Data*Spider*

Connects software to software

Getting started

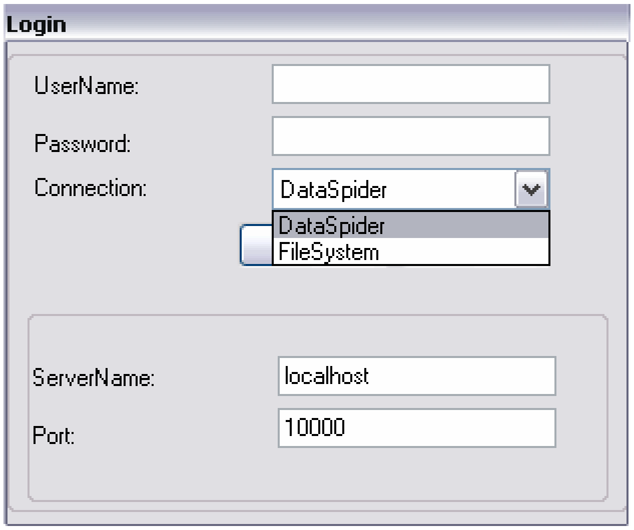
Installation

Run the installer. This will install the Data*Spider* service and the Data*Spider*GUI management console on your PC. Beneath the main installation directory a set of folders is created to accommodate the configuration files. Initially the configuration files for the main service and logging are put here. You can upload additional files later using the GUI.
Included in the setup is a SqlServerCE runtime. Install this if you want to log to the underlying database.
After installation has finished start the GUI to adjust the standard settings to fit your configuration. In the next section this is described in more detail.
Adjust the Data*Spider* service config file in the installation directory with a text editor to reflect the selected path chosen at installation.
When you're finished you can start the Data*Spider* service. In the log pane of the GUI you can monitor the proceedings. With the service running you can test the correct working of the services in relation to the programs using it.

Using the Data*Spider* GUI

(The GUI is still in its infancy. Sometimes you'll have to repeat the steps described below before the desired changes are committed. No validation is done on the entered data and don’t expect a confirmation dialog to be displayed first.)

Start the program from the start menu or the desktop shortcut. Log in to the program with the username and password provided by your vendor.



UserName:

Password:

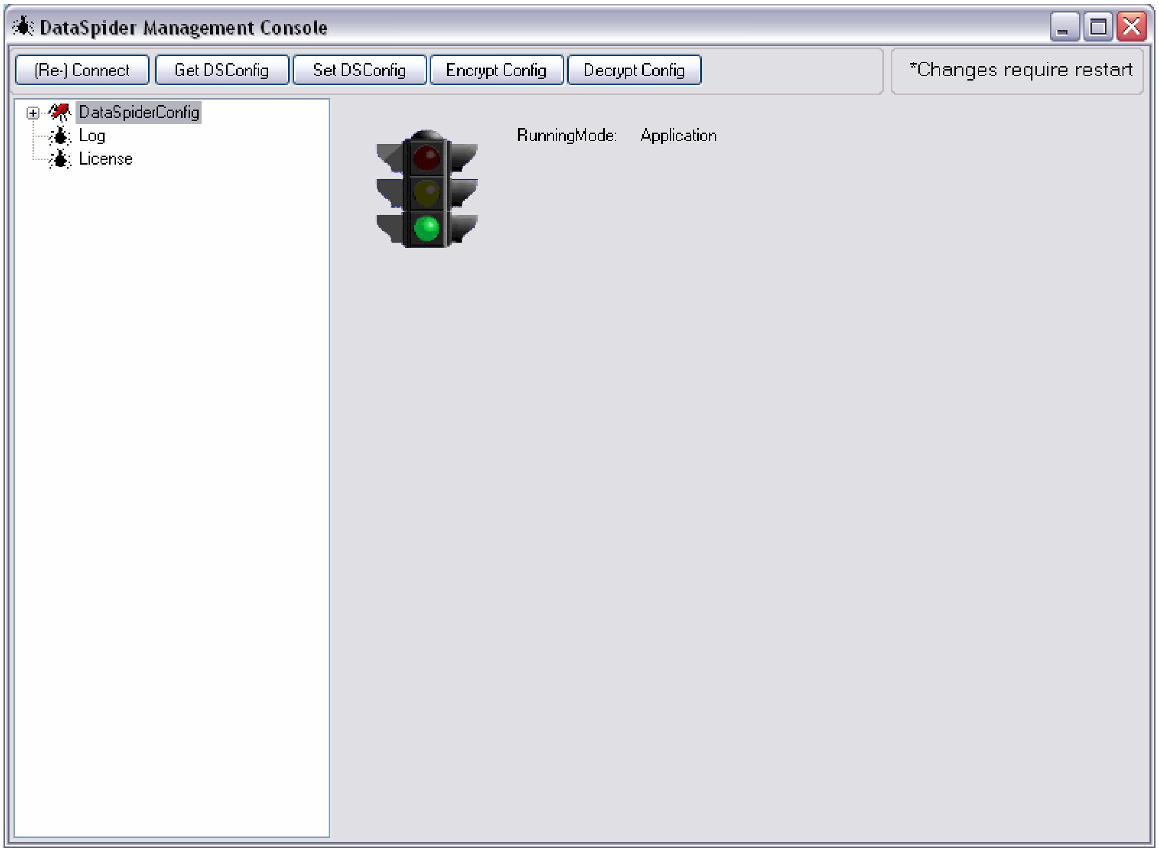
Connection: DataSpider

ServerName: localhost

Port: 10000

You can choose to connect to the Data*Spider* service or directly to the filesystem. (The first time you run the GUI after installation you connect to the filesystem first as the connection to the Data*Spider* service requires it to be started.) For this you either enter the name and port of the server the Data*Spider* service is running on (Can also be a remote PC) or enter the path to the DataSpider.config file. This file by default resides in %installationpath%/config/communicator.

After you logged into the console you are presented with the main screen:



