# DataFlux® Data Management Studio 2.6

## Installation and Configuration Guide

# Contents

# Accessibility

## Accessibility Features of DataFlux Data Management Studio

## Overview

For information about the accessibility of DataFlux Data Management Studio, see the *DataFlux Data Management Studio Users Guide*.

For information about the accessibility of any of the products mentioned in this document, see the documentation for that product.

DataFlux Data Management Studio has not been tested for compliance with U.S. Section 508 standards and W3C web content accessibility guidelines. If you have specific questions about the accessibility of SAS products, contact `accessibility@sas.com` or SAS Technical Support.

## Documentation Format

If you need this document in an alternative digital format, contact `accessibility@sas.com`.

# Recommended Reading

- *DataFlux Data Management Studio Users Guide*

- *DataFlux Web Studio User's Guide*

- *DataFlux Web Studio Installation and Configuration Guide*

- *DataFlux Authentication Server Administrator's Guide*

- *DataFlux Authentication Server User's Guide*

- *DataFlux Data Management Server Administrator's Guide*

- *DataFlux Data Management Server User's Guide*

- *DataFlux Expression Language Reference Guide for Data Management Studio*

- *SAS Federation Server Administrator's Guide*

- *SAS Federation Server User's Guide*

- *DataFlux Migration Guide*

- DataFlux Quality Knowledge Base Online Help and SAS Quality Knowledge Base Online Help

For a complete list of SAS books, go to support.sas.com/bookstore. If you have questions about which titles you need, please contact a SAS Book Sales Representative:

SAS Books
SAS Campus Drive
Cary, NC 27513-2414
Phone: 1-800-727-3228

Fax: 1-919-677-8166
E-mail: sasbook@sas.com
Web address: support.sas.com/bookstore

1

# Installing DataFlux Data Management Studio

## Installing the Software for the First Time

DataFlux Data Management Studio is available through SAS delivery channels. For information about installing this product, see your SAS Software Order E-mail (SOE).

The default installation path under Windows is: `SASHome\<product_instance_name>\<version>`.

The default installation path under UNIX is: `SASHome/<product_instance_name>/<version>`.

For configuration instructions, see "Configuring DataFlux Data Management Studio" on page 7. To install add-on products, see "Installing Add-On Products" on page 51.

# Upgrading Existing Installations of the Software

## Migrating to SAS 9.3

The following table summarizes the upgrade paths when the target system is SAS 9.3.

| Migrating From | Migrating To | Installed with the SAS Deployment Wizard | Migration Instructions |
|---|---|---|---|
| DataFlux Data Management Studio or DataFlux Data Management Server 2.1 or 2.2 | DataFlux Data Management Studio or DataFlux Data Management Server 2.3 or 2.4 | Yes | Migration is automated. Use the SAS Deployment Wizard to install the SAS package that includes 2.3 or 2.4. The wizard uses the deployment registry to locate the 2.1 installation and install 2.3 or 2.4 into that same location. Configuration files are preserved by default. |
| DataFlux Data Management Studio or DataFlux Data Management Server 2.1 or 2.2 | DataFlux Data Management Studio or DataFlux Data Management Server 2.3 or 2.4 | No | Migration is manual. Use the SAS Deployment Wizard to install the SAS package that includes 2.3 or 2.4. Then see the *DataFlux Migration Guide* for instructions about migrating from versions 2.2 to 2.3 or 2.4. |

| Migrating From | Migrating To | Installed with the SAS Deployment Wizard | Migration Instructions |
|---|---|---|---|
| DataFlux Data Management Studio or DataFlux Data Management Server 2.3 | DataFlux Data Management Studio or DataFlux Data Management Server 2.4 | Yes | Migration is automated. Use the SAS Deployment Wizard to install the SAS package that includes 2.4. The wizard uses the deployment registry to locate the 2.3 installation and install 2.4 into that same location. Configuration files are preserved by default. |
| DataFlux Federation Server 2.1 | DataFlux Federation Server 3.1 | No | Migration is manual. See the administration documentation for DataFlux Federation Server 3.1. |
| DataFlux Authentication Server 2.1 | DataFlux Authentication Server 3.1 | No | Migration is manual. See the administration documentation for DataFlux Authentication Server 3.1. |

## Migrating to SAS 9.4

The following table summarizes the upgrade paths when the target system is SAS 9.4.

| Migrating From | Migrating To | Installed with the SAS Deployment Wizard | Migration Instructions |
|---|---|---|---|
| SAS 9.4 offerings that include DataFlux Data Management Studio 2.4 and DataFlux Data Management Server 2.4 | SAS 9.4 offerings that include DataFlux Data Management Studio 2.5 and DataFlux Data Management Server 2.5 | Yes | Migration is automated. Use the SAS 9.4 Deployment Wizard to install the SAS offering that includes 2.5. The wizard uses the deployment registry to locate the 2.4 installation and install 2.5 into that same location. Configuration files are preserved by default. |
| Any SAS 9.3 installation of DataFlux Data Management Studio and DataFlux Data Management Server (all versions) | SAS 9.4 offerings that include DataFlux Data Management Studio 2.5 and DataFlux Data Management Server 2.5 | Yes or No | Migration is manual. Use the SAS 9.4 Deployment Wizard to install the SAS package that includes 2.5. Then see the *DataFlux Migration Guide* for instructions about migrating to 2.5. |
| DataFlux Federation Server 3.1 | SAS Federation Server 3.2 | No | Migration is manual. See the administration documentation for SAS Federation Server 3.2. |
| DataFlux Authentication Server 3.1 | DataFlux Authentication Server 3.2 | No | Migration is manual. See the administration documentation for DataFlux Authentication Server 3.2. |

# System Requirements

You can review system requirements for SAS products at the following location:

`http://support.sas.com/documentation/installcenter/index.html`

Select the release. A search window for that install center documentation appears. Search for your product name. A results page appears with links to the system requirements for your software.

# 2

# Configuring DataFlux Data Management Studio

## About Configuration Files

When the software starts, it determines which configuration options are in effect by reading a series of configuration files, by looking in the environment, and by reading the command line. If two settings of the same name exist in different configuration settings, then the order in which the settings are read determines which value is used. The last value read is used as the configuration setting.

DataFlux Data Management Studio reads configuration settings in this order:

1  The app.cfg file in the etc folder where the software is installed.

2  The app.cfg file in a user folder, such as drive:\Documents and Settings \USERNAME\Application Data\DataFlux\DataManagement\VERSION.

3   The application-specific configuration files in the etc folder, such as ui.cfg or dmserver.cfg.

4   The application-specific configuration files in a user folder.

5   The macros folder in the etc folder. The default path to the macros folder can be overridden with BASE/MACROS_PATH setting in the preceding configuration files.

6   The macros folder in a user folder.

7   The environment variables.

8   The command-line options if applicable.

# DataFlux Folder Permissions for Users

Users must be able to access files that are associated with DataFlux Data Management Studio. At some sites, an administrator might have to manually set the following permissions for individual users.

