



THE
POWER
TO KNOW.

DataFlux[®] Data Management Studio 2.5

Installation and Configuration Guide

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2013. *DataFlux® Data Management Studio 2.5: Installation and Configuration Guide*. Cary, NC: SAS Institute Inc.

DataFlux® Data Management Studio 2.5: Installation and Configuration Guide

Copyright © 2013, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

For a hard-copy book: No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

For a web download or e-book: Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others' rights is appreciated.

U.S. Government Restricted Rights Notice: Use, duplication, or disclosure of this software and related documentation by the U.S. government is subject to the Agreement with SAS Institute and the restrictions set forth in FAR 52.227-19, Commercial Computer Software-Restricted Rights (June 1987).

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

October 2013

SAS provides a complete selection of books and electronic products to help customers use SAS® software to its fullest potential. For more information about our e-books, e-learning products, CDs, and hard-copy books, visit support.sas.com/bookstore or call 1-800-727-3228.

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

Contents

<i>Accessibility Features of DataFlux Data Management Studio</i>	<i>v</i>
<i>Recommended Reading</i>	<i>vii</i>
Chapter 1 • Installing DataFlux Data Management Studio	1
Installing the Software for the First Time	1
Upgrading Existing Installations of the Software	1
System Requirements	4
Supported Databases for Data Storage	4
Supported Databases for Repository Storage	6
Chapter 2 • Configuring DataFlux Data Management Studio	11
About Configuration Files	11
DataFlux Folder Permissions	12
Configuration Options	12
Data Access Component Directives	35
Logging Options	40
Chapter 3 • Installing Add-On Products	43
Installing QKBs and Data Packs	43
Installing Supplemental Language Support	44
Setup for the Address Update Add-On	44
Installing Support for the SAP RFC Node	49
Appendix 1 • Contact SAS	51

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

Chapter 1

Installing DataFlux Data Management Studio

Installing the Software for the First Time	1
Upgrading Existing Installations of the Software	1
Migrating to SAS 9.3	1
Migrating to SAS 9.4	3
System Requirements	4
Supported Databases for Data Storage	4
Supported Databases for Repository Storage	6
Overview	6
Database Storage for Repositories	7
File Storage for Repositories	8

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 11](#). To install add-on products, see [“Installing Add-On Products” on page 43](#).

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 <i>DataFlux Migration Guide</i> for instructions about migrating from DataFlux Data Management Platform 2.2 to 2.3 or 2.4.
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.

Migrating From	Migrating To	Installed with the SAS Deployment Wizard	Migration Instructions
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 <i>DataFlux Migration Guide</i> 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.

Migrating From	Migrating To	Installed with the SAS Deployment Wizard	Migration Instructions
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.

Supported Databases for Data Storage

DataFlux Data Management Studio can use DataFlux ODBC 7.0 drivers to access data in the following databases:

Driver	Database Version
Btrieve®	Btrieve 6.15
Btrieve	Pervasive.SQL™ 7.0, 8.5, and 2000
dBASE	Clipper™
dBASE™	dBASE IV and V
dBASE	FoxPro 2.5, 2.6, and 3.0
dBASE	FoxPro 6.0 (with 3.0 functionality only)
dBASE	FoxPro 3.0 Database Containers (DBC)
DB2® Wire Protocol	IBM® DB2 v9.1, v9.5, and v9.7 for Linux®, UNIX, and Windows®
DB2 Wire Protocol	IBM DB2 Universal Database™ (UDB) v7.x and v8.x for Linux, UNIX, and Windows

