simply remove the checkbox from the form to ensure that they're never calculated.   Or you can remove complex report elements like *Path thru Pages, Jumps within the Site, Previous Pages, Exit Pages, Single Page Visits*, etc.   You also might consider removing the check for looking up site names.   Remember that you can create as many different versions of the form as you like, making it possible to offer different people different options.   See the Performance Tips section for more ideas.

Report Element Filters

You can use filters to 'zoom-in' on particular information or to ignore things that aren't useful. Filters are normally applied to an entire report using the <u>Filter</u> tab of the Design window. However, you may also override the report-level filter for any particular report element. This makes it easy to create very comprehensive reports where different sections may focus on specific information.

Tip: Be sure to check the **Include these objects in this report element** section to confirm that at least one type of object is checked. If not, the report element will produce an "Not enough information…" error.

Tip: Be careful when mixing report-level filters and element-level filters because element-level filtering overrides all report-level filters, effectively nullifying them. If you have a report with elements that make use of element-level filtering (the built-in *Total Number of HTML Pages* element, for example), be sure to mirror any report-level filters on each element that overrides the report-level filtering. For example, if you set a report-level **Server name or IP** filter to *123.456.789.123*, you must copy this setting into all elements that override the report-level filter.

See the report-level <u>filter</u> Help section for more information.

Using Hit List on Your Intranet

Hit List has many features that simplify working within an intranet environment.

1.   Proxy Support.   Hit List can <u>resolve URLs into more friendly page titles</u> and display those in addition to or in place of the URLs.   If accessing web pages requires a proxy server, simply enter the name of the appropriate proxy server in the <u>Advanced</u> tab of the Options dialog box.

2.   Site Name Groups.   You can easily assign names to specific IP addresses or address patterns using the <u>Site Names Groups</u> tab of the Database Manager.   For example, you could create a *North America* pattern of *123.456.789.** and a *Europe* pattern of *987.654.321.**.   Then, when running reports that show site Visitors, you would see *North America* and *Europe* listed instead of their IP numbers.

3.   FTP.   Hit List can automatically retrieve logs files from another machine (often a Unix machine) simply by prefixing *FTP:* in front of the log file name field in either the global <u>Options</u> dialog box or in the <u>Log File</u> tab for a specific report.   Hit List will first attempt a normal FTP copy then, if that fails due to your firewall, will attempt a passive connection.   There is no way to specify a FTP proxy server, however.

4.   Reverse DNS Lookups.   Hit List will normally test for the existence of a working DNS server by attempting to resolve the IP of *www.microsoft.com* and, if that fails, *www.netscape.com*.   If both of these fail, Hit List will assume that there is no working DNS and will abort the process.   This keeps Hit List from waiting forever for those who attempt DNS lookups from machines not connected to a network. If this is a problem because of your firewall, you have a few options:

A.   If you also require a proxy server to access web pages, Hit List will skip the DNS check above, assuming that only someone on a LAN or WAN would use a proxy server.

B.   If you have a working DNS but it cannot access external sites like www.microsoft.com, you can 'trick' Hit List by making an entry in your HOSTS file.   For Windows NT users, this file is located in *\winnt\system32\drivers\etc*. For Windows 95 users, this file is located in the *\windows* directory.   To create the 'fake' entry, add a new line to the bottom of the file and add an entry such as *123.345.789.123   www.microsoft.com*

C.   If neither of these work, you can also turn off all reverse DNS lookups (set in both the <u>Options</u> tab of any report and in the <u>Updates</u> tab of the main Options dialog box) and use Site Name Groups to associate IPs and computer names.

Database Manager/Ads

Hit List can create reports that show how well ads running on your site are doing.   These reports can be categorized by the name of the ad, by the HTML page that displayed the ad or by high-level ad groups.   After you define your ad groups, you might want to run the built-in *Advertising Report* or look at the several dozen report elements in the Toolbox that relate to ads.

Ad groups are collections of ads that you might want to analyze as a group.   Frequently this is used to consolidate ads from each company that runs ads on your site.   In this way, you can create a report that is limited to just showing ads from that company.   Or you can organize ads into categories based on the subject matter (News, Sports, Entertainment, etc.).   Each ad is a member of exactly one ad group.

Each ad must have at least one **impression URL** associated with it.   This is normally the URL of a GIF that represents the ad (e.g. */graphics/MyAd.GIF*).   It can also be the results of a CGI, ISAPI or NSAPI program (e.g. */scripts/ShowAds.DLL?SportsAd1*).   You may associate more than one impression URL with each ad in cases where the creative execution of the ad changes over time.   Wildcards are allowed but note that the first ? character in the impressions URL is not interpreted as a wildcard but as a separator between an application (like */scripts/ShowAds.DLL*) and its argument (*SportsAd1* in this example).

