Sitecore StatCenter Module

Administrator's Guide

Installation and configuration advice for administrators and developers

Sitecore.
Content Delivery

www.sitecore.net
info@sitecore.net

Sitecore Corporation
support@sitecore.net
+45 70 23 66 60

Contents

Chapter 1

Introduction

StatCenter for Sitecore CMS is a module developed to display web statistics from within Sitecore. The module uses the IIS log files and stores data structured in a Microsoft SQL Server database. This chapter is an introduction to the module containing features, confirmations, prerequisites and an introduction.

1.1 General Introduction.

One of the most prominent features of StatCenter is its absolute integration with Sitecore, which makes it possible to generate statistics based on friendly URLs, id-based URLs etc. (this is performed invisible to the end-user). StatCenter provides much needed functionality for administrators and editors, though it does not support the data mining functions provided by products like WebTrends.

1.2 Hardware and Software Requirements

StatCenter requires the same hardware and software as Sitecore CMS. Furthermore StatCenter can be run on Microsoft SQL Server. StatCenter 3.4.0 is built for the .Net framework 2.0.

1.3 Prerequisites

Sitecore CMS: Sitecore CMS 6 or later.

License: The module conforms to the general Sitecore license interface.

Multilingual: The module implements the functionality for multiple languages and is translated into Danish and English.

Design: This module uses the same design guidelines as other ISV modules for Sitecore. This design has been validated by Sitecore.

Package: The module is packaged for Sitecore CMS.

Configuration: The module is configured through an interface in the client.

Chapter 2

Installation and Configuration Instructions

This chapter deals with the instructions necessary to Install and configure StatCenter.

2.1 Installation and Configuration of the Sitecore Package

To install StatCenter follow the steps described below.

2.1.1 Windows Vista and the IIS UI

2.1.1.1 Symptoms

Windows Vista did not ship with a user interface to configure logging in the IIS Server Manager.

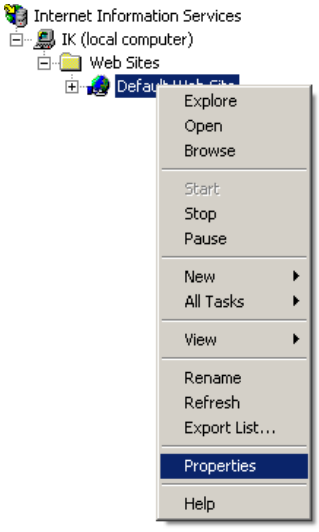
2.1.1.2 Resolution

This download provides source code and binaries to provide you with the configuration interface you need: http://www.iis.net/downloads/default.aspx?tabid=34&i=1328&g=6

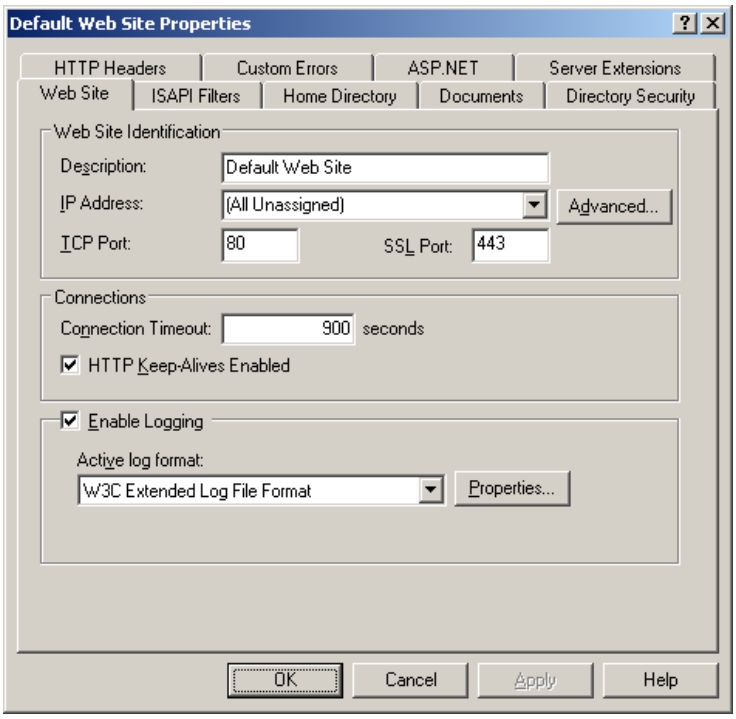
Important: to avoid corruption of the log file please see the following link : Internet Information Services 7.0 removes custom values in the Web.config file.