*Table 2.1   Recommended File System Permissions for DataFlux Data Management Studio Users*

| Folders | Recommended Permissions |
|---|---|
| Files in the DataFlux folder in a user's private storage area, under Microsoft Windows. Example path:<br><br>`C:\Users\[user_ID]\DataFlux` | Full control for all files in this private folder. |
| `DMSTUDIO_HOME\[instance]` | A minimum of Read, Execute, and List Folder Contents for all files and folders in the [instance] folder of the DataFlux Data Management Studio installation directory. |
| `DMSTUDIO_HOME\[instance]\etc` | Full control for all files and folders in the **etc** folder of the DataFlux Data Management Studio installation directory. |

| Folders | Recommended Permissions |
|---|---|
| Any folder specified in a DataFlux repository definition. | Full control for all files and folders in any **Data storage** folder and **File storage** folder specified in a DataFlux repository definition. For more information, see "Understanding Repository Definitions" in the *DataFlux Data Management Studio User Guide*. |

# Configuration Options

Except as noted, the following configuration options can be set for both DataFlux Data Management Studio and DataFlux Data Management Server. These options are typically set in the app.cfg file.

| Option | Purpose | Source | Notes |
|---|---|---|---|
| Base or General Application | | | |
| DEFAULT_UMASK | On UNIX server hosts, specifies the default umask to use when creating output files, such as the output files from jobs. The umask value must be numeric. If not set, the shell's umask value is used. Example: default_umask = 002. | Optional | DataFlux Data Management Servers on UNIX only. |
| ODBC_INI | Overrides the location of the odbc.ini file | Optional | For DataFlux Data Management Servers on UNIX only. |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| BASE/ APP_CONTAINER_D OMAIN | Application container authentication domain | Optional | Identifies the authentication domain expected by application container services. If not specified, "DefaultAuth" is used. |
| BASE/ APP_CONTAINER_L OC | Application container location | Optional | Identifies where to locate the application container services. In most cases this is not required. If it is required, the value is typically an HTTP URI.<br><br>In addition, app.cfg should always set this option to point to the metadata server (iom://<host><port>). |
| BASE/APP_VER | Application version number | Optional | Defaults to 2.6. |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| BASE/ AUTH_DEFAULT_D OMAIN | Default resolved identity domain | Optional | In a metadata configuration, it is possible for the authenticated credentials to resolve to a person that contains multiple logins with the same user ID. |

When more than one log matches the resolved user ID, the authentication domain for the presented credentials is determined by the following:

1. The value of the BASE/AUTH_DEFAULT_DOMAIN option if specified and the specified value matches the authentication domain of one of the logins. If not specified, or no match is found, continue to 2.

2. Use DefaultAuth. If DefaultAuth matches the authentication domain of one of the logins, it is used as the presented credential authentication domain. If no match is found, continue to 3.

3. Use the first matching login.

| Option | Purpose | Source | Notes |
|---|---|---|---|
| BASE/ AUTH_SERVER_LO C | Location of the authenticating server | Optional | Typically, BASE/ AUTH_SERVER_LO C is specified at installation time in the app.cfg file.<br><br>If specified, contains the IOM URI to an authentication server. |

The basic format of the authentication server IOM URI is `iom://` `<host>:<port>`, where <host> is the name of the computer executing the authentication server and <port> is the port to contact the authentication server. If the authenticating server is a DataFlux Authentication Server, then the port should be specified as 21030 unless the default server has been changed. If the authenticating server is a SAS Metadata Server, then the port should be 8561 unless the default server has been changed. For information about valid encodings, see the *SAS National Language Support (NLS): Reference Guide*.

| Option | Purpose | Source | Notes |
|---|---|---|---|
| BASE/ AUTH_SERVER_PA SS | Identifies the password to present when connecting to the authenticating server defined by BASE/ AUTH_SERVER_LO C | | |
| | Typically, BASE/AUTH_SERVER_LOC is specified at installation time in the app.cfg file. Separating the user name (BASE/ AUTH_SERVER_USER) and password (BASE/ AUTH_SERVER_PASS) from the authenticating server location (BASE/AUTH_SERVER_LOC) enables you to run the batch command (dmpexec) with the authenticate option (-a) with individual credentials. For more information see the "Running Jobs from the Command Line" topic in the *DataFlux Data Management Studio: User's Guide*. | | |

| Option | Purpose | Source | Notes |
|--------|---------|--------|-------|
| BASE/ AUTH_SERVER_SS PI | Identifies support for SSPI | Optional | Enables support for single-sign on (SSO) to SAS servers that use the Integrated Object Model (IOM) interface. The SAS Metadata Server uses the IOM interface, for example. The default is NO (no support for single-sign on). Specify YES to enable single-sign-on connections to SAS servers from DataFlux Data Management Studio. Add this option to a user's ui.cfg file or the default ui.cfg file for DataFlux Data Management Studio. There is no reason to set this option in the configuration files for DataFlux Data Management Servers. |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| BASE/ AUTH_SERVER_US ER | Identifies the user name to present when connecting to the authenticating server defined by BASE/ AUTH_SERVER_LO C | Optional | |
| | Typically, BASE/AUTH_SERVER_LOC is specified at installation time in the app.cfg file. Separating the user name (BASE/ AUTH_SERVER_USER) and password (BASE/ AUTH_SERVER_PASS) from the authenticating server location (BASE/AUTH_SERVER_LOC) enables you to run the batch command (dmpexec) with the authenticate option (-a) with individual credentials. For more information see the "Running Jobs from the Command Line" topic in the *DataFlux Data Management Studio: User's Guide*. | | |
| BASE/DMSTUDIO | Studio indicator | Optional | If this option is set to true (value of 1), it indicates that it is using the dmstudio process (not processes started by the application, like dfwfproc, for example). The user should not adjust this or override the value. |
| BASE/ DATE_FORMAT | Specific date formats | Optional | If specified, it is iso8601. |
| BASE/EMAILCMD | Specifies the command used to send e-mail | Required | Can include %T and %B where %T is replaced with the recipient and %B is a file containing the body of the message; also used by monitor event as well as architect nodes. |

| Option | Purpose | Source | Notes |
| --- | --- | --- | --- |
| BASE/EXE_PATH | Path containing executables | Optional | Calculated. |
| BASE/FTPGETCMD | Specifies the command used for Ftp Get Functionality | Required | Should default in the install, as follows:<br>■ %U: Replace with user name.<br>■ %P: Replace with password.<br>■ %S: Replace with server.<br>■ %T: Replace with local directory.<br>■ %F: Replace with Files to download, multiple separated by spaces.<br>■ %L: Replace with the log file to pipe the output. |
| BASE/FTPPUTCMD | Specifies the command used for Ftp Put Functionality | Required | BASE/FTPPUTCMD |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| BASE/ JOB_LOG_ENCODI NG | Encoding for the job log on a DataFlux Data Management Server | Optional<br><br>**Note:** This option must be set on the DataFlux Data Management Server where jobs are executed. It has no effect on DataFlux Data Management Studio job logs. | By default, the log is written in the encoding associated with the locale of the process for the executed job. For English-speaking organizations, this might be LATIN-1 or UTF-8. If a log line contains characters that cannot be represented in the encoding, the log line is not written to the log file. This option enables you to assign the encoding of the job log. |
| BASE/ LIBRARY_PATH | Path for Java JAR dependencies | Optional | Determined by start-up code (DFEXEC_HOME/ lib). |
| BASE/ LOGCONFIG_PATH | Full path to the log configuration file | Optional | Must be set in the configuration file or it defaults to logging.xml in the `etc` directory. |
| BASE/ LOGEXCEPTIONS | Exception logging | Optional | Exception logging defaults to off. |
| | Set this option to either 1 or a combination of letters. A setting of 1 lists the modules loaded when the exception occurred, some information about those modules, and the call stack that caused the error. A setting with letters can include: m = do not show module info, V=turn verbose on, U=install the Unhandled Exception Filter, C=install the continue Exception Filter, f=do not install the first chance exception filter. This must be set before starting the application of interest, as this setting is read only at start-up. | | |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| BASE/ MACROS_PATH | Path for system macros.cfg file | Optional | If not specified, this file is located in the `etc` subfolder of the installation folder. |
| BASE/ MESSAGE_LOCALE | Error message locale | Optional | If not specified, it is determined from the system locale. |
| BASE/ MESSAGE_LEVEL | Error level of messages | Optional | 0 (or not specified) - normal messages; 1 - includes source file and line number in messages. |
| BASE/ MESSAGE_PATH | Path to the message directory | Optional | Determined by start-up code. |
| BASE/ MONITOR_FREQUE NCY | Enables the logging of job node statistics while a job is running. | Disabled by default | If this option is disabled (or its value is -1), then node statistics are logged only when the job has finished executing. |
| | If you want to log node statistics while the job is running, specify the number of milliseconds that the software should wait between logging statistics. The higher the frequency, the more run-time details about node execution are logged in to the job's log file. However, the additional collection and logging of information affects the job's performance. | | |
| BASE/ PLUGIN_PATH | Path used by all subsystems to find plug-ins | Optional | Determined by start-up code. |
| BASE/ PRIMARY_LICENSE | Primary licensing method | Required by base | Must be set in the configuration file as DATAFLUX or SAS |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| BASE/ PRIMARY_LICENSE _LOC | Location of the primary license file or server | Required by base | Must be set in the configuration file. |
| BASE/ REPOS_DDL_LINE_ PREFIX | Format the output of the DDL file that is generated for a repository from the Repository Definition dialog box | Must be set in the configuration file. | For this macro and BASE/ REPOS_DDL_LINE_ SUFFIX only. Specifying ^p as a value causes a line break. |
| BASE/ REPOS_DDL_LINE_ SUFFIX | Format the output of the DDL file that is generated for a repository from the Repository Definition dialog box. | Must be set in the configuration file. | For this macro and BASE/ REPOS_DDL_LINE_ PREFIX only. Specifying ^p as a value causes a line break. |
| BASE/ REPOS_FREQ_CAC HE_MAX | Specifies the number of frequency distributions to cache | Optional. Use only if necessary. | This option prevents memory from overflowing when trying to read a large number of values into memory for frequency distributions. |
| BASE/ REPOS_SYS_PATH | System path for repository configuration files | Optional | Automatically determined. |
| BASE/ REPOS_USER_PAT H | User directory for repository configuration files | Optional | Automatically determined by dfcurver. |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| BASE/ REPOS_FILE_ROOT | Overrides the root of the repository for URI lookups | Optional | If specified, this is used as the root for the repository when resolving the URI. The path in the URI is concatenated to this path to give the actual filename of a URI. |
| BASE/ REPOS_EVENT_WA IT_QUERYMS | Repository event that processes the wait time between processing queries. | Optional | Specifies how frequently in milliseconds to query the repository for changes in the event table. This might need to be changed due to slow servers or IT issues. This is an overriding value and the default is used if no value is set by the user. A setting of -1 disables events from client. |
| BASE/ REPOS_EVENT_CL EAN_TIMEMIN | Repository event processor that removes all events older than X minutes before start-up. | Optional | |
| BASE/ ROWSET_SIZE | Suggested RowSet Size. | Optional | If specified, the value calculates the maximum number of rows each rowset collection should contain. |
| BASE/ SECONDARY_LICE NSE | Secondary licensing method | Required by base | Must be set in the configuration file as DATAFLUX or SAS. |