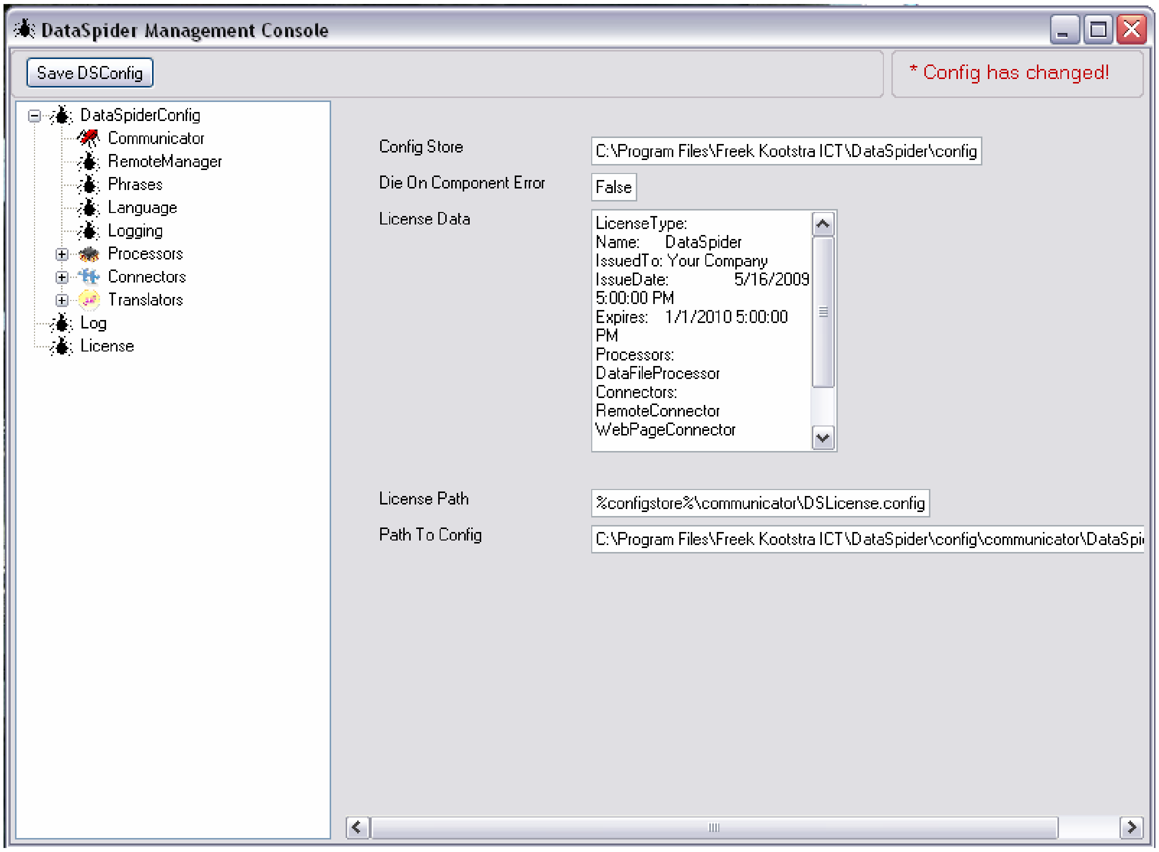
*Changes require restart

RunningMode: Application

- To switch between connections hit (Re-)Connect. You can also connect to another instance.
- To download the main configuration file in its entirety hit Get DSConfig.
- To upload the main configuration file in its entirety hit Set DSConfig.

From the tree on the left you can browse the several configuration items:

- Communicator: The main engine
  1. Config Store: Path to the folders with the configuration files
  2. Die On Component Error: Set to True the Data*Spider* service will halt when an error occurs in one of the running services. The default is False.
  3. License Data: Let's you view the contents of the license. Data*Spider* service will not start when the license is not valid.
  4. License Path: Path to the license file
  5. Path to config: The full path to the main configuration file.



* Config has changed!

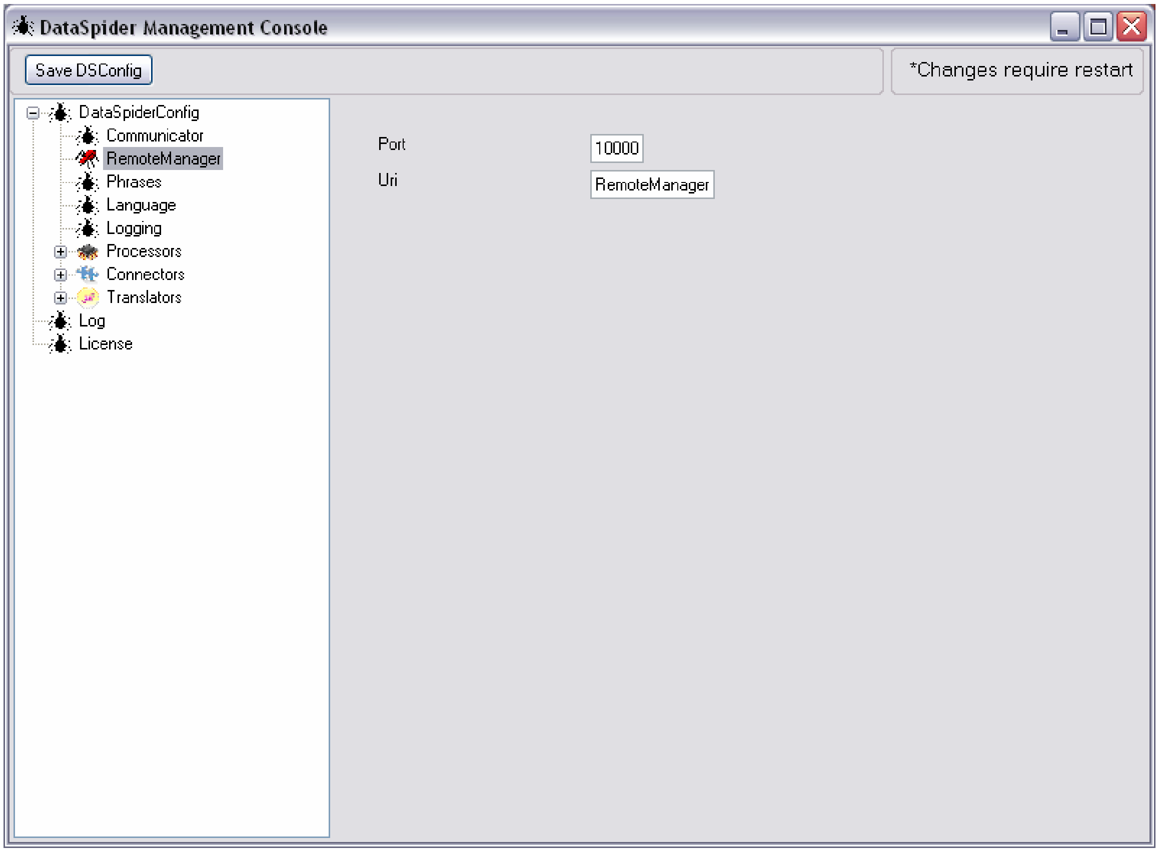
Config Store: C:\Program Files\Freek Kootstra ICT\DataSpider\config

Die On Component Error: False

License Data:
LicenseType:
Name: DataSpider
IssuedTo: Your Company
IssueDate: 5/16/2009 5:00:00 PM
Expires: 1/1/2010 5:00:00 PM
Processors:
DataFileProcessor
Connectors:
RemoteConnector
WebPageConnector

License Path: %configstore%\communicator\DSLicense.config

Path To Config: C:\Program Files\Freek Kootstra ICT\DataSpider\config\communicator\DataSpi



*Changes require restart

Port: 10000

Uri: RemoteManager

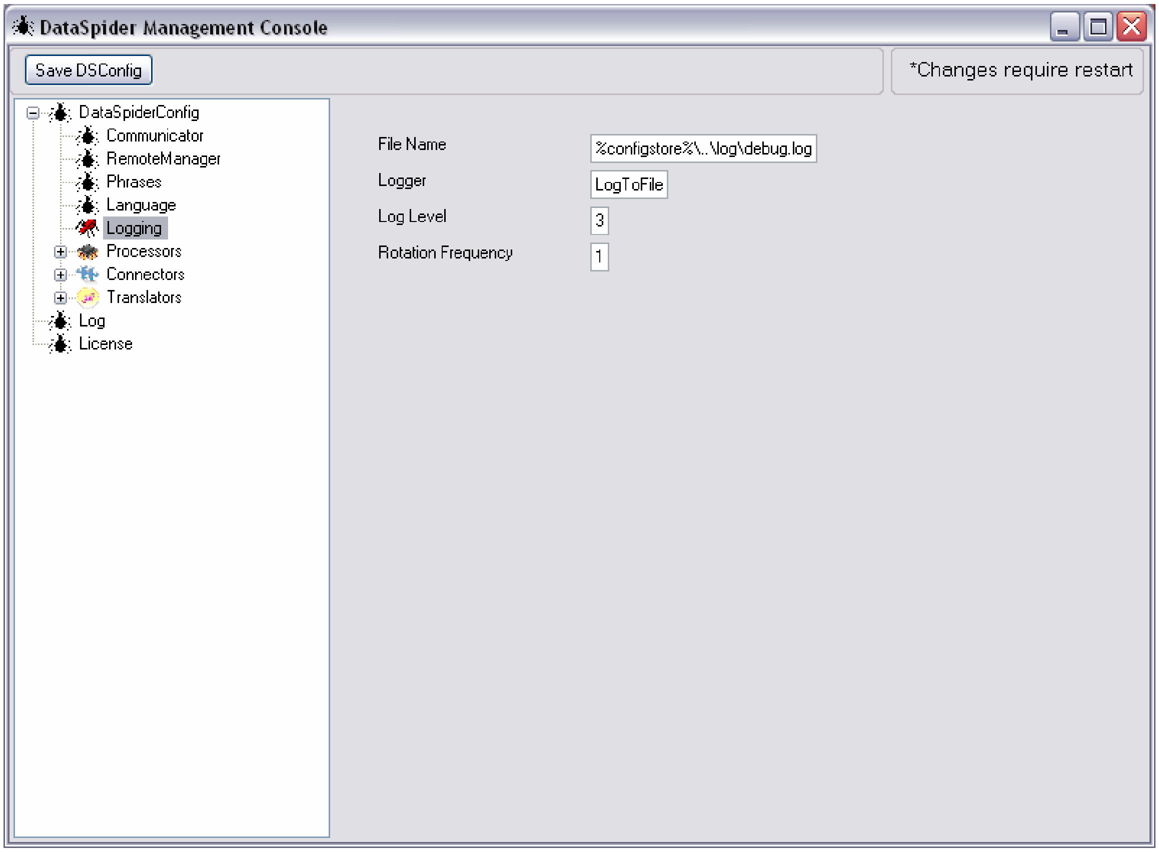
- Remote Manager: Provides access to the service
    1. Port to connect on
    2. Uri is always 'RemoteManager'