2.1.2 Setup logfiles in the Internet Information Server

To setup the log-files you must open the IIS-manager and open the properties for the Default Web Site (shown below).



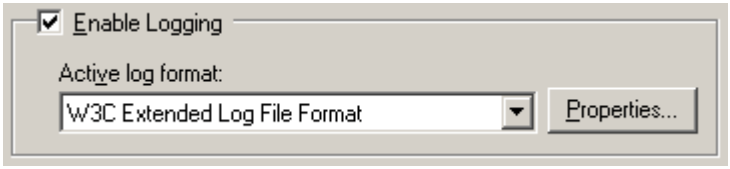
Under the Web Site tab make sure the Enable Logging checkbox is checked then in the drop down list labelled Active log format select W3C Extended Log File Format then click OK (as shown below).



Note: The W3C-format is the only format supported by StatCenter.

2.1.3 Configuring Extended Properties.

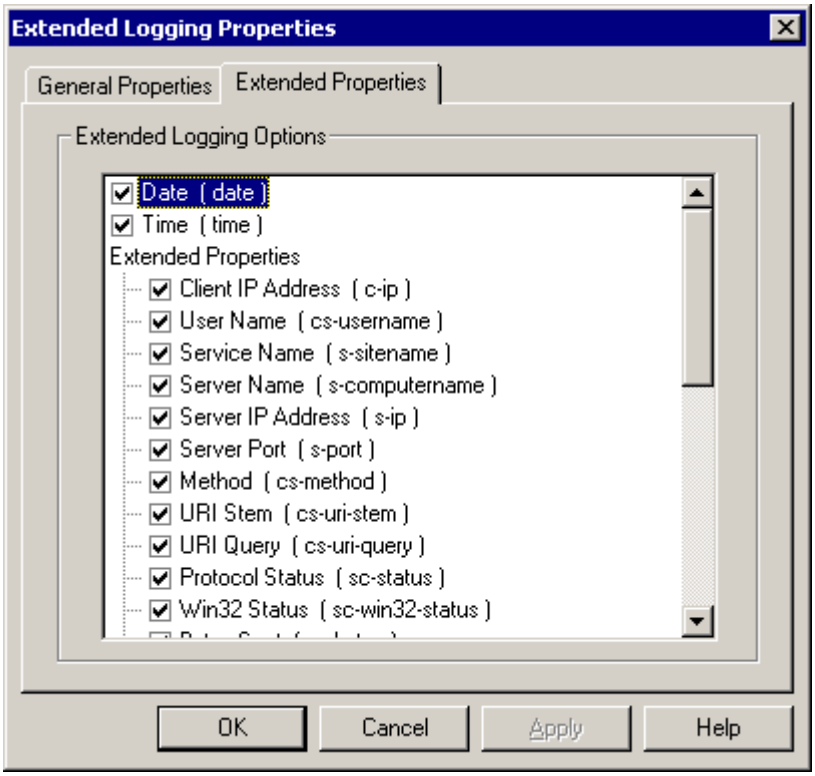
Select the Properties button in the Enable Logging section (as shown below)



Select the Extended Properties tab and make sure the settings conform to the following table.

Item	Description	Exact name
Date & time	used for period based searches	date; time
Client IP Address	used for tracking users host-name	c-ip
URI Stem	used for identifying the page visited	cs-uri-stem
URI Query	used for identifying the parameters transferred to the page	cs-uri-query
Host	used for multisites	cs-host
Cookie	used for identifying user and integration towards mailing list	cs(Cookie)
Referrer	used in later version for identifying most popular referrer	cs(Referer)

User-Agent	used in later versions for identifying most popular browsers	cs(User-Agent)



2.1.4 Install the package

Install the Sitecore-package for StatCenter. The package will install assemblies and the module used to present statistical data within the Sitecore client.

2.1.5 Further configuration

After the package installation you should grant write access rights for _Sitecore_root_path_/Sitecore modules/shell/StatCenter/ folder to the ASPNET user if you use IIS 5.0 (Windows XP), or to the NETWORK SERVICE user if you use IIS 6.0.