| Option | Purpose | Source | Notes |
| --- | --- | --- | --- |
| BASE/ SECONDARY_LICE NSE_LOC | Specifies the location of the secondary license file or server. | Required by base | Must be set in the configuration file. |
| BASE/SORTBYTES | Specifies the bytes to use when sorting | Optional | |
| BASE/ SORTMERGES | Enables merge during sort | Optional | |
| BASE/SORTTEMP | Specifies the temporary path for sorts | Optional | |
| BASE/ SORTTHREADS | Specifies the number of sort threads | Optional | |
| BASE/ SORT_KEEPVAR | Specifies the temporary file variable width to fixed width conversion. | A significantly advanced parameter that should rarely be manipulated. | A Boolean that when set to true indicates that the default behavior of the temporary data file is to honor the variable width record indicator at temporary data file creation time. If set to false (default), the temporary data file sort support converts a variable width file to a fixed width file if the record does not contain any string fields or the lengths of the string fields in a record are within a threshold with regard to the overhead necessary to sort variable width records. Set to true to mimic pre-2.4 behavior. |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| BASE/TEMP | Temporary directory | Optional | If not specified, it inherits the value of the TEMP environment variable. |
| BASE/TEXTMINE_LITI_LANG_LOCATION | Doc extraction node option | Optional | This is the install location of Teragram liti files. This allows them to be in Teragram provided languages instead of in the DataFlux install. |
| BASE/TIME_BASE | Whether to use GMT time | Optional | If this is set to GMT (not the default), the current date returns in GMT. This affects anything that uses the current date timestamp. |
| BASE/UPDATE_LEVEL | Application update level | Optional | Defaults to 0. Could be used as a minor revision number. |
| BASE/USER_PATH | Path for user configuration files | Optional | Automatically determined by dfcurver. |
| Data Access Component Logging | | | |
| DAC/DFTKLOGFILE | DFTK logging | Optional | Filename. |
| DAC/DISABLESYSCATENUM | Enumeration of SYSCAT DSNs | Optional | When set to "yes", 1, or "true," this setting disables the listing the SYSCAT type DSNs into DSNs that are on that server. |

| Option | Purpose | Source | Notes |
| --- | --- | --- | --- |
| DAC/ DFTKDISABLECEDA | Disables CEDA support | Optional | "Yes" turns it on. |
| DAC/ DFTK_PROCESS | Run DFTK out of process | Optional | "Yes" turns it on; off by default. |
| DAC/ DFTK_PROCESS_T KPATH | TKTS path for DFTK out of process | Optional | Path that defaults to a `core/sasext` directory off the executable directory. |
| DAC/DSN | DSN directory for TKTS DSNs | Optional | Path that defaults to DFEXEC_HOME/etc/ dftkdsn. |
| DAC/ SAVEDCONNSYSTE M | Location of system-saved connections | Optional | Defaults to DFEXEC_HOME/etc/ dsn. |
| DAC/ SAVEDCONNUSER | Location of user-saved connections | Optional | Defaults to the user settings folder, the folder where all of the application-specific settings supplied by a user are stored, such as the following path under Windows 7: `C: \Users\ [username] \AppData \Roaming \DataFlux\dac \9.x` |
| DAC/TKTSLOGFILE | TKTS logging | Optional | Filename. |
| | Address Update (NCOA) (in dfncoa_appcfg.h) | | |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| NCOA/ AUDIT_OPERID | Specifies the operator ID for an NCOA audit run | Optional | Add this option to either ncoa.cfg or app.cfg. If you do not specify an operator ID, AuditOper is specified automatically. |
| NCOA/DVDPATH | Path to the unpacked and unzipped NCOA data | Required | Resides in macros/ncoa.cfg. |
| NCOA/ QKBPARSEDEFN | Path to the QKB parse definition used for Address Update | Optional | Default is "Name (Address Update)". Resides in `macros/ ncoa.cfg.` |
| NCOA/QKBPATH | Path to the QKB used for Address Update name parsing | Required | Resides in macros/ncoa.cfg. |
| NCOA/USPSPATH | Path to the USPS CASS/DPV/etc data | Required | Resides in macros/ncoa.cfg. |
| NCOA/ REPOSCONNECTIO N | Specifies the connection string used to connect to the Address Update repository | Required | Overrides NCOA/ REPOSDSN. One or the other is required. This is typically set by the Address Update Admin utility. Resides in app.cfg. |
| NCOA/REPOSDSN | Specifies DSN used to connect to the Address Update repository | Required | Is overridden by NCOA/ REPOSCONNECTIO N. One or the other is required. This is typically set by the Address Update Admin utility. Resides in app.cfg. |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| NCOA/ REPOSPREFIX | Table prefix used on the Address Update tables. | Required | This is typically set by the Address Update Admin utility. Resides in app.cfg. |
| NCOA/REPOSTYPE | Specifies the repository type | Required | Valid values are: 0 (Guess), 1 (ODBC), 2 (DFTK). If the value is 0, the node attempts to determine the type from the connection string. This is typically set by the Address Update Admin utility. Resides in app.cfg. |
| NCOA/ DFAV_CACHE_SIZE | Set verify cache percentage for the USPS data. | Optional | The higher the value, the more data is cached. The faster the processing, the more memory is used. The default is 0. Resides in macros/ ncoa.cfg. |
| NCOA/ DFAV_PRELOAD | Set verify preload options for the USPS data. | Optional | Valid values are "ALL" or an empty string. Using "ALL" requires a large amount of memory. Resides in macros/ ncoa.cfg. |
| | | Pooling | |