Ads may also have **click-through URLs** associated with them.   These are the URLs, on your web site, that visitors jump to when they click an ad.   This is normally a <u>redirect</u> program like Redirect.EXE/DLL that we provide or Redir.dll from Microsoft.

Tip: If your ads go directly to another web site, this jump is not recorded in your logs and, therefore, not tracked by Hit List.   You should consider using a redirect program instead because it will both make a record in your log and jump the visitor to the appropriate place.   You may associate more than one click-through URL with each ad in cases where the destination of the ad changes over time.   Wildcards are allowed but note that the first ? character in the click-through URL is not interpreted as a wildcard but as a separator between an application (like */scripts/redirect.DLL*) and its argument (*?http://www.AnotherPlace.com/,* for example).

Creating a New Group
Click the **New** button next to the **Ad Groups** list and enter then name of the new Group.

Modifying an Existing Group
Select the group in the **Ad Groups** list and click the **Change** button next to the **Groups** list.   Enter the group's new name.

Deleting an Existing Group
Select the group in the **Ad Groups** list and click the **Delete** button next to the **Groups** list.   This removes the group but does not delete any data from the database.

Creating a New Ad
Select the ad group that the ad belongs to then click the **New** button next to the **Ad Names** list and enter then name of the new ad.

Modifying an Existing Ad
Select the ad in the **Ad Names** list and click the **Change** button next to it.   Enter the ad's new name.

Deleting an Existing Group
Select the ad in the **Ad Names** list and click the **Delete** button next to it.   This removes the ad but does not delete any data from the database.

Adding a New Impression or Click-Through URL

Select the ad that the new URL belongs to from the **Ad Names** list then click the **New** button next to the **Impression URLs** and **Click-Through URLs** list.   Enter the complete URL or URL pattern.   If the URL requires a query (e.g. */cgi-bin/AdClicker.exe?SciFiAd5*), simply separate the query from the application with a question mark as shown in the example.

Modifying an Impression or Click-Through URL

Select the URL or pattern from the **Impression URLs** and **Click-Through URLs** list then click the **Change** button next to it.   Enter the new URL or URL pattern.

Deleting an Impression or Click-Through URL

Select the URL or pattern from the **Impression URLs** and **Click-Through URLs** list then click the **Delete**   button next to it.   This disassociates the URL from the currently selected ad but does not delete any data from the database.

See Also:
Object Types
Groups
Browsers
Operating Systems
Tools
Site Name Groups
App Argument Groups

Using Cookies

A cookie is simply an extra bit of information that a web server asks a web browser to transmit with every request. It is most frequently used to assign each site visitor a unique ID code so the web server can recognize when that same person (actually machine) comes back to the site. Such cookies are called *persistent cookies* because they persist even after the visit has ended since they are stored on the visitor's hard disk drive. Sites may also use *session cookies* that exist only during one visit. Cookies are very flexible so the content of the cookie (the *cookie value*) is entirely determined by the web server.

Hit List supports cookies in a number of ways. The built-in *Cookie Analysis* report provides a good starting point, showing you the number of unique cookies, a list of the most commonly seen cookies and, if you're using persistent cookies, the difference between calculating visitors solely on unique IP addresses versus using a combination of IP addresses and cookies.

If your site uses persistent cookies, you should turn the Use cookies when calculating visits switch on in the Updates tab of the Options dialog box. You should not use this option if your site has some other kind of cookie.

Hit List recognizes cookies in Microsoft IIS logs and standard NCSA combined (extended) logs.

Microsoft IIS logs should have the cookie after the referrer field and before the parameter field. For example:

```
157.55.69.103, -, 12/6/96, 7:08:22, W3SVC, WEBSERVER, 206.129.192.10, 10, 286,
14167, 200, 0, GET, /info/default.asp, Mozilla/2.0 (compatible; MSIE 3.0;
Windows 95), http://www.yourcompany.com/default.htm,
35bebd61b31211cfbdcd00c04fd611cf, -,
```

where

```
35bebd61b31211cfbdcd00c04fd611cf
```

is the value of the persistent cookie.

NCSA format logs should include the cookie as a quoted value after the user-agent field. For example,

```
205.218.110.166 - - [08/Dec/1996:15:02:10 -0800] "GET /info/index.html
HTTP/1.0" 200 14912 "http://www.yourcompany.com/index.html " "Mozilla/3.0Gold
(Win95; I)" "35bebd61b31211cfbdcd00c04fd611cf"
```

Hit List will not correctly process a log that includes cookies but not referrer and user-agent fields. However, you do not need to do anything special for Hit List to recognize logs in either of the above formats. You may freely combine logs that have cookies with those that do not.