Furthermore you should enable the Sitecore KeepAlive function (a scheduled task) if you want to perform scans on old log files which might be a time consuming process

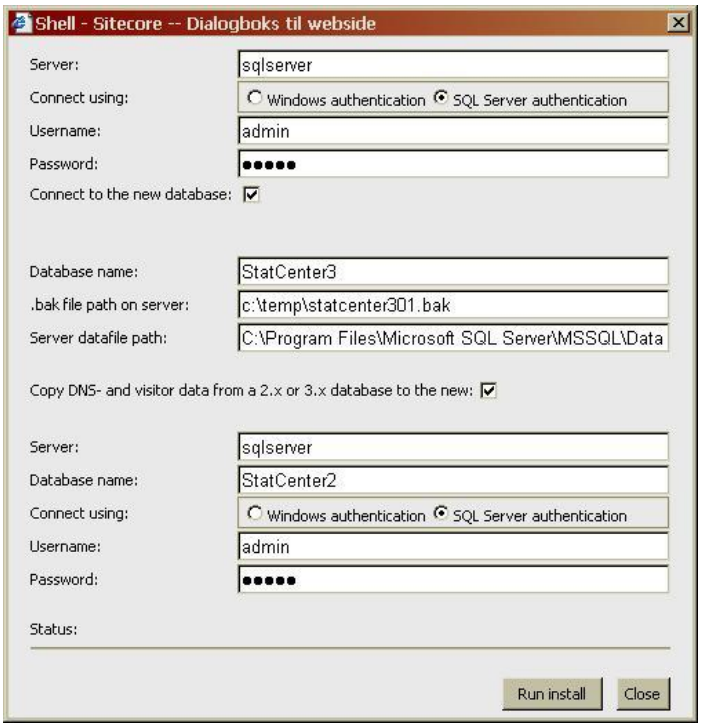
Install the StatCenter database, via the install dialog in the File menu.

Configure StatCenter, via the Configuration dialog in the File menu.

2.2 Installation and configuration of the StatCenter Database.

2.2.1 Installation

Start StatCenter, close the popup saying that StatCenter is not configured, and open the dialog File » Database administration » Install. Fill and run it.



1) First fill the connection fields to the sqlserver (Server, Connect Using. Username and Password), where you want to install the database.
2) Mark Connect to the new database if you want StatCenter to connect to the new database immediately.
3) Database name gives the name of the new database.

   Note: If the database already exists it will be overwritten!
4) Put the database backup file

   \Sitecore modules\shell\statcenter\database\statcenter301.bak

   in a folder on the sqlserver, and set the path in .bak file path on server.
5) To copy DNS- and visitor data from an old StatCenter database, check the Copy DNS- and visitor data from a 2.x or 3.x database to the new checkbox, and fill in the connection fields to point the source database.
6) Click Run Install.

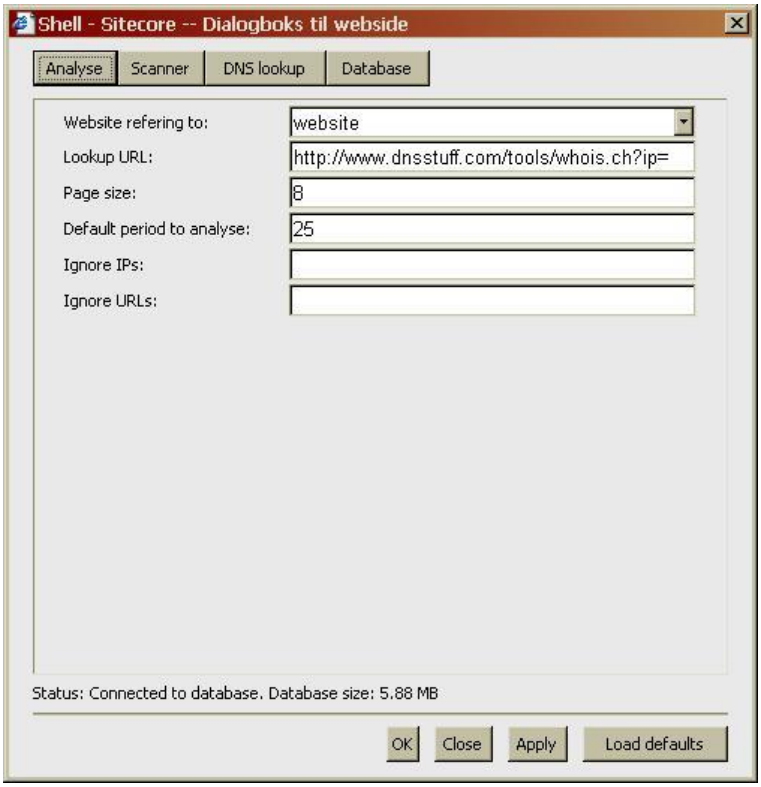
2.2.2 Configuration of the Database Installation

Open the configuration dialog from the File menu.