**Note:** For puddle options, the name of the puddle is placed after 'POOLING/' (for example, POOLING/WFEJOB/MAXIMUM_PROCESSES). If no puddle name is specified, it is globally applied to all puddles. Here are a few puddles: WFEJOB - batch jobs on DMServer, WFESVC - Process services on DMSERVER, APISVC - DFAPI services (in the works)

| Option | Purpose | Source | Notes |
|--------|---------|--------|-------|
| POOLING/ CHILD_MAXIMUM_L AUNCHES | Throttling for launches | Optional | When specified, the number of concurrent child process launches is limited by this value. If the current child launch request exceeds the specified value, the launch waits until the number of launching processes is below the specified value. If zero or not specified, there is no limit of concurrent child launches. |
| POOLING/ GET_PROCESS_TI MEOUT | Acquire process time-out | Optional | Default is no time-out. Specifies the length of time, in seconds, the process requester should wait for a process to become available. If zero, the requester waits definitely. The acquire process time-out is in terms of the acquisition of a process and the process pooling handshaking. It does not consider the time required by the requester to complete application-level initialization. This is a puddle option. |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| POOLING/ IDLE_TIMEOUT | Idle process time-out | Optional | Default is 0. Specifies the length of time, in seconds, a process remains idle before it is terminated. If zero, idle processes are not terminated. This is a puddle option. |
| POOLING/ MAXIMUM_ERRORS | Maximum number of pooled process errors before process is terminated | Optional | Default is 0 (never terminate it). This controls how many times a process can fail (when it is reused for something else) before it is terminated. This is a puddle option. |
| POOLING/ MAXIMUM_PROCES SES | Maximum number of concurrent pooled processes | Optional | If 0, the number of concurrent pooled processes is unlimited. Default is unlimited. If POOLING/ GET_PROCESS_TI MEOUT is set, it waits for that amount of time to get a new process if needed. This is a puddle option. |
| POOLING/ MAXIMUM_USE | Maximum number of pooled process uses before process is terminated. | Optional | Default is 0 (unlimited). The maximum number of times a pooled process can be used. After the pooled process has been used the specified number of times, it is terminated. This is a puddle option. |

| Option | Purpose | Source | Notes |
|--------|---------|--------|-------|
| | | Process Flow | |
| WFE/ CANCEL_TIMEOUT | Amount of time to give remote processes to cancel in milliseconds | Optional | When user selects cancel, this is the amount of time to wait for remote nodes to exit gracefully before terminating them. |
| WFE/ ENGINE_THREAD_L IMIT | Specifies the thread pool limits for the workflow engine. | Optional | Use this setting to limit the number of engine threads. The default is 0, meaning unbounded, which defers to the system for the thread pool limits. The optimal setting is the number of processors + 1. |
| WFE/ STATUS_FREQUEN CY | How frequently to update status | Optional | The default is 250 milliseconds. This is how long to wait before obtaining status from a remote node. Setting to -1 disables polling for status (which might yield better performance). |
| | | Profile | |
| PROF/ DEBUG_MODE | Frequency distribution engine debug mode | Optional | Possible values include 0, not debug mode, or 1 debug mode. The default is not debug mode. The log is located at C: \Documents and Settings\<USER ID> \Local Settings\Temp. |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| PROF/ LOCK_RETRIES | SQLite repository connection attempts | Optional | Specifies the number of times to retry SQLite repository connection when a connect attempt times out or -1 to retry until a connection is established. |
| PROF/ PER_TABLE_BYTES | Frequency distribution engine per table bytes | Optional | Any numeric value. Default is -1 (frequency distribution engine default). |
| | | QKB | |
| CUSTOMIZE/ DISABLE_FILE_NOT IFICATIONS | Temporarily disables notifications | Optional | Read by ui.cfg when Customize starts. When QKB developers make numerous small changes to files in an editor while Customize is open, Customize sends a notification that warns that the file is being changed and provides a list of all the definitions that are affected. To temporarily disable these notifications, edit ui.cfg by adding *CUSTOMIZE/ DISABLE_FILE_NOT IFICATIONS=1*. |

| Option | Purpose | Source | Notes |
|--------|---------|--------|-------|
| QKB/ ALLOW_INCOMPAT | Allow Data Jobs to run even when the software detects that a QKB definition invoked by the job was saved by a version of the software that is later than the current version of the software. | Optional | Default is NO. The default behavior is for these definitions to fail to load. Results obtained when this option is turned on are undefined. |
| QKB/COMPATVER | Tells DataFlux Data Management Studio which version of Blue Fusion to use when running a data job. | Optional | Possible values: dfpower82, dmp21, and dmp22. Default: dmp22. This is for customers who want to use the latest version of DataFlux Data Management Studio but who want the outputs of their QKB-related Data Job nodes (for example, matchcodes) to be exactly the same as the outputs for earlier versions). |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| QKB/ON_DEMAND | Loads QKB definitions on demand | Optional | Default is YES. The application start-up creates a Blue Fusion pool that sets the option for all consumers (Profile, Explorer, and Nodes) with the exception of the Expression Engine, which has its own initialization. |
| | | | Set this option to No to find errors within definitions and to see error messages specific to Pattern Logic nodes. |
| | If used for a data job, this option allows QKBs to load definitions as needed instead of loading all of them at the beginning. These definitions are kept in memory for future runs that reuse that process. | | |
| | If used for a service, each loaded service loads its own QKB. Once a service is loaded into memory and runs at least once, it keeps the QKB loaded from the previous run and does not have to be loaded again. Note that a QKB in memory is not shared across different services or threads, so each initiation of either a new service or a new thread for an existing service will cause the QKB to be loaded. This could have an implication on memory and performance. | | |
| QKB/PATH | Path to QKB | Required by QKB products | Path is set to the default QKB defined in application. |

| Option | Purpose | Source | Notes |
|--------|---------|--------|-------|
| QKB/SURFACEALL | Surfaces all definitions in the Data Job interface, even definitions for which the "Surface" flag is unchecked in Customize. | Optional | Default is NO. Note that the application start-up creates a Blue Fusion pool that sets the option for all consumers (Profile, Explorer, and Nodes) with the exception of the Expression Engine, which continue to have its own initialization. |
| Architect Client (UI) settings | | | |
| ARCHITECT/ AutoPassThru | Client option to set mappings | Optional | Maintained by client; choices are 0 (target), 1 (Source and Target), and 2 (All). |
| Architect nodes, and so on (Defined in ids.h) | | | |
| CLUSTER/BYTES | Specifies the bytes use when clustering | Optional | |
| CLUSTER/LOG | Specifies whether a clustering log is needed | Optional | |
| CLUSTER/TEMP | Specifies the cluster temporary path | Optional | |
| FRED/LOG | Specifies whether a FRED log is needed | Optional | |
| JAVA/CLASSPATH | Specifies the Java classpath | Optional | |
| JAVA/DEBUG | Specifies the Java debug options. | Optional | |

| Option | Purpose | Source | Notes |
|--------|---------|--------|-------|
| JAVA/DEBUGPORT | Specifies the port to remotely debug Java. | Optional | |
| VERIFY/PRELOAD | Preloads defined state data into memory for address verification. | Optional | Set in the app.cfg file. Values can be *ALL*, a two-letter state abbreviation, or multiple state abbreviations separated by spaces. |
| VERIFY/USELACS | Enables or disables the LACSLink processing | Optional | Locatable Address Conversion System (LACS). |
| VERIFY/USEELOT | Enables or disables the eLOT processing | Optional | |
| VERIFY/USPS | Specifies the USPS data path | Required by USPS | Maintained by USPS installation. |
| VERIFY/UPSPINST | Determines whether the USPS data is installed or if sample data is being used | Required | Maintained by USPS installation. |
| VERIFYINTL/CFG | Verifies the international addresses | Required by international verification | Path maintained by component installation. |
| VERIFYWORLD/ CONFIGFILE | Specifies the path to the SetConfig.xml file that is used by the Address Verification (World 2) data job node. Use this option to change the default location of this file. | Required by Address Verification (World 2) data job node | For more information about this file, see the Address Verification (World 2) node in the *DataFlux Data Management Studio Users Guide*. |
| VERIFYWORLD/DB | Specifies the Platon data path | Required for Platon | Path maintained by component installation. |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| VERIFYWORLD/ UNLK | Specifies the Platon library universal unlock code | Required for Platon | Path maintained by component installation. |
| WEBSERVICE/ CONFIG_FILE | Specifies a user-defined configuration file for the Web Service node and the HTTP Request node. This file can be used to increase the time-out value, for example. | Optional | For more information about the user-defined configuration file, see the FAQ topic: "What Can I Do About Time-Out Errors in Data Jobs with the Web Service Node or the HTTP Request Node?" in the *DataFlux Data Management Studio: User's Guide*. |
| dfIntelliServer | | | |
| DFCLIENT/CFG | Used for dfIntelliServer | Required | Maintained by dfIntelliServer installation; typical location is 'C:\Program Files\DataFlux\dfIntelliServer\etc\dfclient.cfg; modify the dfclient.cfg file to point to the server and port. |
| Repository | | | |
| REPOS/ CREATE_SPEC_PATH | Specifies how to create the repository table or index | Optional | This specification provides a means of configuring the commands to create tables and indexes in the repository. |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| REPOS/ FORCE_FILE_BASED | Repository SQLite usage | Optional | If set to true, all SQLite access goes through dfsqlite instead of DAC. |
| REPOS/ LOCK_RETRIES | Specifies the number of attempts to connect to a SQLite repository | Optional | Number of times to retry SQLite repository connection when a connect attempt times out or -1 to retry until a connection is established. |
| REPOS/ TABLE_LIST_PATH | Repository XML table definition | Optional | The directory that should contain XML files for any tables the repository library should add on creation or update. If set, look here for XML files that contain repository table definitions; if not set, look in DFEXEC_HOME/etc/ reposcreate. |
| Other | | | |
| BY_GROUP | Specifies to sort the job by the selected group | Optional | When a key differs, an indicator is returned informing the caller that the retrieved row begins a new group with the same key. This is useful when clustering by a single key or similar processing. |