Port and Servername are taken from the login-screen so basically you don't change anything here unless you're unhappy with the Port number.

- Phrases: Used for multi-language support (Currently not used).

- Language: the language setting for multi-language support (Currently not used).

- Logging:



File Name %configstore%\..\log\debug.log
Logger LogToFile
Log Level 3
Rotation Frequency 1

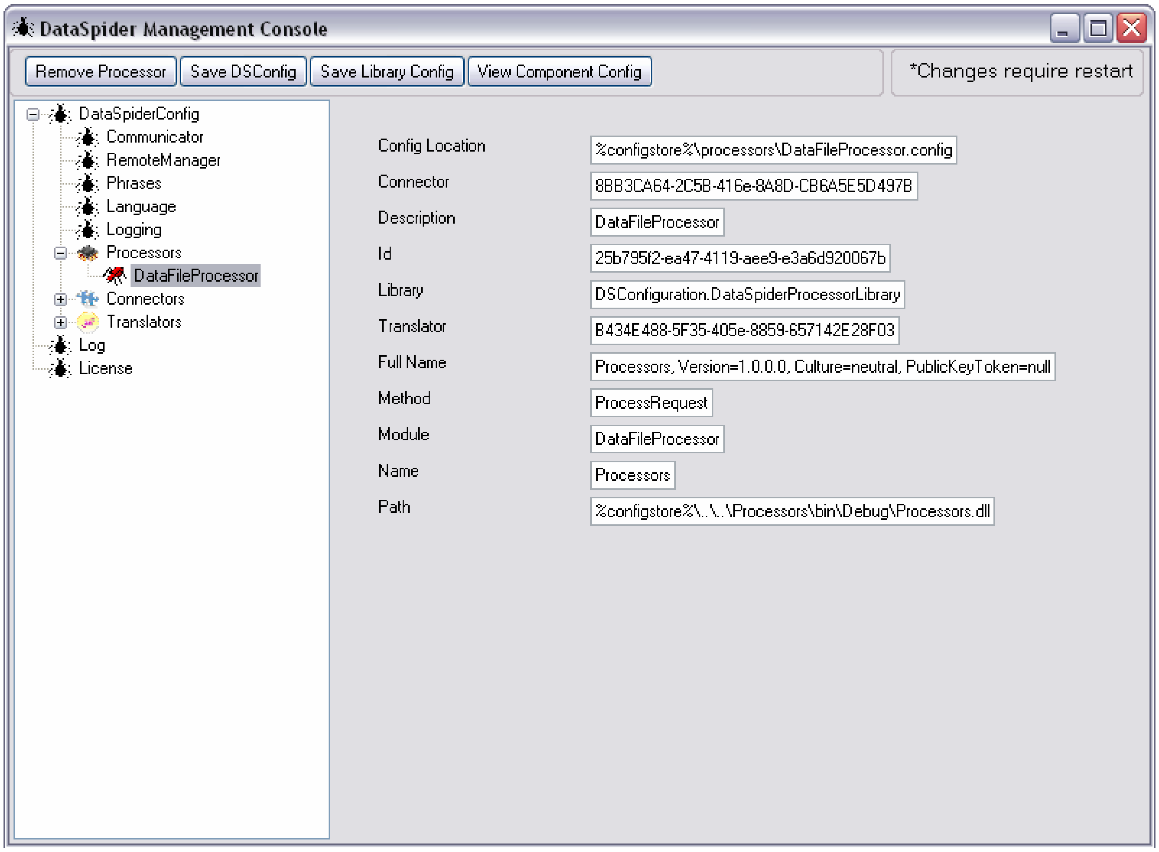
1. File Name: Name of the file for LogToFile or LogToSqlServerCE logger. When logging to Windows Eventlog the Filename setting is ignored.
2. Logger: You can choose between logging to flatfile (LogToFile), database (LogToSqlServerCE) or Windows Eventlog (LogToEventlog).
3. Loglevel:

    - Errors only = 0
    - Warnings = 1
    - Informational = 2
    - Debug = 3

4. Rotation frequency:
    - No Rotation = 0
    - Hourly = 1
    - Daily = 2
    - Weekly = 3
    - Monthly = 4

    Choose the rotation frequency based on the expected throughput and the loglevel for the logs not growing too big.

Processors: These are the services running in the Data*Spider* service. Each has its own set of configuration items based on the task they perform. Beneath is an example:



Config Location	%configstore%\processors\DataFileProcessor.config
Connector	8BB3CA64-2C5B-416e-8A8D-CB6A5E5D497B
Description	DataFileProcessor
Id	25b795f2-ea47-4119-aee9-e3a6d920067b
Library	DSConfiguration.DataSpiderProcessorLibrary
Translator	B434E488-5F35-405e-8859-657142E28F03
Full Name	Processors, Version=1.0.0.0, Culture=neutral, PublicKeyToken=null
Method	ProcessRequest
Module	DataFileProcessor
Name	Processors
Path	%configstore%\..\..\Processors\bin\Debug\Processors.dll

For the module to be loaded into the Data*Spider* service successfully it has to be configured here:

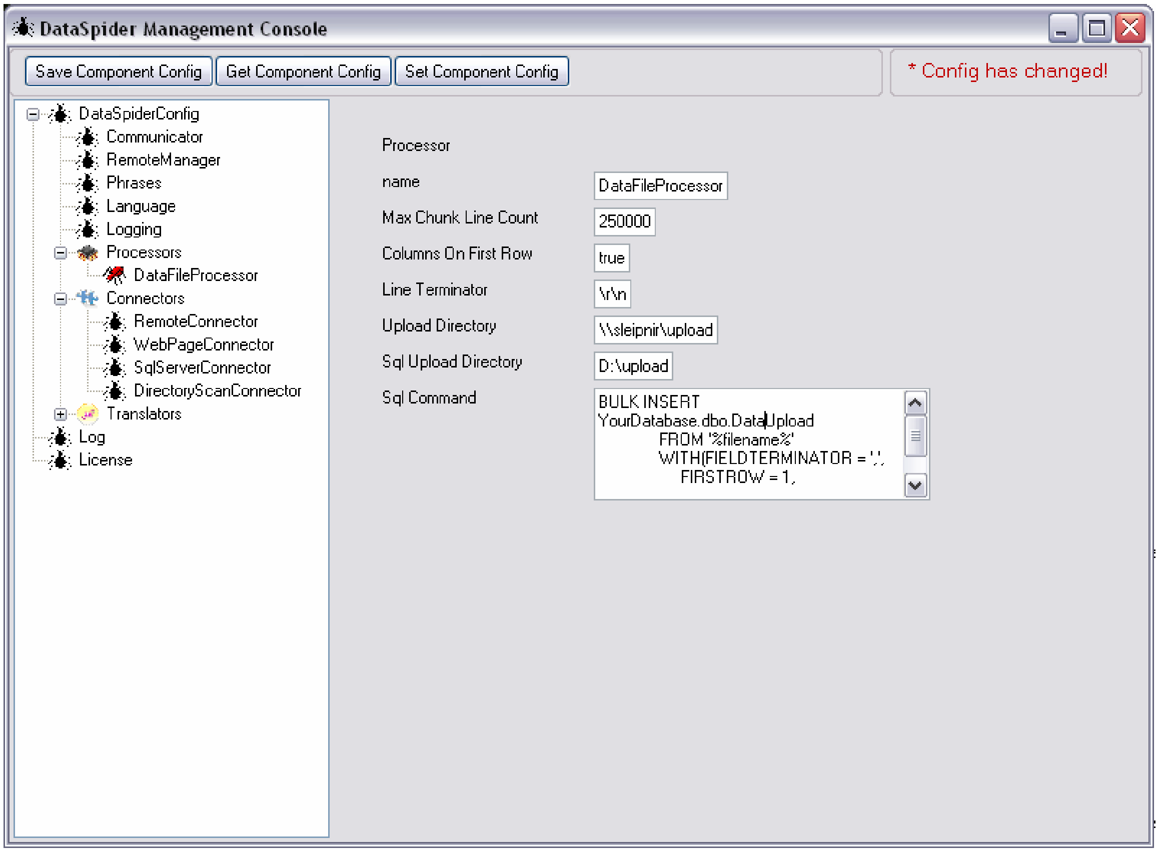
- Config location: The full or relative path to the configuration file. Use '%configstore%' to indicate the path is relative to the main config store.

- Connector: The Guid of the related connector. This should be the same as the Id of the connector.
- Description: Any meaningful word will suffice, but usually the name of the processor.
- Id: The Guid identifying this processor.
- Library: DSConfiguration.DataSpiderProcessorLibrary
- Translator: The Guid of the related translator. This should be the same as the Id of the translator.

From here changes should be committed by hitting 'Save Library Config':

- Full name: The full name of the library. Should not be changed.
- Method: The name of the method to be called by the Communicator.
- Module: The name of the module in the library
- Name: The name of the library
- Path: The full or relative path to the compiled dll.

From here you can drill down to the specific configuration by hitting 'View Component Config':



Processor

name: DataFileProcessor

Max Chunk Line Count: 250000

Columns On First Row: true

Line Terminator: \r\n

Upload Directory: \\sleipnir\upload

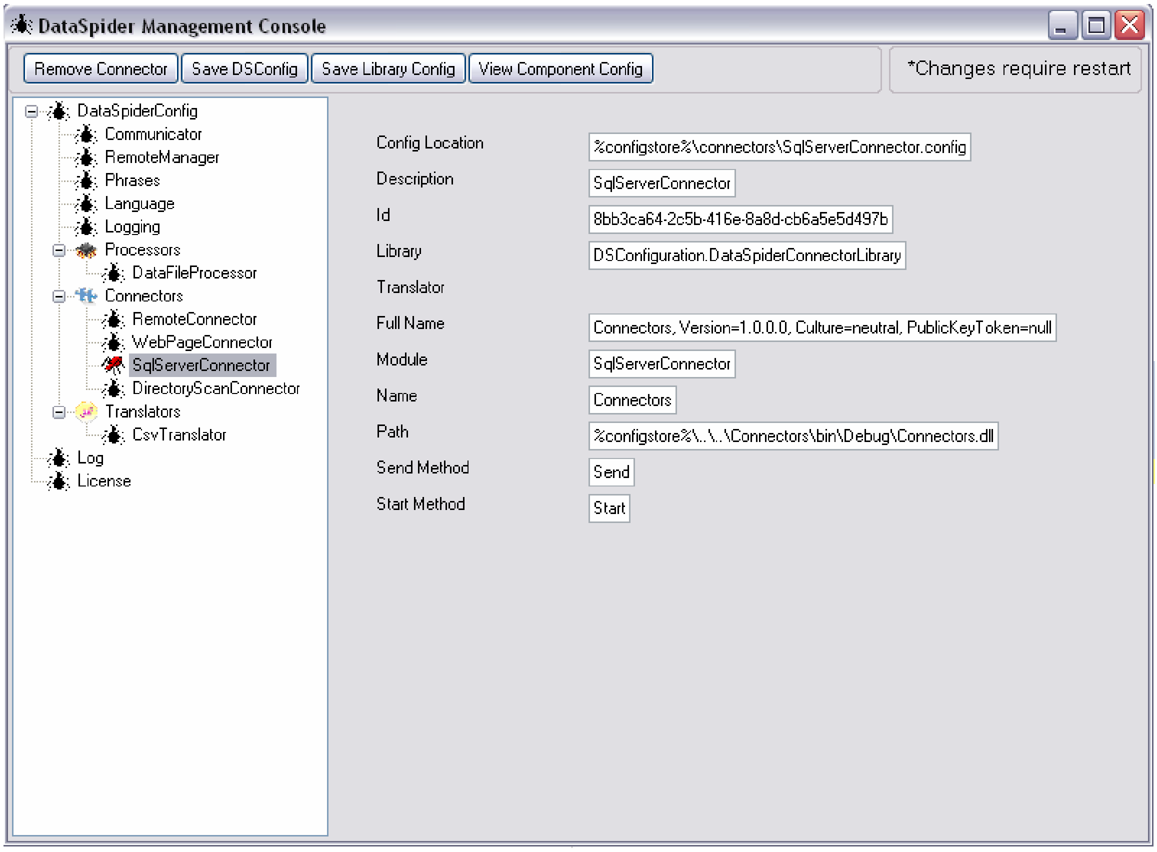
Sql Upload Directory: D:\upload

Sql Command:

```
BULK INSERT
YourDatabase.dbo.DataUpload
         FROM '%filename%'
         WITH(FIELDTERMINATOR = ',',
            FIRSTROW = 1,
```

This screen will not be discussed as these configuration files vary widely depending on the tasks to be performed. You can use the Get and Set Component Config buttons at the top to enter or change the standard configuration file. In the management console you can make minor adaptations afterwards.

Connectors: These provide the connection between the calling programs and the data or service providers. Usually there are two for each service, one for the caller and one for the provider. The screen is very similar to that for the processors:

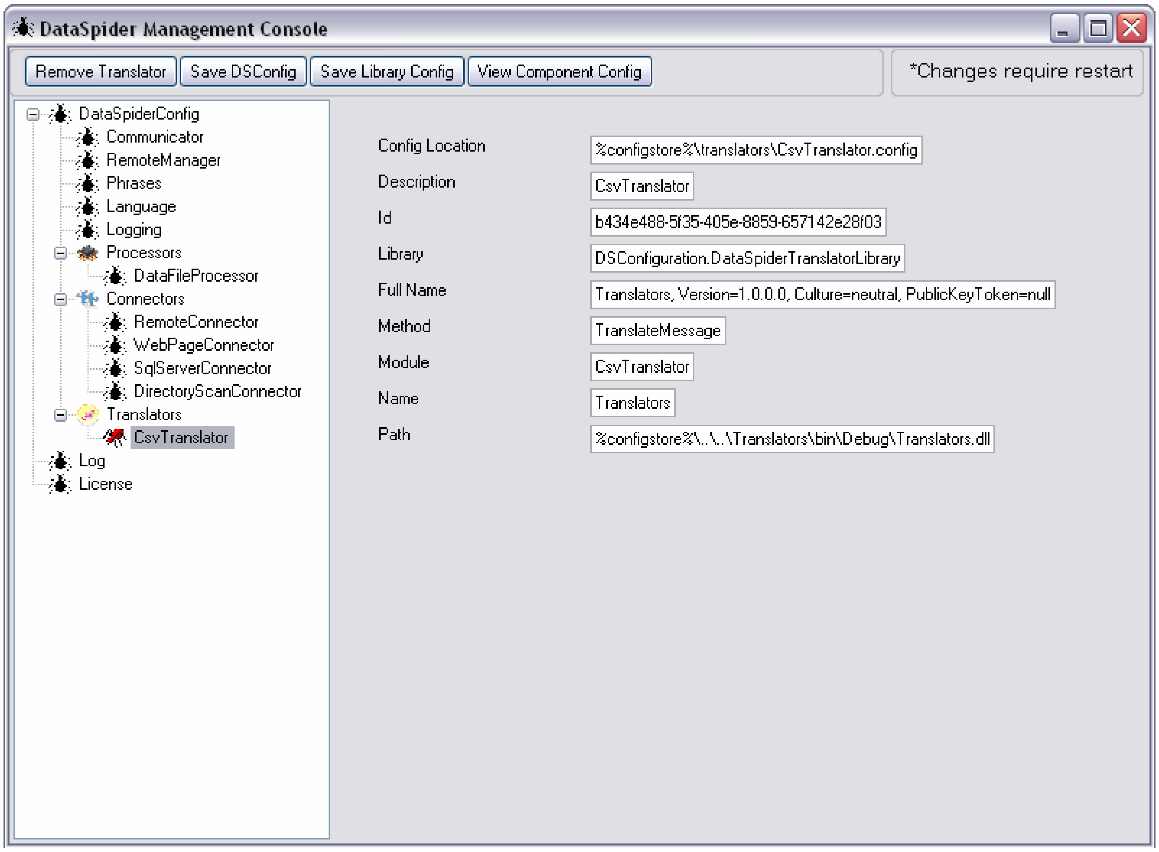


Config Location %configstore%\connectors\SqlServerConnector.config
Description SqlServerConnector
Id 8bb3ca64-2c5b-416e-8a8d-cb6a5e5d497b
Library DSConfiguration.DataSpiderConnectorLibrary
Translator
Full Name Connectors, Version=1.0.0.0, Culture=neutral, PublicKeyToken=null
Module SqlServerConnector
Name Connectors
Path %configstore%\..\..\Connectors\bin\Debug\Connectors.dll
Send Method Send
Start Method Start

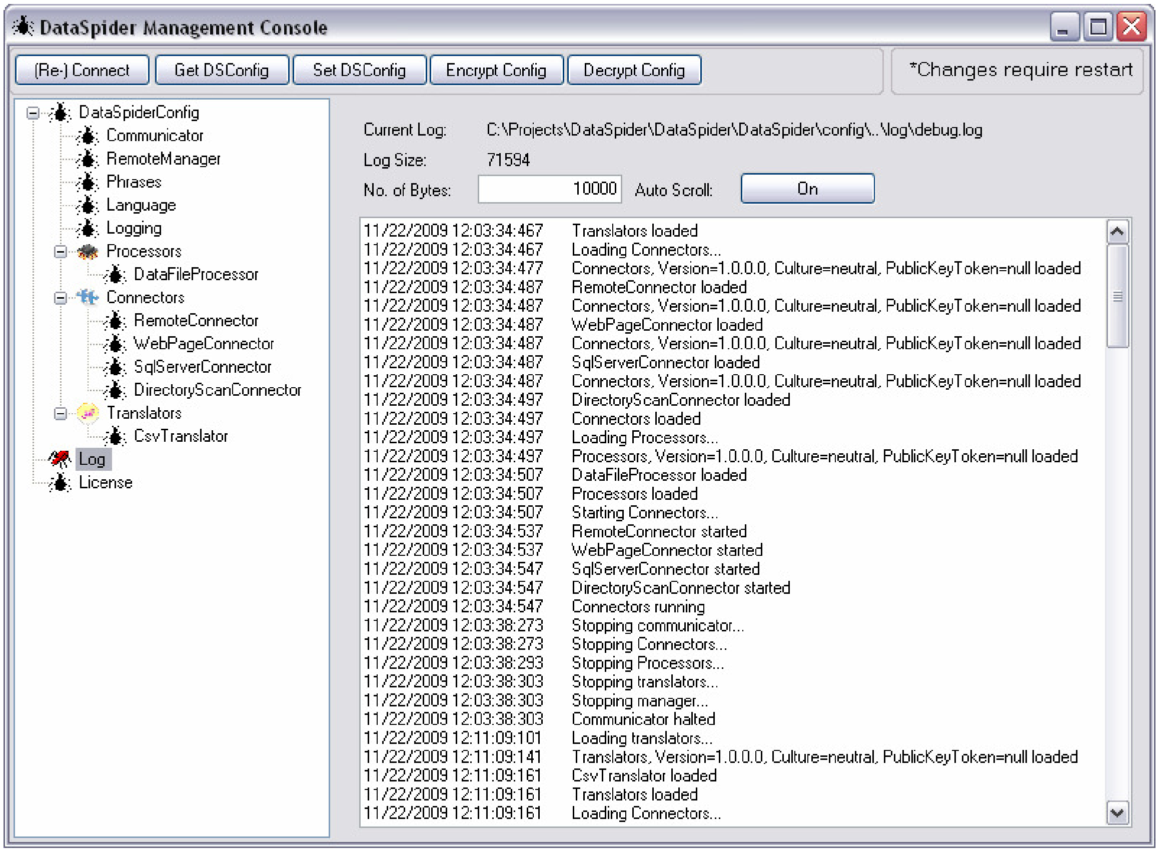
Main difference is that you have to provide two methods one for starting the connector and one for sending out the (return-) messages.
Also from here you can drill down to the Component configuration.

Translators: These provide for the translation of the incoming and outgoing message to suit the specific format and needs of the communicating programs. These translators also vary widely and will not be discussed in detail here. The screen is also very similar to that for the processors:

As the translators always have one method 'Translate', they don't have to be configured separately here.
Also from here you can drill down to the Component configuration.



Config Location	%configstore%\translators\CsvTranslator.config
Description	CsvTranslator
Id	b434e488-5f35-405e-8859-657142e28f03
Library	DSConfiguration.DataSpiderTranslatorLibrary
Full Name	Translators, Version=1.0.0.0, Culture=neutral, PublicKeyToken=null
Method	TranslateMessage
Module	CsvTranslator
Name	Translators
Path	%configstore%\..\..\Translators\bin\Debug\Translators.dll



Current Log: C:\Projects\DataSpider\DataSpider\DataSpider\config\..\log\debug.log

Log Size: 71594

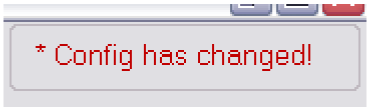
No. of Bytes: 10000 Auto Scroll: On

```
11/22/2009 12:03:34:467    Translators loaded
11/22/2009 12:03:34:467    Loading Connectors...
11/22/2009 12:03:34:477    Connectors, Version=1.0.0.0, Culture=neutral, PublicKeyToken=null loaded
11/22/2009 12:03:34:487    RemoteConnector loaded
11/22/2009 12:03:34:487    Connectors, Version=1.0.0.0, Culture=neutral, PublicKeyToken=null loaded
11/22/2009 12:03:34:487    WebPageConnector loaded
11/22/2009 12:03:34:487    Connectors, Version=1.0.0.0, Culture=neutral, PublicKeyToken=null loaded
11/22/2009 12:03:34:487    SqlServerConnector loaded
11/22/2009 12:03:34:487    Connectors, Version=1.0.0.0, Culture=neutral, PublicKeyToken=null loaded
11/22/2009 12:03:34:497    DirectoryScanConnector loaded
11/22/2009 12:03:34:497    Connectors loaded
11/22/2009 12:03:34:497    Loading Processors...
11/22/2009 12:03:34:497    Processors, Version=1.0.0.0, Culture=neutral, PublicKeyToken=null loaded
11/22/2009 12:03:34:507    DataFileProcessor loaded
11/22/2009 12:03:34:507    Processors loaded
11/22/2009 12:03:34:507    Starting Connectors...
11/22/2009 12:03:34:537    RemoteConnector started
11/22/2009 12:03:34:537    WebPageConnector started
11/22/2009 12:03:34:547    SqlServerConnector started
11/22/2009 12:03:34:547    DirectoryScanConnector started
11/22/2009 12:03:34:547    Connectors running
11/22/2009 12:03:38:273    Stopping communicator...
11/22/2009 12:03:38:273    Stopping Connectors...
11/22/2009 12:03:38:293    Stopping Processors...
11/22/2009 12:03:38:303    Stopping translators...
11/22/2009 12:03:38:303    Stopping manager...
11/22/2009 12:03:38:303    Communicator halted
11/22/2009 12:11:09:101    Loading translators...
11/22/2009 12:11:09:141    Translators, Version=1.0.0.0, Culture=neutral, PublicKeyToken=null loaded
11/22/2009 12:11:09:161    CsvTranslator loaded
11/22/2009 12:11:09:161    Translators loaded
11/22/2009 12:11:09:161    Loading Connectors...
```