In the status area you will be able to see whether the database connection could be established and if so how much space it currently takes up.

If StatCenter has just been installed, the default configuration is a good starting point. Load it by clicking the Load defaults button at the bottom of the dialog box.

Under Analyse you can administer the following settings:

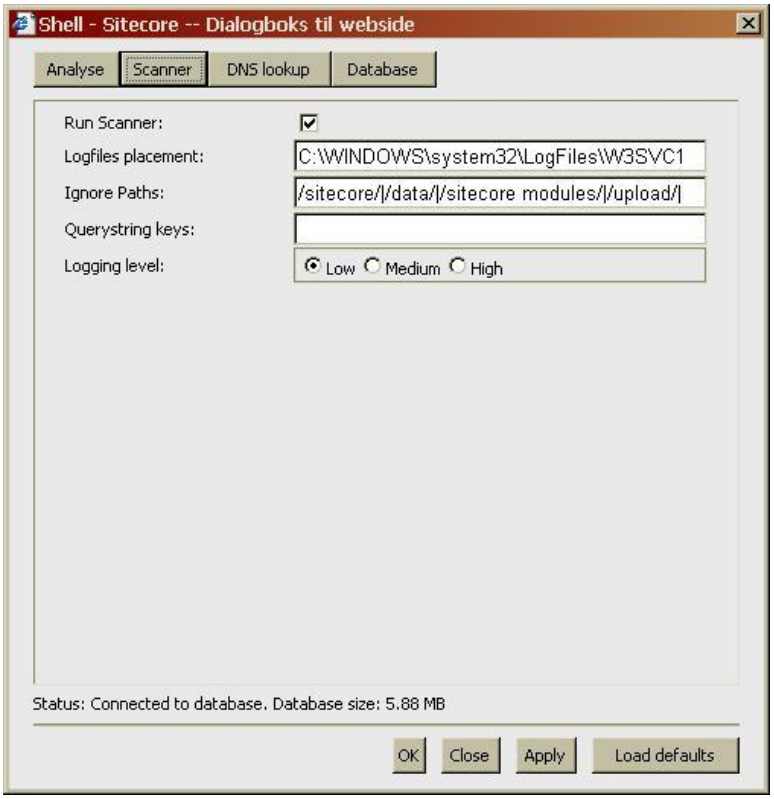


- Website referring to.
  This is used when a Sitecore solution goes from a single site to a multisite configuration. Before there was only one site, namely the website, but now more sites are added, and in order to keep using the website data different possibilities arise. The site from the single site is named website in the multisite solution. The following options exists in relation to specifying the behavior of website when using a multisite solution:
  - No site: website is not used in the solution at all (the data is scanned, but it is not showed in the statistical views) and the site is not appearing to the user.
  - Website: It is used as a single separate site.
  - All other items are the sites in the current solution. If one of these is chosen, website will be counted as part of the chosen site.
- Lookup URL.
  This specifies the path to a service which can be used for the Lookup externally function.
- Page size.
  This is the default value defining how many items to display in lists viewed by the users.
- Default period to analyze.
  This is used to define the default number of days to perform statistical searches for
- Ignore IPs.
  IP-addresses that will be ignored when searching. This can be used for ignoring hits from

googlebots etc. Separate with pipes **"|"**. Note that each user can add their own IPs to ignore, but the values here are ignored globally.

- Ignore URLs.
  URLs that will be ignored when searching. This can be used for ignoring hits from googlebots etc. Separate with pipes **"|"**. Note that each user can add their own URLs to ignore, but the values here are ignored globally.

  Under Scanner you can administer the following settings:



  - Run Scanner.
    Turn off the scanner by unchecking this field.
  - Log files placement.
    The full paths to the folders of the log files that are to be scanned. If you are not sure where they have been placed look in the IIS Manager at the properties for log files.

    Note: It is possible to state more than one folder for log files, just separate the paths with pipes **"|"**.
  - Ignore paths.
    Paths that will be ignored when searching. We recommend that you ignore /Sitecore, /data and /Sitecore modules. Separate paths with pipes **"|"**.
  - Query string keys.
    In some Sitecore solutions different query variables are used to find the specified page to load. If the solution e.g. differentiate pages the following way: /default.aspx?article={41B4958B-E25A-4224-A7FF-7ABCE477CAF0} then the page entered is not default.aspx, but the page with the GUID equal to the value of the article. Different query variables can be used, e.g. article, post etc.. Note the following precedence when deciding which Sitecore item to map the entries to:

- First the variables written in the textbox in the order from left to right, as in "queryVar1|queryVar2|…..|queryVarX".
- Then per default the queryVariable "path" (is used no matter if you write it or not!)
- Finally the path used to access the page (e.g. default.aspx above).