| Option | Purpose | Source | Notes |
|--------|---------|--------|-------|
| COLLATION | Specifies how things are collated | Optional | A value of UPPER sorts uppercase letters first, then the lowercase letters. A value of LOWER sorts lowercase letters first, then the uppercase letters. If you do not select a collation value, then the user's locale-default collation is selected. Linguistic collation allows for a locale-appropriate collating sequence. |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| DMSERVER/NAME | Specifies the name of the metadata definition of the DataFlux Data Management Server that is stored on the SAS Metadata Server. When the DataFlux Data Management Server is started, it uses the name to query the SAS Metadata Server for configuration information. | This option is ignored when BASE/AUTH_SERVER_LOC identifies a DataFlux Authentication Server rather than a SAS Metadata Server. | For DataFlux Data Management Servers only.<br><br>This option is specified by default when the DataFlux Data Management Server is installed as part of SAS Visual Process Orchestration. |

When the option BASE/AUTH_SERVER_LOC in app.cfg identifies a SAS Metadata Server, the DataFlux Data Management Server retrieves and sets the following values:

- DMSERVER/SOAP/LISTEN_HOST

- DMSERVER/SOAP/LISTEN_PORT

- DMSERVER/SOAP/SSL

- DMSERVER/SECURE

If the SAS Metadata Server cannot locate a metadata definition based on the name, then the DataFlux Data Management Server does not start.

If any of the preceding options have values in the DataFlux Data Management Server's dmservre.cfg file, then the local values override the values that are supplied in metadata. For this reason, it is recommended that you comment-out these options in dmserver.cfg.

To access the named metadata definition on the SAS Metadata Server, one of two conditions must be met. You can ensure that the process owner of the Data Management Server has a user definition on the SAS Metadata Server. Otherwise, the named metadata definition needs to be available to the PUBLIC group.

| Option | Purpose | Source | Notes |
|---|---|---|---|
| EXPRESS_MAX_STRING_LENGTH | Specifies the maximum size of strings declared in expression nodes | Optional | Default maximum length of any string in this node is 5,242,880 bytes (5MB). This enables specifying a larger value in bytes. If performance issues arise, the suggested setting is 65536 bytes. |
| EXPRESSION/ UDFDIR | Specifies where to look for UDF files | Optional | If not specified, look for UDF files in **installationdi r/etc/udf**. |
| JAVA/COMMAND | Command used to launch Java | Optional | Default is Java. This is the command used to launch the Java proxy process. The Java command must be compatible with launching from the command line. |
| | Here are some examples: `JAVA/COMMAND = javaJAVA/ COMMAND = java -Djavax.net.ssl.trustStore=C: \Store\jssecacertsJAVA/COMMAND = java - Djavax.net.ssl.trustStore="C:\Cert Store \jssecacerts"JAVA/COMMAND = "C:\Program Files \Java\jre6\bin\java"` | | |
| MDM/ REPOSITORY_ROOT_FOLDER | Allows the **foundations/ master_data** to be overwritten by the end user when putting the contents of [INSTALL_ROOT]/ share/mdm into a repository. | Optional | Name and location of the root folder for Master Data Management within a repository. |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| MONITOR/ BULK_ROW_SIZE | Specifies the number of rows in a bulk load. | | Default value of a bulk load is 1000. This value can be changed to enhance the performance of jobs that monitor business rules when those jobs include row-logging events. |
| MONITOR/ DUMP_JOB_DIR | Specifies a directory to store temporary jobs created by the business rule monitor. | Optional | By default, this option is not set and the Monitor does not store temporary jobs on disk. |
| NODUPS | Specifies no duplicates. | | When duplicate keys are encountered, only one of the rows containing the duplicate key is returned to the caller. |
| SAP_LIBPATH | Specifies the location of SAP RFC libraries on UNIX only. | Optional | These shared libraries are installed to support the SAP Remote Function Call node, a data job node in DataFlux Data Management Studio. For more information, see "Installing Support for the SAP RFC Node " on page 60 . |

| Option | Purpose | Source | Notes |
|--------|---------|--------|-------|
| STABLE | Ensures that when duplicate keys are encountered, the duplicate rows are returned in the order in which they were entered. | | The stable implementation is accomplished by adding a row counter key to the end of the selected key(s). The row counter is hidden from the caller and is not returned. The addition of a unique portion to the key adversely affects BY Group and No Duplicate processing. Therefore, when the Stable feature is requested with either BY Group or No Duplicate processing, the BY Group and No Duplicate processing are delayed until post sort. |

| Option | Purpose | Source | Notes |
|---|---|---|---|
| STEPENG/ PROFILEBYNODE | Specifies the performance profiler by node instance. | Use only for design and testing. Do not use in a production environment | When set to Yes, this provides each node instance and how many milliseconds were spent on each operation (prepare, pre-execute, execute), and how many times each was entered. The ID corresponds to the iid field in the job's XML file, and includes the job name so that you can see embedded jobs. |

To turn on this functionality, go into your configuration files. To profile real-time services, update dfwsvc.cfg. To profile batch jobs, update dfwfproc.cfg. To profile from studio, update ui.cfg. To profile all three, update app.cfg. The results are written to the log under the DF.RTProfiler heading at trace level. An example of the output is: `NX,inner2.ddf,1,0,5` where the values represent action type or operation (either NX - cumulative time spent processing rows, PR - time preparing, or PX - time pre-executing), job name, instance (iid field in the XML file), milliseconds, and entries (the number of times that you have entered that code).

| Option | Purpose | Source | Notes |
|---|---|---|---|
| STEPENG/ PROFILEBYTYPE | Specifies the performance profiler by node type. | Use only for design and testing. Do not use in a production environment | When set to Yes, this setting provides you each node type along with how many milliseconds were spent on each of three operations (prepare, pre-execute, execute), and how many times each was entered. |
| | To turn on this functionality, go into your configuration files. To profile real-time services, update dfwsvc.cfg. To profile batch jobs, update dfwfproc.cfg. To profile from studio, update ui.cfg. To profile all three, update app.cfg. The results are written to the log under the DF.RTProfiler heading at trace level. An example of the output is: `NX,ARCHITECT_EMBEDDED_JOB,0,5` where the values represent action type or operation (either NX - cumulative time spent processing rows, PR - time preparing, or PX - time pre-executing), node type, milliseconds, and entries (the number of times that you have entered that code). | | |

# Data Access Component Directives

The Data Access Component (DAC) enables you to connect to data using Open Database Connectivity (ODBC) and Threaded Kernel Table Services (TKTS). ODBC database source names (DSNs) are not managed by the DAC, but by the Microsoft ODBC Administrator. However, TKTS DSNs are managed by the DAC, and TKTS connections are stored in a TKTS DSN directory. The DAC reads a list of settings to determine what actions to take, where to find various things that it needs, and how to make the proper connections.

