

# LogProcess v1.0

## User Guide



DIGIEVO.CO.UK

## Index

Pages	Subject
	Introduction
3	<ul style="list-style-type: none"><li>• What is LogProcess?</li></ul>
	Installation & Configuration
4	<ul style="list-style-type: none"><li>• System requirements</li></ul>
4-5	<ul style="list-style-type: none"><li>• Installing the utility</li></ul>
6	<ul style="list-style-type: none"><li>• Configuring LogProcess</li></ul>
7	<ul style="list-style-type: none"><li>• Configuring IIS</li></ul>
8	<ul style="list-style-type: none"><li>• Scheduling LogProcess to run</li></ul>
	Management
9	<ul style="list-style-type: none"><li>• Adding sites to be process</li></ul>
9	<ul style="list-style-type: none"><li>• Monitoring LogProcess' operation</li></ul>
9-10	<ul style="list-style-type: none"><li>• What LogProcess produces</li></ul>
	Current Known Issues
11	<ul style="list-style-type: none"><li>• Issues</li></ul>

## **Introduction**

### **What is LogProcess?**

In a nutshell: *LogProcess is a simple utility designed to automate the use of AWStats in a Windows based hosting environment.*

AWStats is an excellent web server log file analysis suite that takes log files in a standardised format and produces very useful (and thorough) graphical HTML analysis pages. AWStats is written in PERL so you will need a suitable PERL environment to use it.

For more information on AWStats;

- <http://awstats.sourceforge.net/>
- <http://www.activestate.com/>

### **Problem:**

The basic procedure to get up and running with AWStats on Windows is very simple; all that you need do is download and install ActiveState (or any other) PERL and download and extract AWStats into a folder.

To generate a report all you have to do is tailor a text based configuration file to suit your needs and run AWStats passing the path to a log file as an argument. AWStats will then produce a database file containing the analysis. Another AWStats utility is then used to build the HTML report pages.

Whilst this all works fine I manage a large number of websites on several Windows servers and have various automation systems that manage the processes of setting up new sites from creating folders to setting up DNS and IIS. Ideally I needed a single script/program that could be executed periodically that would have the following capabilities;

1. Process any number of domains without having to maintain a list
2. Automatically generate an AWStats configuration file suitable for sites the first time they are processed
3. Merge any number of daily log files into a single log file that AWStats can accept
4. Periodically move processed log files from their source locations where they are generated by IIS into an archive folder tree for moving to CDR etc.
5. Automatically build folder structures to contain the HTML reports on a monthly basis and also build an default.asp index page that links to all the months for which statistics have been generated.

LogProcess v1.0 achieves all of this with very little change to basic standard IIS setups.

## Installation & Configuration

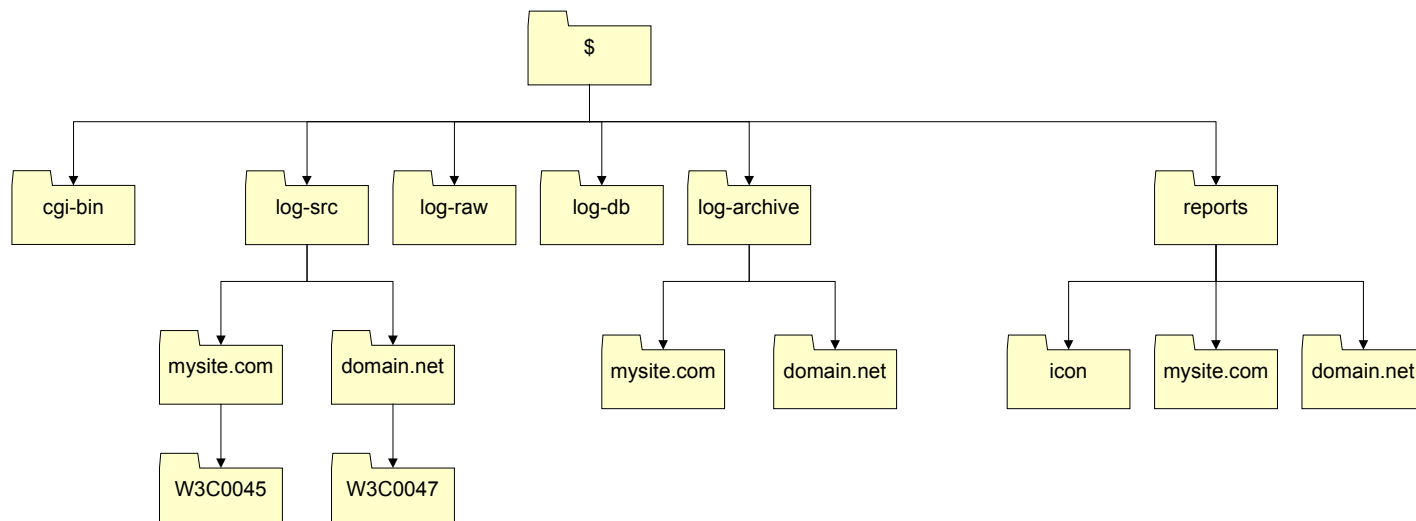
### System requirements?