Driver	Database Version
DB2 Wire Protocol	IBM DB2 v9.1 and v10 for z/OS
DB2 Wire Protocol	IBM DB2 UDB v7.x and v8.1 for z/OS
DB2 Wire Protocol	IBM DB2 UDB V5R1, V5R2, V5R3, V5R4, V6R1, and V7R1 for iSeries®
Greenplum Wire Protocol	Greenplum Database 3.1, 3.2, 3.3, 4, and 4.1
Informix® (client)	Informix Dynamic Server 9.2, 9.3, 9.4, 10.0, 11.0, 11.5, and 11.7
Informix Wire Protocol	Informix Dynamic Server 9.2, 9.3, 9.4, 10.0, 11.0, 11.5, and 11.7
MySQL™ Wire Protocol	MySQL 5.0x, 5.1, and 5.5
Oracle® (client)	Oracle 8.0.5 and higher (32-bit only)
Oracle (client)	Oracle 8i R2 and R3 (8.1.6 and 8.1.7)
Oracle (client)	Oracle 9i R1 and R2 (9.0.1 and 9.2)
Oracle (client)	Oracle 10g R1 and R2 (10.1 and 10.2)
Oracle (client)	Oracle 11g R1 and R2 (11.1 and 11.2)
Oracle Wire Protocol	Oracle 8i R2, R3 (8.1.6 and 8.1.7)
Oracle Wire Protocol	Oracle 9i R1 and R2 (9.0.1 and 9.2)
Oracle Wire Protocol	Oracle 10g R1 and R2 (10.1 and 10.2)
Oracle Wire Protocol	Oracle 11g R2 (11.2)
PostgreSQL® Wire Protocol	PostgreSQL 8.2, 8.3, 8.4, 9.0, and 9.1
Progress OpenEdge Wire Protocol	Progress OpenEdge 10.1x, 10.2x, and 11.0
Salesforce	Salesforce (API Version 23)
SQL Server Legacy Wire Protocol	Microsoft® SQL Server® 7.0
SQL Server Legacy Wire Protocol	Microsoft SQL Server 2000
SQL Server Legacy Wire Protocol	Microsoft SQL Server 2005
SQL Server Legacy Wire Protocol	Microsoft SQL Server 2008 R1, R2
SQL Server Wire Protocol	Microsoft SQL Server 2000

Driver	Database Version
SQL Server Wire Protocol	Microsoft SQL Server 2005
SQL Server Wire Protocol	Microsoft SQL Server 2008 R1, R2
Sybase® Wire Protocol	Sybase Adaptive Server® 11.5 and 11.9
Sybase Wire Protocol	Sybase Adaptive Server Enterprise® 12.0, 12.5x, 15.0, 15.5, and 15.7
Sybase IQ	Sybase IQ 15.0, 15.1, 15.2, and 15.3
Teradata®	Teradata 12.0, 13.0, and 13.1
Teradata	Teradata V2R6.0, V2R6.1, and V2R6.2
Text	Text Files
XML	XML Documents (tabular and hierarchical formatted)

Supported Databases for Repository Storage

Overview

To add or update a repository, access the Repository Definition window to create a repository definition file (.RCF file). There are two main sections of the Repository Definition window, as shown in the following display:

Edit Repository Definition

Name: Web Repository ORA

Data storage
Specify a database for the storage of items requiring a database.

☐ Database file
Location: Browse...

☒ Database connection
Name:
Table prefix (optional):

File storage (optional)
Specify a folder for the storage of file-based items.

Folder:

☐ Connect to repository at startup
☐ Private

Database Storage for Repositories

A DataFlux Data Management Studio repository supports two types of storage: database storage and file storage.

database storage

specifies a database for the storage of data explorations, profiles, and all objects in the Business Rule Manager (rules, tasks, custom metrics, and so on). Supported databases include SQLite and other database formats. For a list of supported formats, see Database Storage for Repositories.

file storage

specifies a separate storage location for objects that are stored as files, such as data jobs, process jobs, queries, *.sas files (SAS code files), and Entity Resolution Output files (*.sri files). For a list of supported formats, see File Storage for Repositories.