**Note:** The default DAC directives should be satisfactory for most sites. Change these settings only if you have special needs.

There are three types of DAC Configuration Directives:

configuration options

Most of the DAC's settings come from configuration values as specified in app.cfg or other configuration files that are read in a specific sequence. It is recommended to put these settings in the app.cfg file, but they can be put in macro.cfg or another application-specific configuration file. See "Configuration Options" on page 9 for a more comprehensive list of possible configuration values.

registry

The DAC only looks in the registry for DSN-specific items. These are found in either or both HKEY_CURRENT_USER\Software\DataFlux\dac\{version}\{dsn} or HKEY_LOCAL_MACHINE\Software\DataFlux\dac\{version}\{dsn}. In this case, "dsn" is a string with the same name as the DSN to be affected.

environment variables

You can identify the environment variables with $NAME in the following table. The application using the DAC typically sets environment variables to appropriate locations. On Windows, $DFEXEC_HOME is set to the application's root directory.

| Setting | Description | Windows | UNIX |
|---|---|---|---|
| Command file execution | Specifies a text file with SQL commands (one per line). These commands run in turn on any new connection that is made. For example, they can be used to set session settings. This is implemented only for the ODBC driver. | The DAC/SAVEDCONNSYSTEM configuration value specifies the path to the saved connections. The DAC checks for files with the same filename as the DSN and a .sql extension. | The same as Windows. |

| Setting | Description | Windows | UNIX |
|---|---|---|---|
| Connection and login timeout | Enables you to specify the time in seconds for a connection time-out and a login time-out. | The string type value is located in USER \\*dsn_name* \\logintimeout (or connecttimeout) where dsn_name is the name of the dsn. The string value is the number of seconds for the time out. If the string type value is not found in USER\\. then the DAC looks in the SYSTEM \\*dsn_name* \\logintimeout. | Not supported. |
| DAC logging | Determines whether to create a log file for DAC operations. | This is controlled through the effective logging configuration files, which use the SAS logging facility. The logger name "DAC" should be used. | The same as Windows. |
| DFTK log file | Specifies the log file that interacts with the DFTKSRV layer and is useful only for debugging issues specific to dftksrv. This setting is applicable only if you are running DFTK out of process. | Look for config value DAC/DFTKLOGFILE. | The same as Windows. |
| Disable CEDA | Specifies whether to disable CEDA. This setting is applicable only to tkts connections. | Look for DAC/ DFTKDISABLECEDA configuration value, which should specify any non-null value (for example, yes). | The same as Windows. |

| Setting | Description | Windows | UNIX |
|---|---|---|---|
| Multiple active results sets (MARS) | Specifies whether to allow a connection to open multiple result sets at the same time. This is applicable only when using the SQL Server Native Client driver on Windows against a SQL Server 2005 (or later). | Look in USER \\*dsn_name*\\mars for dword value of 1 where *dsn_name* is the name of the dsn. If the DAC does not find it, it looks in the SYSTEM\\*dsn_name*\\mars. | Not applicable. |
| Oracle NUMBER(38) handling | If connected to Oracle (only), NUMBER(38) columns are treated as INTEGER by default. To override that functionality, treat them as REAL. | Look in USER \\oranum38real for dword value of 1. If it is not found in USER, look in SYSTEM \\oranum38real. | Look in $HOME/.dfpower/ dsn.cfg for a line dsn_name = oranum38real where *dsn_name* is the name of the DSN. |
| Read uncommitted | For SQL Server only, specifies that reading data from cursors is allowed to read uncommitted rows. This can alleviate lockups in SQL Server (applies to DMP version 2.3 and later). | Look in USER \\*dsn_name* \\readuncommitted for dword value of 1. Where *dsn_name* is the name of the dsn. After looking in USER, look in SYSTEM. | Look in $HOME/.dfpower/ dsn.cfg for a line dsn_name = readuncommitted where *dsn_name* is the name of the DSN. |
| Run DFTK out of process | Specifies whether to run TKTS out of process, allowing you to perform troubleshooting. | Look for config value DAC/ DFTK_PROCESS. This should specify any non-null value (for example, yes). | The same as Windows. |

| Setting | Description | Windows | UNIX |
|---------|-------------|---------|------|
| Suffix for CREATE TABLE statements | This enables you to provide a string that is appended to every CREATE TABLE statement. If you include %T in this string, it is substituted with the table name. | Look in USER \\*dsn_name* \postcreate for a string where *dsn_name* is the name of the dsn. If it is not found in USER, look in SYSTEM. | Not supported. |
| System saved connection | Specifies where to find system-saved connections. | Look in DAC/ SAVEDCONNSYSTE M config value for a path. If it is not found, the DAC checks for the DFEXEC_HOME environment variable in the $DFEXEC_HOME \etc\dsn subdirectory, which is in the installation directory. | Look for the same config value. If not found, look in the $DFEXEC_HOME/et c/dsn subdirectory, which is in the installation directory. |
| TK Path | Specifies where threaded kernel files are located. This setting is applicable only if you are running Data Factory Took Kit (DFTK) out of process. The dftksrv path and core directory should be specified. | Look for config value DAC/ DFTK_PROCESS_P ATH. If still not found, set it to $DFEXEC_HOME \bin; $DFEXEC_HOME \bin\core\sasext. | Same as for Windows, but sets it to $DFEXEC_HOME/lib/ tkts if it is not found. |
| TKTS DSN directory | Specifies the path where TKTS DSNs are stored in XML files. | Look in config value DAC/DSN. If it is not found, look in $DFEXEC_HOME \etc\dftkdsn\. | Look in the same config value. If it is not found, look in $DFEXEC_HOME/et c/dftkdsn/. |

| Setting | Description | Windows | UNIX |
|---|---|---|---|
| TKTS log file | Specifies the log file that is produced by the TKTS layer and is useful for debugging tkts issues. | Look for config value DAC/TKTSLOGFILE. | The same as Windows. |
| TKTS startup sleep | Specifies how much time in seconds to delay between the start of the dfktsrv program and the booting of threaded kernel. This setting is applicable only if you are running DFTK out of process. | The DAC checks the registry for a tktssleep value. | Not supported. |
| User saved connection | Specifies where to find user-saved connections. | Look in DAC/SAVEDCONNUSER config value for a path. If not found, look in the user's application settings directory, which is usually found in the %APPDATA% directory, in the DataFlux\dac\*version* subdirectory.<br><br>The %APPDATA% location is C:\Users\userid\AppData\ for Windows 7 and C:\Documents and Settings\userid\Application Data\ for Windows XP. | Look for the same config value. If not found, look in the $HOME/.dfpower/dsn directory. |

| Setting | Description | Windows | UNIX |
|---------|-------------|---------|------|
| Use braces | Specifies whether to enclose DSN items in braces when they contain reserved characters. Most drivers do not support use of braces. However, this might be an option if problems occur with saving credentials. | The DAC looks in USER\usebraces for dword value of 1. If it is not found there, the DAC looks in SYSTEM\usebraces. | Look in $HOME/.dfpower/ dsn.cfg for a line dsn_name = usebraces where dsn_name is the name of the DSN. |

## Logging Options

The following log files are provided for DataFlux Data Management Studio:

- Studio log

- Platform log

- DAC log

- TKTS log

If enabled, the TKTS log captures TKTS events such as BASE data sets and data sent across the wire to SAS Federation Server. The TKTS log provides a default set of information that cannot be customized.

The Studio log, the Platform log, and the DAC log are enabled by default. These logs use the SAS Logging Facility, a flexible and configurable framework that you can use to collect, categorize, and filter events. Then you can write them to a variety of output devices. The logging facility supports problem diagnosis and resolution, performance and capacity management, and auditing and regulatory compliance.

The logging facility framework categorizes and filters log messages in SAS server and SAS programming environments. It also writes log messages to various output devices.

In the server environment, the logging facility logs messages based on predefined message categories such as the following:

- Admin for administrative messages

- App for application messages

- Perf for performance messages

Messages for a category can be written to files, consoles, and other system destinations simultaneously. The logging facility also enables messages to be filtered based on the following thresholds, in order from most inclusive to least inclusive: TRACE, DEBUG, INFO, WARN, ERROR, and FATAL. The configuration files for the Studio log, the Platform log, and the DAC log contain a template for logging options. You need only to uncomment certain lines to turn on the corresponding options. For a full description of these options, see the *SAS Logging: Configuration and Programming Reference*. There are also some DataFlux options for logs. See the logging options in "Configuration Options" on page 9.