Log: A simple built-in logviewer. for more advanced log analysis use a separate program. A special logviewer to handle large flat files is included in the Setup (-> Tools -> Logviewer). For database logging use the manager for the specific database product. Logs are not encrypted. No sensitive data is ever logged except when running Debug-level: This stores the processed messages in a separate folder: All sensitive data inside the messages will be exposed.

License: Let's you view the license and its validity.

After making changes the text in the upper right corner changes and turns red:

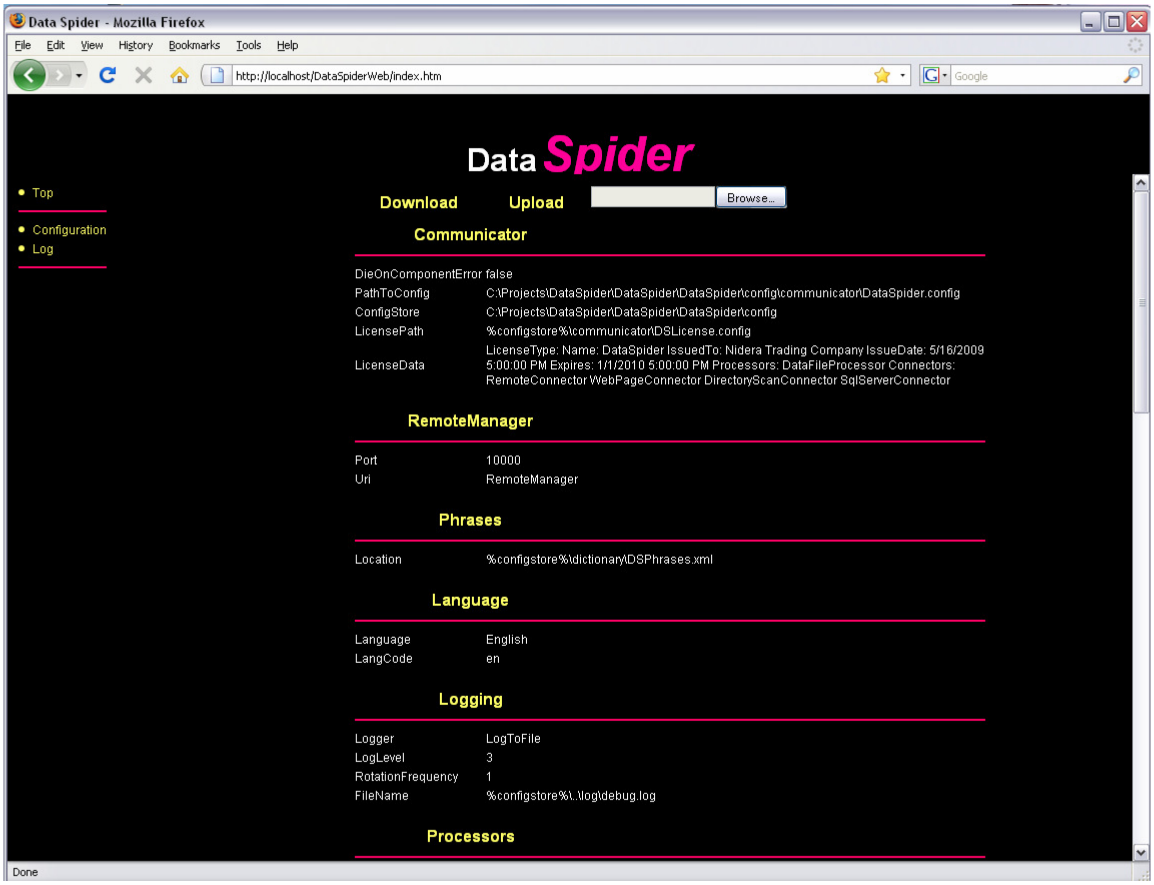


* Config has changed!

If you want to save your changes do so before leaving the specific screen, otherwise they are discarded. Use the appropriate 'Save' button to do so.

Web-interface

Data*Spider* comes with a simple Asp-Net web-interface for viewing the configuration and a basic logviewer:


Data Spider - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://localhost/DataSpiderWeb/index.htm
Google
Data Spider
Top
Configuration
Log
Download
Upload
Browse...
Done

Communicator

DieOnComponentError	false
PathToConfig	C:\Projects\DataSpider\DataSpider\DataSpider\config\communicator\DataSpider.config
ConfigStore	C:\Projects\DataSpider\DataSpider\DataSpider\config
LicensePath	%configstore%\communicator\DSLicense.config
LicenseData	LicenseType: Name: DataSpider IssuedTo: Nidera Trading Company IssueDate: 5/16/2009 5:00:00 PM Expires: 1/1/2010 5:00:00 PM Processors: DataFileProcessor Connectors: RemoteConnector WebPageConnector DirectoryScanConnector SqlServerConnector