LogProcess assumes the following minimum requirements are met; these are in addition to those required to run AWStats;

1. NT4, 2000, XP or .NET
2. 32Mb free memory
3. The MS .NET Framework

### Installing the utility

LogProcess uses a folder structure that looks like the following;



You must create this folder structure as follows:

- A) Create a root folder (\$). This folder can be anything that you like and should not be in a place where it is available to the world via IIS. If it is not possible to locate the root folder 'outside' or 'web scope' then use NTFS file permissions to ensure that the IUSR\_<machine> and IWAM\_<machine> accounts do not have access to folders other than 'reports'.

Preferably the entire folder structure should be placed in a private location on the server and a new website or virtual server created in IIS whose root folder is the 'reports' folder.

An alternative scheme if you do not want the stats for all sites to be available under a single URL would be to add a new virtual folder to each site that had stats to point to its own folder under reports and not map a general website/virtual folder to the reports folder at all. This scheme may be nature is more complex in terms of automation.

B) Create the level one folders. The purpose of these folders is as follows;

Folder	Purpose
cgi-bin	The AWStats application will be installed in here. Configuration and log files will be generated in here.
log-src	You must create a folder here for each domain that you wish to generate statistics for and set IIS to generate its log files in these folders.
log-raw	This folder is used for temporary files by AWStats.
log-db	This folder is used for database files by AWStats.
log-archive	Processed log files are periodically moved into this folder, as files are archived subfolders are automatically created to match the folder names in log-src.
reports	HTML Statistics reports are saved into this folder, as reports are generated subfolders are automatically created to match the folder names in log-src.

The subfolders shown in the diagram under *log-archive* and *reports* are in fact generated automatically by LogProcess as it needs them.

C) Copy the following files from LogProcess.zip into your cgi-bin folder;

- a. awstats.prototype.conf
- b. config.xml
- c. logprocess.exe

D) Follow the AWStats installation instructions placing all the required AWStats files into the cgi-bin folder except for the 'icons'.

E) Create a second level folder called 'icons' in your 'reports' folder and copy the AWStats icons into this folder.

F) Create a standard user account on the NT server called AWStats and set a safe password.

G) Secure the folder structure

The following NTFS permissions are appropriate for the folders;

	IUSR_<machine> IWAM_<machine>	AWStats	Administrator, etc	IIS Service A/C
cgi-bin	No access	RW + Execute	Full Control	RW + Execute
log-src	No access	RW	Full Control	RW
log-raw	No access	RW	Full Control	R
log-db	No access	RW	Full Control	R
log-archive	No access	RW	Full Control	R
reports	R	RW	Full Control	R

## Configuring LogProcess – Overview

To configure LogProcess to run the first time the following things need to be done;

1. Set the paths to the PERL, AWStats and the level one folders in config.xml
2. Make any desired changes to awstats.prototype.conf
3. Set IIS to generate logs in the correct format

## Configuring LogProcess – Config.xml

You can open the config.xml file into notepad or any application that can edit text or xml files.