The following table specifies the scope, level thresholds, and configuration location for each of these logs:

| Name | Scope | Level Thresholds | Configuration Location |
|------|-------|------------------|------------------------|
| Studio | DataFlux Data Management Studio events | TRACE, DEBUG, INFO, WARN, ERROR, and FATAL | By default, these are enabled. To disable, see note below. |
| Platform | Engine events | TRACE, DEBUG, INFO, WARN, ERROR, and FATAL | By default, these are enabled. To disable, see note below. |
| DAC | Data access events | TRACE, DEBUG, INFO, WARN, ERROR, and FATAL | By default, these are enabled. To disable, see note below. |
| TKTS | TKTS events such as BASE data sets and data sent across the wire to SAS Federation Server | Not configurable | `drive:\Program Files\DataFlux \DMStudio \studio1\etc \app.cfg` |

| Name | Scope | Level Thresholds | Configuration Location |
|------|-------|------------------|------------------------|

**Note:** By default, Studio, Platform, and DAC are enabled. To turn them off, comment out the BASE/LOGCONFIGPATH line in ui.cfg and in batch.cfg and restart the application. For configuration, go to `drive:\Program Files\DataFlux\DMStudio\studio1\etc\platform.log.xml` and `drive:\Program Files\DataFlux\DMStudio\studio1\etc\dfwfproc.log.xml`.

By default, all Windows XP log files are written to `drive:\Documents and Settings\USERNAME\Application Data\DataFlux\DMStudio\[version]\Logs`. The default location in Windows 7 is `drive:\Users\[username]\AppData\Roaming\DataFlux\DMStudio\[version]\Logs`.

All Studio, Platform, and DAC logs by default are all logged to the platform log file. This log file is prefixed with platform_. The DAC logging can be redirected to its own log file by uncommenting the appender-ref setting in the logger section for DF.DAC and by uncommenting the RollingFileAppender section for DacFile. The latter contains the filename pattern that controls where and what name the DAC log file takes. The Studio logging can be redirected to its own log file by uncommenting the appender-ref setting in the logger section for DF.DMStudio and by uncommenting the RollingFileAppender section for DMStudioFile. The latter contains the filename pattern that controls where and what name the Studio log file takes.

The Studio, Platform, and DAC logging all use a logging level of INFO by default. This level is set in the root logger section of the platform.log.xml configuration file. The level value can be changed here to change the default for all three logs. In addition, the logging level for the DAC and Studio logs can be set separately in the DF.DAC and DF.DMStudio logger sections respectively by uncommenting (and possibly updating) the level values in each.

The platform.log.xml file controls all of Studio, Platform, and DAC logging within the DataFlux Data Management Studio application except for job runs. When you run a job, the application creates a separate dfwfproc process for that job run. Therefore, you should modify the dfwfproc.log.xml to configure logging for jobs.

By default, the Platform log contains run-time statistics for job nodes. See the BASE/MONITOR_FREQUENCY option in for details about

these statistics. To disable all job node statistics from the Platform log, add the following block immediately before the <root> section at the bottom of *.log.xml file:

```
<logger name="DF.Monitor">

<level value="OFF"/>

</logger>
```

# 3

# Installing Add-On Products

## Installing QKBs and Data Packs

A Quality Knowledge Base (QKB) is a collection of files that define rules, patterns, and transformations that are used for data cleansing. Data Packs are collections of third-party enrichment data, such as Loqate international address data. QKBs and DataPacks are not installed as part of the SAS Deployment Wizard process. If you have licensed a QKB or a Data Pack, go to SAS support site to download and install these

products. Use the Release Notes for your QKBs to verify that they are compatible with your version of SAS software.

`Download QKBs`

`Download Datapacks`

You can also download these and other products from `SAS DataFlux Software` section of the tech support page.

After you install a QKB, see the "Using Quality Knowledge Bases" chapter in the *DataFlux Data Management Studio: User's Guide* to enable access that QKB. After you install a data pack, see the topic for that feature in the *DataFlux Data Management Studio: User's Guide* to enable access to that data pack.

## Installing Supplemental Language Support

If you plan to use DataFlux Data Management Studio for data that includes East Asian languages or right-to-left languages, you must install additional language support. To install these packages:

1   Select **Start ▸ Settings ▸ Control Panel**.

2   Double-click **Regional and Language Options**.

3   In the Regional and Language Options dialog box, click the **Languages** tab.

4   Under Supplemental Language Support, select the check boxes for **Install Files for complex script and right-to-left languages (including Thai)** and **Install files for East Asian languages**.

5   The Microsoft Windows installer guides you through the installation of these language packages.

# Setup for the Address Update Add-On

## Overview

The DataFlux® Address Update add-on enables you to use the United States Postal Service (USPS) NCOALink® system to identify and update customer address information. This section describes how to perform some of the setup tasks for the Address Update Add-on to DataFlux Data Management Studio. The setup tasks described here are performed outside of the application. Also see the Address Update Add-On chapter in the *DataFlux Data Management Studio User's Guide*, which describes setup tasks required in the interface.

## Obtain a License for DataFlux Products and USPS Components

If you are receiving this DataFlux product as part of a SAS® license, the DataFlux license file is provided with your SAS installation media. If you need assistance in obtaining a license, contact your SAS account representative or access the DataFlux Unlock Code page at `support.sas.com/contact/additional/dfunlockreq.html`.

Contact the USPS to obtain the required licensing and certification.

If you are using the DataFlux Data Management Server to execute address update jobs, you must also have the appropriate DataFlux licenses and USPS licenses on that server.

## Install DataFlux Software

Install the following DataFlux software if this has not been done:

- DataFlux Data Management Studio

- DataFlux Quality Knowledge Base (QKB) for Contact Information 2011A

■ Address Update Add-on installer for DataFlux Data Management Studio

If you are using the DataFlux Data Management Server to execute address update jobs, you must also install the following when you are ready to use the server.

■ DataFlux Data Management Server: See the DataFlux Data Management Server Administrator's Guide.

■ Address Update Add-on installer for Data Management Server, which updates Data Management Server to support the Address Update Add-on.

Address Update jobs must be able to access the DataFlux USPS Verify data from the computer where the jobs are executed. The Verify data is available for Microsoft Windows and UNIX systems. The path to this data is specified in the ncoa.cfg file, which is described later. Be sure to record the installation path to this data. For example, the default installation path for Windows is `C:\Program Files\DataFlux\VerifyData\`

## Install Address Update Data from USPS

Address Update jobs must be able to access the USPS NCOALink® data from the computer where the jobs are executed. If you are using the Data Management Server to execute the address update jobs, you must obtain and license this data from the USPS and install the data so that it is accessible to both the DataFlux Data Management Studio and the Data Management Server. You can copy this information to both computers, or make the information available on a shared file system, as long as you follow the terms of the license. The path to this data is specified in the ncoa.cfg file, which is described later.