The following databases can be used for the database storage section of a repository:

Database	ODBC Driver
DB2 V8.x, V9.1, V9.5, and V9.7 for Linux, UNIX, and Windows	X

Database	ODBC Driver
DB2 Universal Database (UDB) v7.x for Linux, UNIX, and Windows	X
Informix Server 10.0, 11.0, and 11.5	X
Microsoft SQL Server 2008 R1, R2	X
Microsoft SQL Server 2005	X
Microsoft SQL Server 2000 Enterprise Edition (64-bit)	X
Microsoft SQL Server 2000 Desktop Engine (MSDE 2000)	X
Microsoft SQL Server 2000 Service Packs 1, 2, 3, 3a and 4	X
Oracle 11g R1, R2 (11.1, 11.2)	X
Oracle 10g R1, R2(10.1, 10.2)	X
Oracle 9i R1, R2 (9.0.1, 9.2)	X
Sybase Adaptive Server 12.0, 12.5x, 15, and 15.5	X
Teradata 12.0 and 13.0	X

Note: Due to locking issues, file-based repositories and Microsoft Access based repositories are not recommended for production systems.

File Storage for Repositories

The file storage location can be any physical path that is accessible by the software. Because DataFlux Data Management Studio is a Windows application, any location with a supported version of Windows can be used for file storage. For more information, see [“System Requirements” on page 4](#).

It is possible to specify a UNIX location for file storage if this location is accessible to DataFlux Data Management Studio as a network drive. The following versions of UNIX are supported:

- AIX: Version 5.3 and 6.1 on POWER architectures
- HP-UX PA-RISC: HP-UX 11iv2 (11.23), 11iv3 (11.31)
- HP-UX Itanium: HP-UX 11iv2 (11.23), 11iv3 (11.31)
- Linux for x86 (x86-32)
- Linux on x64
- Solaris on SPARC: Version 8, 9, 10

- Solaris on x64: Version 10

Chapter 2

Configuring DataFlux Data Management Studio

About Configuration Files	11
DataFlux Folder Permissions	12
Configuration Options	12
Data Access Component Directives	35
Logging Options	40

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

The following table outlines the recommended operating system permissions for users.

Directories	Users	Default Permissions
DataFlux-home	Administrators, Installer	Full control
DataFlux-home \Data Management Studio	Process user	Read and Execute, List Folder Contents
DataFlux repositories (which can exist anywhere)	Process user	Read and Execute, List Folder Contents

Users need to be able to access the content and make changes to the **DataFluxhome/etc** location. Also, they should have access to their own user location in **%APPDATA%/DataFlux/**.

Configuration Options

The main configuration options for DataFlux Data Management Studio are as follows:

Option	Purpose	Source	Notes
Base or General Application			
ODBC_INI	Overrides the location of the odbc.ini file		UNIX only.
BASE/ APP_CONTAINER_ DOMAIN	Application container authentication domain	Optional	Identifies the authentication domain expected by application container services. If not specified, “DefaultAuth” is used.

Option	Purpose	Source	Notes
BASE/ APP_CONTAINER_ LOC	Application container location	Optional	<p>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.</p> <p>In addition, app.cfg should always set this option to point to the metadata server (iom://<host><port>).</p>
BASE/APP_VER	Application version number	Optional	Defaults to 2.5.

Option	Purpose	Source	Notes
BASE/ AUTH_DEFAULT_ DOMAIN	Default resolved identity domain	Optional	<p>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:</p> <ol style="list-style-type: none"> 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_L OC	Location of the authenticating server	Optional	<p>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 <i>SAS National Language Support (NLS): Reference Guide</i>.</p>

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.
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.
BASE/EXE_PATH	Path containing executables	Optional	Calculated.

Option	Purpose	Source	Notes
BASE/FTPGETCMD	Specifies the command used for Ftp Get Functionality	Required	<p>Should default in the install, as follows:</p> <ul style="list-style-type: none"> • %U: Replace with user name. • %P: Replace with password. • %S: Replace with server. • %T: Replace with local directory. • %F: Replace with Files to download, multiple separated by spaces. • %L: Replace with the log file to pipe the output.
BASE/FTPPUTCMD	Specifies the command used for Ftp Put Functionality	Required	BASE/FTPPUTCMD
BASE/ JOB_LOG_ENCODING	Encoding for the job log on a DataFlux Data Management Server	Optional	<p>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.</p> <p>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.</p>
BASE/ LIBRARY_PATH	Path for Java JAR dependencies	Optional	Determined by start-up code (DFEXEC_HOME/lib).

Option	Purpose	Source	Notes
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.
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.