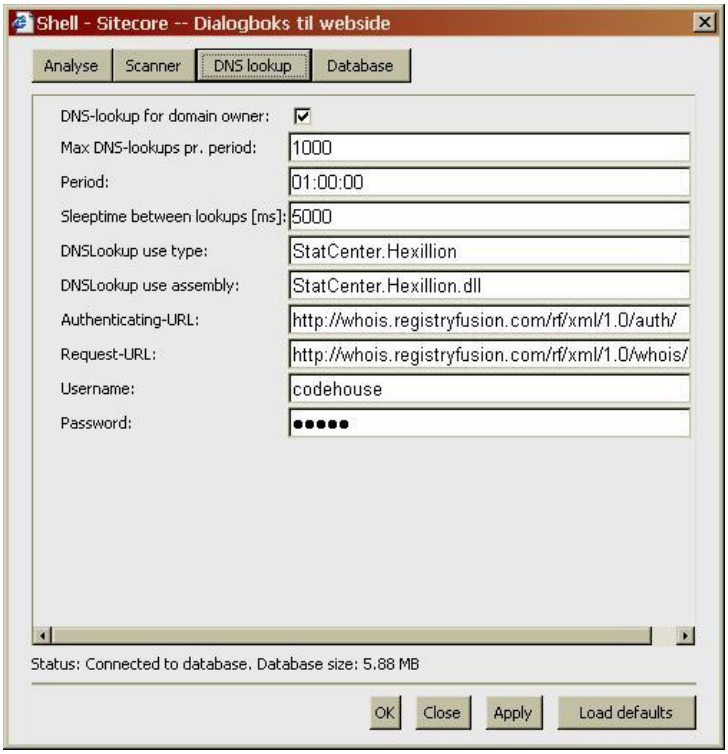
- Logging level
  Set the level of logging:

  Low: Start, files scanned and end are logged.

  Medium: Same as Low but including the line scanned and time between each scanner status is logged.

  High: Same as Medium but including the time used for DB, the file calls, and the longest DB call are logged.

Under DNS lookup you can administer the following settings:

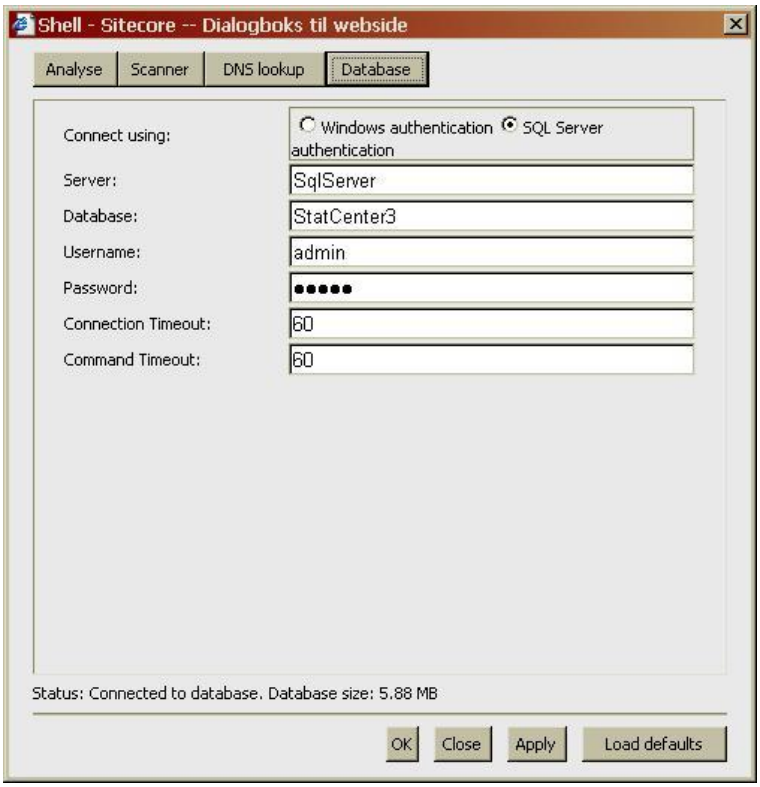


- DNS-lookup for domain owner
  Whether or not to look up the owners name using reverse DNS. This might be very time consuming for the scanner and thus may decrease performance dramatically. Since the various whois-servers around the world have set boundaries as to how many request a specific IP-address may perform per day we propose following setup. Please read the WhoIs section for information on using Hexillion.
- Max. DNS-lookups pr. Period
  Set how many DNS lookups can be performed in the time span given in Period.