### **Sample config.xml**

```
<LogProcess>
<SourceRoot>c:\webroot\awstats\log-src</SourceRoot>
<RawTemp>c:\webroot\awstats\log-raw</RawTemp>
<ReportRoot>c:\webroot\awstats\reports</ReportRoot>
<CGIBIN>c:\webroot\awstats\cgi-bin</CGIBIN>
<PERL>c:\perl\bin</PERL>
<Archive Day="1">c:\webroot\awstats\log-archive</Archive>
</LogProcess>
```

!! You must not change the structure of the XML in any way or LogProcess will not function correctly.

Update the first four lines to the correct paths to your level one folders that you created, the fifth line should be the full path to the PERL executable (shown path is correct for ActiveState PERL).

The archive element allows you to specify the day of the month on which you would like processed log files to be moved into the archive folder as well as the path to that folder. If you set Day="0" then this will disable archival completely.

## Configuring LogProcess – awstats.prototype.conf

LogProcess uses this file as a template to build new configuration files whenever it notices a new domain that it has not processed log files for yet. Within the file there are two lines which read...

```
<<LOGFILE>>
and
<<DOMAIN>>
```

... do not change these lines. Otherwise you can make any changes that you need to ensuring that you follow the AWStats documentation.

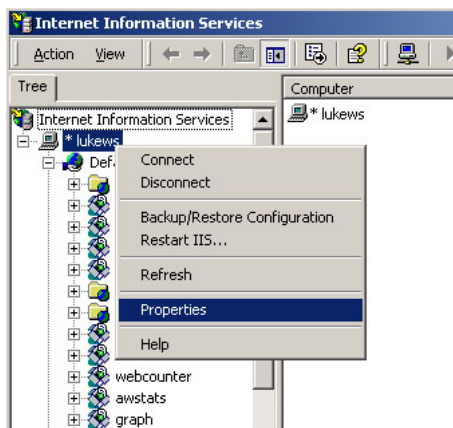
Note that once you are up and running changing this file will not change configuration files for domains that have already been processed at least once. For the new changes to take hold for existing domains you will need to delete their configuration files that LogProcess generated so that it will re-generate them.

Note that the section entitled 'log format' is already correctly set up for Windows/IIS to use with LogProcess in the supplied awstats.prototype.conf file.

## Configuring LogProcess – IIS Setup

In order for IIS to support all of the features of AWStats (namely referrers and traffic transfer totals) some small changes need to be made to IIS.

*!! Please note these instructions may differ slightly for NT4*

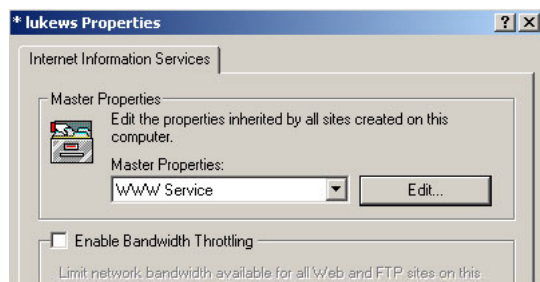


Launch the ISM (Internet Services Manager) and right click the webserver, then select 'Properties'.

The master properties window will then open. Ensure that 'WWW Service' is displayed and click 'Edit'.

A standard website properties dialogue will be shown but be careful with any changes you make as they will be immediately applied to all websites on this server.

The default tab to be displayed is 'Website'. Click the properties button under the 'Logging' section at the bottom of the tab.



The 'Extended Logging Properties' window will be displayed. Select the second tab 'Extended Properties'. Ensure that only the properties shown in the list below and exactly those are selected in your list before clicking OK on each property screen until you can exit the ISM.

Properties to tick: Date, Time, Client IP Address, User Name, Server IP Address, Server Port, Method, URI Stem, URI Query, Protocol Status, Bytes Sent, Bytes Received, User-Agent, Referer.

## Scheduling LogProcess to Execute

To actually have LogProcess process any new log files you must run LogProcess. You can do this interactively by simply double-clicking LogProcess and a standard Windows GUI will appear and give you options and visual feedback whilst process.

Alternatively you can use an **at** command or Scheduler in 2000/XP/.NET. The use of the windows scheduler is recommended over **at** as it allows you to specify the user account that you want LogProcess to run under (the service account you created during setup) – this affords us better security than **at** which will run tasks under the system account.