RemoteManager

Port	10000
Uri	RemoteManager

Phrases

Location	%configstore%\dictionary\DSPhrases.xml

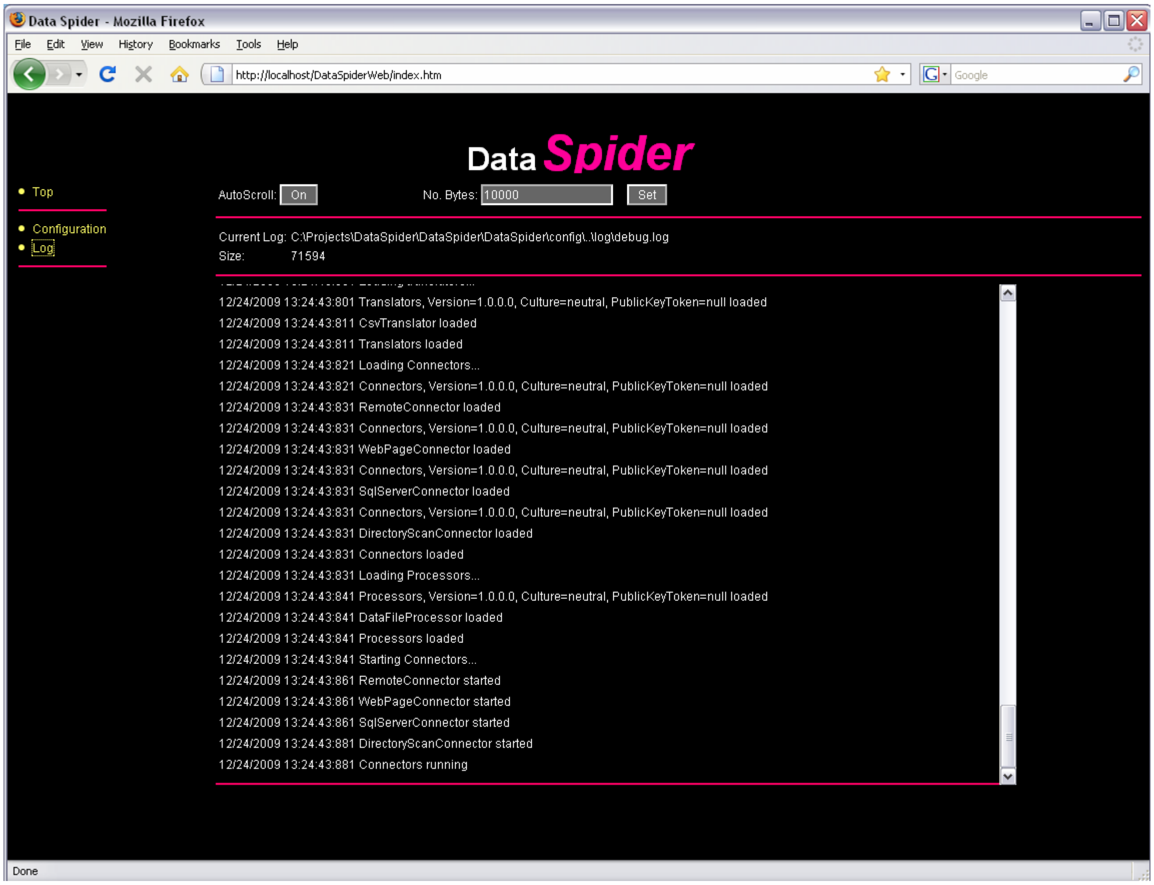
Language

Language	English
LangCode	en

Logging

Logger	LogToFile
LogLevel	3
RotationFrequency	1
FileName	%configstore%\..\log\debug.log

Processors


Data Spider - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://localhost/DataSpiderWeb/index.htm
Google
Data Spider
Top
Configuration
Log
Set
Done

AutoScroll: On

No. Bytes: 10000

Current Log: C:\Projects\DataSpider\DataSpider\DataSpider\config\..\log\debug.log
Size: 71594

```
12/24/2009 13:24:43:801 Translators, Version=1.0.0.0, Culture=neutral, PublicKeyToken=null loaded
12/24/2009 13:24:43:811 CsvTranslator loaded
12/24/2009 13:24:43:811 Translators loaded
12/24/2009 13:24:43:821 Loading Connectors...
12/24/2009 13:24:43:821 Connectors, Version=1.0.0.0, Culture=neutral, PublicKeyToken=null loaded
12/24/2009 13:24:43:831 RemoteConnector loaded
12/24/2009 13:24:43:831 Connectors, Version=1.0.0.0, Culture=neutral, PublicKeyToken=null loaded
12/24/2009 13:24:43:831 WebPageConnector loaded
12/24/2009 13:24:43:831 Connectors, Version=1.0.0.0, Culture=neutral, PublicKeyToken=null loaded
12/24/2009 13:24:43:831 SqlServerConnector loaded
12/24/2009 13:24:43:831 Connectors, Version=1.0.0.0, Culture=neutral, PublicKeyToken=null loaded
12/24/2009 13:24:43:831 DirectoryScanConnector loaded
12/24/2009 13:24:43:831 Connectors loaded
12/24/2009 13:24:43:831 Loading Processors...
12/24/2009 13:24:43:841 Processors, Version=1.0.0.0, Culture=neutral, PublicKeyToken=null loaded
12/24/2009 13:24:43:841 DataFileProcessor loaded
12/24/2009 13:24:43:841 Processors loaded
12/24/2009 13:24:43:841 Starting Connectors...
12/24/2009 13:24:43:861 RemoteConnector started
12/24/2009 13:24:43:861 WebPageConnector started
12/24/2009 13:24:43:861 SqlServerConnector started
12/24/2009 13:24:43:881 DirectoryScanConnector started
12/24/2009 13:24:43:881 Connectors running
```

Copy the files on the CD to an IIS webfolder. Create an application and set the access security. When this is on the same PC as Data*Spider* was installed on the correct AspNet version 2.0 can be selected, otherwise install it separately either from the CD or directely from the Microsoft download site when needed.

Security

The entire config store is encrypted. Never open and subsequently save these files with an editor or they will be corrupted. You can store sensitive data like passwords in these files if you want to. Although a professional algorithm is used, remember that security is never 100% tight and thus no guarantees are given on this. Always limit the access to the folders with Windows security settings. Make sure you have a backup of the configuration stored safely in case of hardware failure or other accidents

For access to network-drives, databases and other objects the account the Data*Spider* service runs in should be granted the appropriate rights. This can also be used for (database) connections configured as 'Integrated Security'.

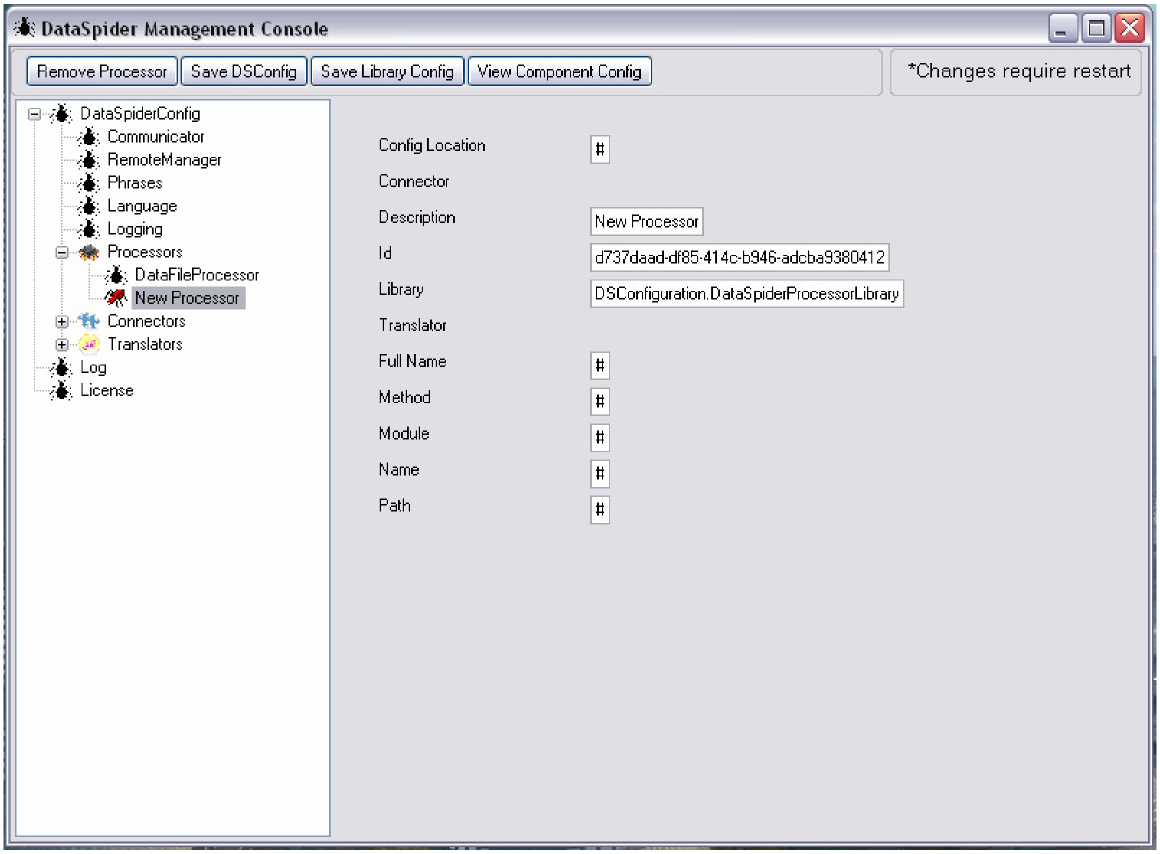
Be sure to adequately protect the PC the Data*Spider* service runs on to prevent others from gaining unwanted access to any of these objects. The more levels of security you implement the better the data is protected.