- Period
  How long time a DNS lookup session is allowed to run.
- Sleeptime between lookups [ms]
  The amount of time between DNS lookup sessions in milliseconds. Must be set to at least 2000 to work properly.
- DNSLookup use type
  The class to use for DNS lookups. StatCenter.Hexillion is the standard class.
- DNSLookup use assembly
  The assembly to use for DNS lookups. StatCenter.Hexillion.dll is the standard assembly.
- Authentication-URL
  http://whois.registryfusion.com/rf/xml/1.0/auth/ is the standard.
- Request-URL
  http://whois.registryfusion.com/rf/xml/1.0/whois/ is the standard.
- Username
  The username given by the provider.
- Password
  The password given by the provider.

NOTE: It is recommended that you set up the ReverseDns task only after having scanned all existing log-files. The reason for this is that Sitecore's scheduler will otherwise instantiate multiple threads because attempting to perform reverse DNS resulting in DNSLookupTime is being overruled. When the log-files have been scanned you should go to the status page and perform and calculate the preliminary frequency. If totalDNS saved is 12.000 and DNSLookupTime is 2.000 you should allocate approximately 7 hours in frequency. After the first scan has been performed you can set the frequency to a more appropriate period – for instance every five minutes.

Under Database you can administer the following settings:

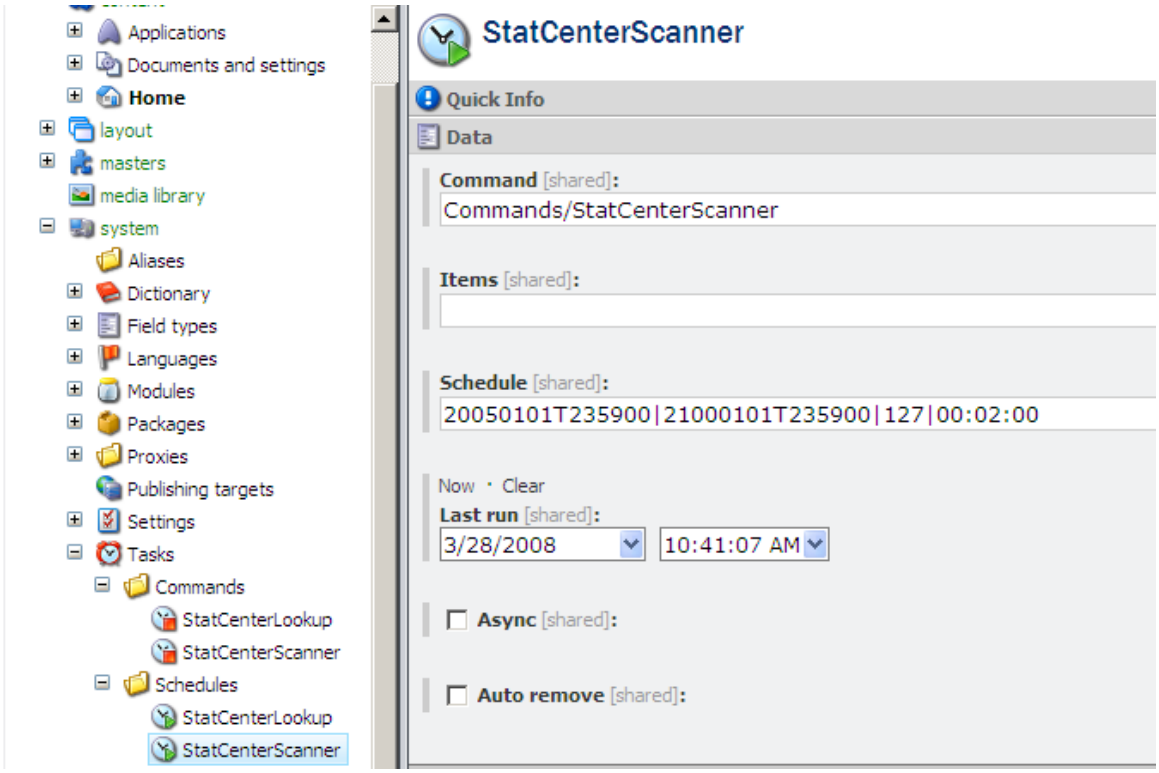


- Connect using
  Choose if the database connection is made through Windows authentication or SQL Server authentication.
- Server
  The name of the server instance.
- Database
  The name of the database.
- Username
  The username to use for accessing the database.
- Password
  The password for the specified user.
- Connection Timeout
  Sets the time before timeout when StatCenter connects to the database.