When launching LogProcess as a scheduled task either from **at** or the windows scheduler you can use the following switches;

- s Silent – Do not display the GUI (you will generally want this!)
- h Historic – LogProcess always processes any unprocessed log files but will only generate the current months HTML reports. Use this option to force generation of HTML reports for all months that data exists for in the 'log-db' folder.

In most cases you will simply want `-s` but may want to use `-h` the first time that you process logs to generate historic months reports.

### Testing

If you have followed all the steps above then AWStats/LogProcess should be ready to go. Follow the instructions in the next section to add one or more sites to be processed and check that there are at least a few lines in the log files and double-click LogProcess to run it interactively with its GUI.

This should give you a reasonable indication of whether things are working correctly.



## Management

### Adding sites to process

To add a new site to be process needs only two simple steps. For this guide I will add a site whose domain is mysite.net

1. In your 'log-src' level one folder create a folder for the domain called 'mysite.net' – it is important that this name matches the domain name as the name of the folder is used in the process of creating the AWStats configuration file.
2. Using ISM right click the site that is being added and select 'properties'. Click the properties tab under the logging section. IIS allows you to specify a folder in which it will place its log files, you can change this by using the browse button – select the folder you created in step 1.

IIS Will actually create a folder called W3SVC4 (where the '4' is the 4<sup>th</sup> website configured on the server) and then place daily log files within that. LogProcess can cope with this and will actually search for the W3SVCx folder within the folder that you created in step 1 for it's log files.

Note: Although it is possible to tell IIS to use the same log file for periods longer than a day this is not necessary as LogProcess will combine the multiple files where they exist. No application (including LogProcess) can access the file that IIS is currently writing to so this is another key reason for having daily files.

### Monitoring LogProcess' operation

When ever LogProcess is run it will append all of the output that would be generated by the AWStats tools as well as some of its own messages to a log file called 'audit.log'. If for some reason your stats don't appear to be generating properly then you can check this file for problems.

If you have sites that are particularly busy or are doing full DNS lookups then you ought to check the file to see how long each days processing takes. I recently discovered a site on one of my servers taking 3 ½ hours to process its files – the owner promptly lost DNS lookup in their stats!

### What LogProcess Produces

When LogProcess runs it co-ordinates AWStats into producing many files. First any unprocessed files from each domain within 'log-src' are combined into a single file per site which is stored in 'log-raw'.

AWStats is then called on each of those combined files one at a time. In this process a file is either created or amended in 'log-db' additionally depending on the settings within the domains configuration file a dnscache file may be generated also in 'log-db'. After this process the temporary file in 'log-raw' is removed.

Next AWStats is called again to produce the HTML reports for each domain. This process reads from the data files in 'log-db' which AWStats produced and saves a number of HTML pages into a folder decided upon by LogProcess.

The folder the report is saved to will be created (if it doesn't already exist) under a matching folder in 'reports' for the source folder in 'log-src'. The name of the folder will be in the format of MMYYYY as appropriate for the report.

Next all default.asp index pages in each domain folder within 'reports' are erase and regenerated to take account of any new months that have been added to the domain.

All log files that have been processed in 'log-src' are renamed from file.log to file.old, this prevents LogProcess from processing the files more than once although this is primarily a performance tuning thing as AWStats would ignore duplicate entries anyway.

The last step is archival. If the value for archival in config.xml is anything other than 0 and the day of the month is the same as that value then archival begins. In archival all .old files within the 'log-src' folder are moved over to the 'log-archive' folder and the correct folders will automatically be generated. It's fine to complete empty the 'log-archive' folder once you have copied it's contents off to CDR or wherever else you archive the logs to.

**Current Known Issues****Version 1.0**

Number	Reported	Description	Workaround/Resolution	Fix Status
LP/001	22-03-2003 Chris Walters	<p>Space in folder names within log-src causes LogProcess to produce no output.</p> <p>Cause: Log process spawns CMD.exe whilst processing and executes PERL within the shell instance to invoke AWStats. The parsing of the command line arguments appears to be subtly different between interactive instances of CMD.exe and indirected (piped) instances of CMD.exe with .NET.</p>	Do not use spaces within folder names!	V1.1 Will check folder names during processing and report to audit log.