Setup a Service

To add services you need the appropriate license. Without a license for these services, Data*Spider* won't start.

To setup a service you will need at least one processor, one or two connectors and a translator. Data*Spider* ships with standard components for basic functions. If these are not sufficient you can write your own by inheriting from the base classes or you can have us build it for you.

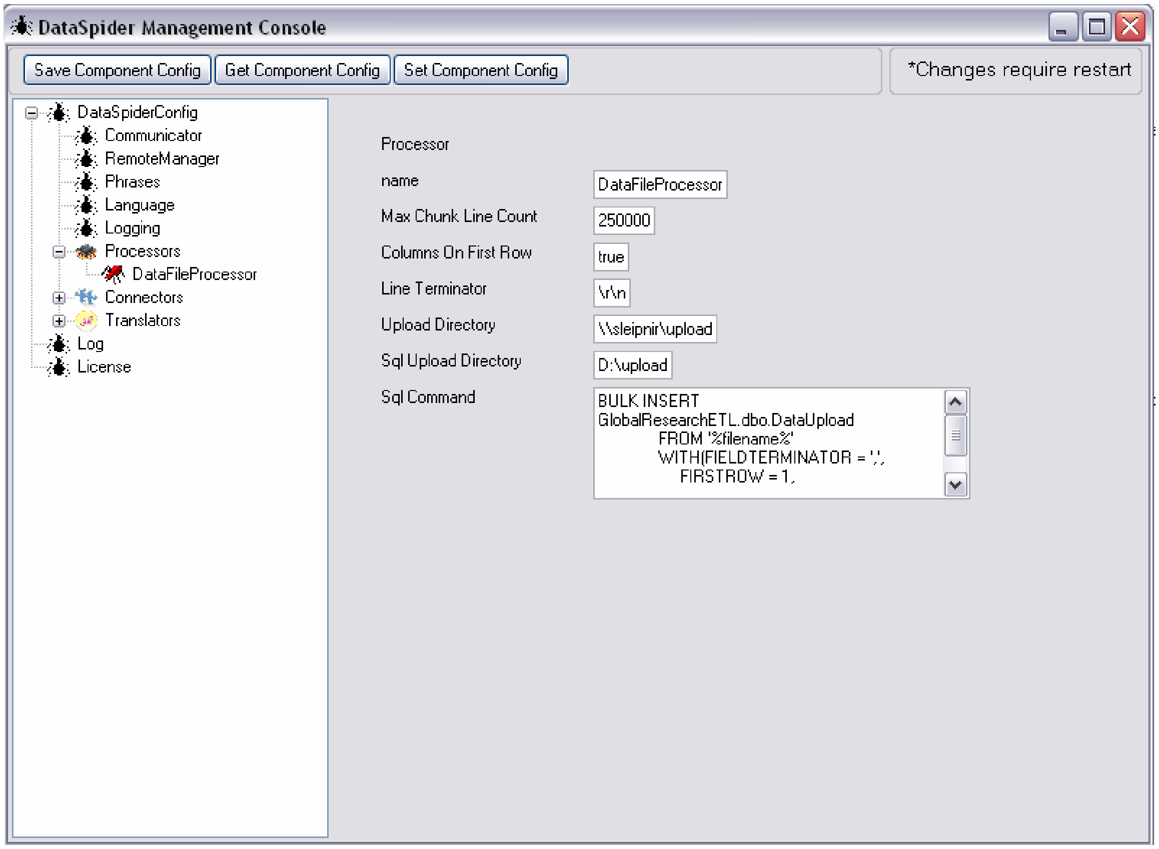
The steps for adding the components are identical for the three types. It's described in detail for Processors. Connect to the Filesystem. In the GUI navigate to the Processors node and hit Add Processor. An empty Processor is added to the tree. Navigate to the new Processor:



Config Location	#
Connector	
Description	New Processor
Id	d737daad-df85-414c-b946-adcba9380412
Library	DSConfiguration.DataSpiderProcessorLibrary
Translator	
Full Name	#
Method	#
Module	#
Name	#
Path	#

As you can see some entries are pre-filled. Adjust the other entries with the values described in the section above. Don't forget to save the library settings separately.
When you're done, navigate back to the new component and hit 'View Component Config'. In the presented screen hit 'Set Component Config' to upload the configuration file associated with the new component. Hit 'Save Component Config' to commit the changes. In the next section the configuration of the standard components that ship with Data*Spider* are described.
Repeat these steps for all required components.

DataFileProcessor



Processor

name: DataFileProcessor

Max Chunk Line Count: 250000

Columns On First Row: true

Line Terminator: \r\n

Upload Directory: \\sleipnir\upload

Sql Upload Directory: D:\upload

Sql Command:

```
BULK INSERT
GlobalResearchETL.dbo.DataUpload
          FROM '%filename%'
          WITH(FIELDTERMINATOR = ',',
             FIRSTROW = 1,
```

Bulk inserts a csv-file into a database table

- Name: Always DataFileProcessor
- Max Chunk Line Count: The number of lines to read per batch from the upload file
- Columns on first rows: True/False
- Line terminator: Either \n (Unix-style) or \r\n (Windows-style)
- Upload directory: Path to put the file, can be an UNC-path

- Sql Upload directory: Path to the upload file as seen from the database server.
- Sql Command: The command in the correct dialect for this database server.

RemoteConnector

Connector for Remoting

- Name: Always RemoteConnector
- Port: Port number
- ChannelName: Unique name of the channel
- Uri: Unique Name of the uri

When configured the Uri to connect to becomes:
tcp://%servername%:%port%/%uri%

WebpageConnector

Retrieves a file from the internet

- Name: Always WebpageConnector
- Url: The Url to retrieve the file from

SqlServerConnector

Connects to a SqlServer

- Name: Always SqlServerConnector
- ConnectString: Server=%server%;database=%database%;User=%user%;Pwd=%pwd% (You can also define Integrated Security when supported)
- Retry Count: The number of retries before throwing an exception
- Retry Interval: Seconds delay between retries
- Non Query: True/False
- Timeout: Timeout in seconds

OdbcConnector

Connects to an Odbc datasource

- Name: Always OdbcConnector
- ConnectString: The connectstring either pointing to an Odbc entry or DSN-less (You can also define Integrated Security when supported)
- Retry Count: The number of retries before throwing an exception
- Retry Interval: Seconds delay between retries
- Non Query: True/False
- Timeout: Timeout in seconds

DirectoryScanConnector

Scans a directory for files and reports the filenames found

- Name: Always DirectoryScanConnector
- Scan Directory: The directory to scan
- Scan Extension: The file type to filter on (May contain wildcards)
- Scan Interval: Delay between scans in milliseconds
- Processor Id: The Guid of the processor to call

CsvTranslator

Transforms one csv file into another

- Name: Always CsvTranslator
- CsvRegex: The Regular Expression to parse the lines from the input file with
- Line terminator: Either \n (Unix-style) or \r\n (Windows-style)
- Column Mappings:
  1. srcindex: The zero-based position of the input column
  2. destindex: The zero-based position of the output column
  3. name: Name of the output column
  4. value: [Optional] The value to fill the output column with. Accepts some standard values: %today% = ShortDateString

ColumnMappings can be as many as required but you can not add or remove them in the management console.

XmlTranslator

Transforms one xml file into another using xslt stylesheets

- Name: Always XmlTranslator
- InputXslt: Path to the xslt file to use for the incoming message
- OutputXslt: Path to the xslt file to use for the reply message

- XsltFill: [Optional] Path to the xslt file to use for the retrieval of the service id from the message
- Params: [Optional] A collection of parameters to feed to the stylesheets

Parameters can be as many as required but you can not add or remove them in the management console.