- Command Timeout
  Sets the time before timeout for SQL procedures. If you have large amounts of data and experience problems when trying to view the statistics, try increasing the value of this setting.

2.2.3 Scheduled tasks

StatCenter uses two scheduled tasks to perform scanning of log-files and DNS lookups. These are located under /core/sitecore/system/Tasks/Schedules in the Sitecore client and are automatically installed with the StatCenter package.

Note: /core is not a node, it is the name of the database. The root item in Sitecore content is /sitecore. If you can see any other items (master, core or web), they are database names.



The scheduled tasks are configured by a pipe separated string as shown in the above image:

20050101T235900|21000101T235900|127|00:02:00

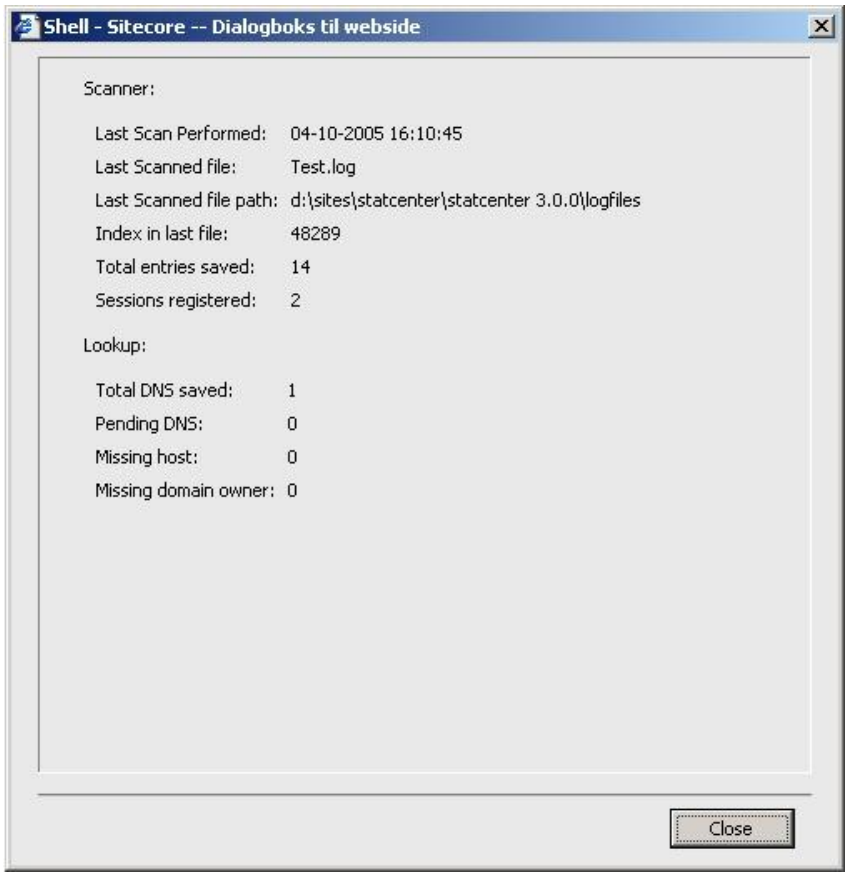
The first value is the start date for the scheduled task and the next value is the end date for the scheduled task. By default these dates are set so StatCenter will begin to perform scans and DNS lookups immediately after being configured and will continue indefinately.

The value 127 implies that the task will be performed every day of the week and the last value indicates how often the task will be performed using the format HH:MM:SS.

Note: It is NOT recommended to set how often the task will be performed to less than 00:01:00 (1 minute) for the StatCenterScanner task and 00:02:00 (2 minutes) for the StatCenterLookup task.

2.2.4 View status

Use the application status window to see how the application is running. This is opened from the File menu.



2.2.5 WhoIs

The use of WhoIs is integrated into StatCenter. This is setup in the Lookup task as described under Configuration.

Note: Performing WhoIs may be rather unpredictable since the WhoIs-service owners may block your IP if you are performing mass lookups.