Address Update processing does not take place unless you have specific data folders in the expected locations with the correct data files. To ensure that this folder structure is created:

1 Create a folder called "NCOA Data" on your drive.

2 Add subfolders called "DVD 1" and "DVD 2" under the NCOA Data folder.

3 Add subfolders called "CLK" and "RV9" under the DVD 2 folder.

This structure yields the following layout:

*C:\NCOA Data\DVD 1\*
*C:\NCOA Data\DVD 2\ CLK\*
*C:\NCOA Data\DVD 2\ RV9\*

To install Address Update Data:

1   Copy all of the files on DVD 1 into the DVD 1\ directory.

2   Unzip all of the ZIP files on DVD 1 into the DVD 1\ directory.

3   Copy all of the files on DVD 2 into the DVD 2\ directory.

4   Unzip the file CLK.ZIP into the CLK\ directory under the DVD 2\ directory.

5   Unzip the file RV9.ZIP into the RV9\ directory under the DVD 2\ directory.

If you are processing flat files, a sample layout looks like this:

| Subfolder | Files | Source |
| --- | --- | --- |
| C:\NCOA Data\ DVD 1\ | clk.flt.1<br>clk.flt.2<br>and so on | |
| C:\NCOA Data\DVD 2\ | clk.a.hf0<br>clk.b.hf0<br>and so on | |
| C:\NCOA Data\DVD 2\ CLK\ | business.lst<br>clk.flt.3<br>and so on | Folder unzipped from CLK.ZIP |
| C:\NCOA Data\DVD 2\ RV9\ | cln.dat<br>cln.txt<br>crd.dat<br>and so on | Folder unzipped from RV9.ZIP |

If you are processing Hash files, a sample layout looks like this:

| Subfolder | Files | Source |
|---|---|---|
| C:\NCOA Data\ DVD 1\ | clk.a.hs1 | Unpacked from CLK.A.HS1.SZP |
| | clk.a.hs2 | Use the dfAddrHu utility |
| | clk.a.hs3 | |
| | ... | |
| | clk.z.hs8 | hs8 files are unzipped from CLK.*.HS8.ZIP files |
| | and so on | |
| C:\NCOA Data\ DVD 2\ | clk.a.hs0 | |
| | clk.b.hs0 | |
| | and so on | |
| C:\NCOA Data\ DVD 2\CLK\ | | Folder unzipped from CLK.ZIP |
| | business.lst | |
| | clk.flt.3 | |
| | and so on | |
| C:\NCOA Data\ DVD 2\RV9\ | | Folder unzipped from RV9.ZIP |
| | cln.dat | |
| | cln.txt | |
| | crd.dat | |

| Subfolder | Files | Source |
|---|---|---|
| | and so on | |

**Note:** You must unpack Hash files (.szp). DataFlux has provided a utility called dfAddrHu .

You can access the dfAddrHu utility in the `C:\Program Files > DataFlux > DM Studio > [version number] bin` directory, with the following syntax: *dfAddrHu [<options>] —i [<input>] —o [<output>].*

To run the utility, specify the path location to the DVD 1 folder as both the input and the output, as follows: `dfAddrHu <path to DVD 1 folder> <path to DVD1 folder>`.

For example, you can specify the paths as: `dfAddrHu "C:\NCOA Data\DVD 1" "C:\NCOA Data\DVD 1"`.

The utility reads through the DVD 1 folder and unpacks all of the .szp files into the same folder.

## Update the ncoa.cfg File

The physical paths for Address Update data, the DataFlux QKB, and USPS data for Verify (CASS, DPV, and LACS) must be specified in the ncoa.cfg file on any computer that executes address update jobs. You should configure the ncoa.cfg file for DataFlux Data Management Studio and confirm that it is working correctly. Then you can copy the ncoa.cfg file to Data Management Server and edit paths if necessary.

For DataFlux Data Management Studio, add the following options to the ncoa.cfg file, located at `C:\Program Files\DataFlux\DMStudio\<version number>\etc \macros`:

NCOA/DVDPATH
    Location where the USPS NCOALINK data is installed

    Example: ncoa/dvdpath = "C:\NCOA Data\"

NCOA/QKBPATH

> The location of the QKB (minimum of QKB for Contact Information 2011A is required)

> Example: ncoa/qkbpath = "C:\Program Files\DataFlux\QKB\CI\2011A\"

NCOA/USPSPATH

> Location of the DataFlux USPS Verify data (CASS, DPV, LACS)

> Example: ncoa/uspspath = "C:\Program Files\DataFlux\VerifyData\data\1105\usps\"

The following optional settings are available for the United States Postal Service (USPS), CASS, and LACS configuration. Set this option for the Address Update node in the ncoa.cfg file. Note that these options do not affect the operation of the Address Verify nodes.

NCOA/DFAV_CACHE_SIZE

> Range: 0 through 100 and indicates how much USPS data to cache. The higher the value, the more data is cached. The faster the processing, the more memory used. The default is 0.

NCOA/DFAV_PRELOAD

> Set dfav/verify preload options for the USPS data. Provide the names of states to preload, to accelerate address verification for those states. Valid values include:

> > "." - do not preload any states (default)
> >
> > "ALL" - preload all states
> >
> > "MIL" - preload military addresses only
> >
> > "CONUSA" - preload the 48 contiguous states
> >
> > "TX FL" - preload Texas and Florida

**Note:** Monitor the impact of the NCOA/DFAV_CACHE_SIZE and NCOA/DFAV_PRELOAD options on the available memory on your system. If you run out of memory, you might consider decreasing the amount of memory that you have allocated to these options.

After you have run a sample address update job and verified that the ncoa.cfg file is working properly, copy the ncoa.cfg file to DataFlux Data Management Server and edit

paths if necessary (for example, ).`C:\Program Files\DataFlux\DMServer` `\<version number>\etc\macros`.

## Update the app.cfg File

DataFlux Data Management Studio has a PAF administration dialog box. For instructions about using this dialog box, see the Online Help under "Administering the Address Update Add-On". The PAF administration dialog box updates the app.cfg file.

Descriptions of the options added by the PAF administration dialog box are as follows:

NCOA/REPOSDSN
   DSN connection for the address update repository

NCOA/REPOSPREFIX
   Table prefix for the tables in this repository, if a prefix has been specified

NCOA/REPOSTYPE
   Value that indicates the type of repository:

   > 0 = NULL (DataFlux Data Access Component (DAC) attempts to determine the repository type from the connect string)
   > 1 = ODBC DSN
   > 2 = Custom DSN

The type depends on how the address update repository was configured in the PAF administration dialog box.

If you are using the DataFlux Data Management Server to execute address update jobs and have run a sample address update job and verified that the app.cfg file is working properly, then copy the app.cfg file to DataFlux Data Management Server and edit paths if necessary (for example, `C:\Program Files\DataFlux\DMServer\<version number>\etc`).

## Setup Continues in DataFlux Data Management Studio

See the DataFlux Data Management Studio Online Help Address Update Add-On section for additional information about setup tasks that must be performed.

## USPS Copyright Disclosure

DataFlux is a non-exclusive interface distributor of the United States Postal Service and holds a non-exclusive license from the United States Postal Service to publish and sell USPS CASS, DPV, and RDI information. This information is confidential and proprietary to the United States Postal Service. The price of these products is neither established, controlled, or approved by the United States Postal Service.

# Installing Support for the SAP RFC Node

If you open a data job in DataFlux Data Management Studio, the **SAP Remote Function Call** node is available in the **Data Integration** folder. This node enables you to read and write data using SAP RFC and Business Application Programming Interfaces (BAPI) provided by SAP. This functionality is useful for the following:

- reading information from SAP that is not readily available in tables

- using the functionality that SAP exposes as BAPIs

- loading data to SAP by using functions provided by SAP that are a standard way of updating objects

- triggering an upload to SAP by using functions that take (for example, an external file input)

In order to use the SAP Remote Function Call node, you must install SAP libraries (DLLs) on the computer where the job that contains the node is executed. For example, if you use DataFlux Data Management Studio to execute a job with an **SAP Remote Function Call** node, then you must install the SAP libraries on the computer where the

software is running. If you execute a job with an SAP Remote Function Call node, then you must install the SAP libraries on the computer where Data Management Server is running.

DataFlux Data Management Studio is a 32-bit Windows application and requires 32-bit SAP libraries. See the SAP documentation for details about installing these libraries.

DataFlux Data Management Servers can be either 32-bit or 64-bit. See the SAP documentation for details about installing the appropriate libraries. Note that for DataFlux Data Management Servers running on UNIX systems, you can set the SAP_LIBPATH option in the app.cfg file to point to the SAP DLLs.

# Appendix 1

## Contact SAS

| Location | Contact | Address |
| --- | --- | --- |
| SAS Institute Inc. | Toll Free: (877) 846-3589<br>Phone: (919) 677–8000<br>Fax: (919) 677–4444 | 100 SAS Campus Drive<br>Cary, NC 27513–2414 USA |
| Technical Support | Phone: 919-677-8008<br>E-mail: techsupport@sas.com<br>Web: http://support.sas.com/techsup/contact/ | |
| Documentation Support | E-mail: yourturn@sas.com | |