This can be avoided by having a managed relationship with one or more of the WhoIs providers. This is rather difficult. Thus we have entered a partnership with the WhoIs-provider, Hexillion, and by simply buying a ticket book and registering your user name and password you will be able to use their functionality. Should you, however, wish to implement your own WhoIs service with your own functionality or integration to another WhoIs-provider this is also possible.

2.2.5.1 Using Hexillion

The RegistryFusion service from Hexillion [http://www.Hexillion.com/] is an advanced interface to the WhoIs system. It enhances the value of your statistics by providing identity and location information about your site's visitors.

For each site visitor, the RegistryFusion service contacts the correct WhoIs server, obtains a record, and distils the record into a useful form.

To sign up for the RegistryFusion service, please visit http://www.Hexillion.com/whois/order/ . Hexillion will provide you with a username and password for accessing the system.

2.2.5.2 Using you own class

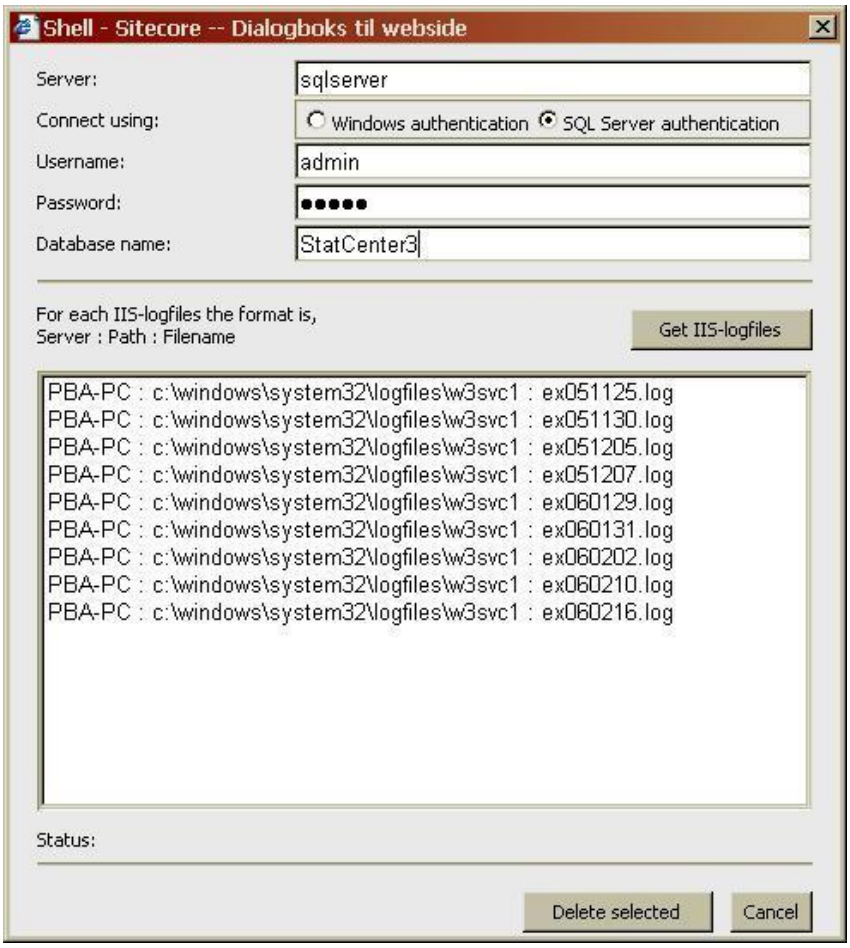
1) Create a new class and let it implement the interface StatCenter.IWhoIs
2) Make sure that Authenticated() returns true
3) In the function GetInformationByIP() you can write the code that return information about a domain. You should return the information in an instance of a StatCenter.WhoIsInformation object.

2.3 Deleting IIS Log Files from the StatCenter database

It is possible to delete IIS-Logfiles that are in the database. All entries in the database from the deleted IIS-Logfiles, are deleted in the same operation, so it will shrink the size of the database considerably.

It is done via the dialog File » Database management » Delete IIS-Logfiles.

StatCenter will scan the deleted IIS-Logfiles again, if these are not removed from the logfile folder specified in the configuration. It is advised to compress the IIS-Logfiles files and keep them for later usage.



Do the following to delete IIS-Logfiles:

1) Point to the StatCenter database.

2) Select the IIS-Logfiles to be deleted.

3) Click the Delete selected button.