Option	Purpose	Source	Notes
BASE/ MONITOR_FREQU ENCY	Enables the logging of job node statistics while a job is running.		This option is 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_LICENS E	Primary licensing method	Required by base	Must be set in the configuration file as DATAFLUX or SAS
BASE/ PRIMARY_LICENS E_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.

Option	Purpose	Source	Notes
BASE/ REPOS_SYS_PATH	System path for repository configuration files	Optional	Automatically determined.
BASE/ REPOS_USER_PATH	User directory for repository configuration files	Optional	Automatically determined by dfcurver.
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_WAIT_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_CLEAN_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_LICENSE	Secondary licensing method	Required by base	Must be set in the configuration file as DATAFLUX or SAS.
BASE/ SECONDARY_LICENSE_LOC	Specifies the location of the secondary license file or server.	Required by base	Must be set in the configuration file.

Option	Purpose	Source	Notes
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, which is the 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.
BASE/TEMP	Temporary directory	Optional	If not specified, it inherits the value of the TEMP environment variable.
BASE/ TEXTMINE_LITI_L ANG_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.

Option	Purpose	Source	Notes
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/DISABLESYSCATE NUM	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.
DAC/DFTKDISABLECE DA	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/dfkdsn.
DAC/SAVEDCONNSYSTEM	Location of system-saved connections	Optional	Defaults to DFEXEC_HOME/etc/dsn.

Option	Purpose	Source	Notes
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)			
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/ REPOSCONNECTION	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/REPOSCONNECTION. 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.	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.	Optional	Valid values are "ALL" or an empty string. Using "ALL" requires a large amount of memory. Resides in macros/ncoa.cfg.
Pooling			
<p><i>Note:</i> 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 DMSEVER, APISVC - DFAPIS services (in the works)</p>			
POOLING/ CHILD_MAXIMUM_LAUNCHES	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.

Option	Purpose	Source	Notes
POOLING/ GET_PROCESS_T IMEOUT	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.
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_ERRO RS	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_PRO CESSES	Maximum number of concurrent pooled processes	Optional	If 0, the number of concurrent pooled processes is unlimited. Default is unlimited. If POOLING/ GET_PROCESS_T IMEOUT is set, it waits for that amount of time to get a new process if needed. This is a puddle option.

Option	Purpose	Source	Notes
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.
Process Flow			
WFE/ CANCEL_TIMEOU T	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_ LIMIT	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_FREQUE NCY	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_BYTE S	Frequency distribution engine per table bytes	Optional	Any numeric value. Default is -1 (frequency distribution engine default).
QKB			
CUSTOMIZE/ DISABLE_FILE_NO TIFICATIONS	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 <i>CUSTOMIZE/ DISABLE_FILE_NO TIFICATIONS=1</i> .
QKB/ ALLOW_INCOMPA T	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.

Option	Purpose	Source	Notes
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).
QKB/ON_DEMAND	Loads QKB definitions on demand	Optional	Default is YES. 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 continues to have its own initialization.
QKB/PATH	Path to QKB	Required by QKB products	Path is set to the default QKB defined in application.
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)			

Option	Purpose	Source	Notes
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	
JAVA/DEBUGPORT	Specifies the port to remotely debug Java.	Optional	
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 <i>DataFlux Data Management Studio Users Guide</i> .
VERIFYWORLD/D B	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 <i>DataFlux Data Management Studio: User's Guide</i> .
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.
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.

Option	Purpose	Source	Notes
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			
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 installationdir/etc/udf .

Option	Purpose	Source	Notes
JAVA/COMMAND	Command used to launch Java	Optional	<p>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:</p> <pre> JAVA/COMMAND = java JAVA/COMMAND = java - Djavax.net.ss l.trustStore= C:\Store \jssecacerts JAVA/COMMAND = java - Djavax.net.ss l.trustStore= "C:\Cert Store \jssecacerts" JAVA/COMMAND = "C:\Program Files\Java \jre6\bin \java"</pre>
MDM/ REPOSITORY_ROOT_FOLDER		Optional	<p>Name and location of the root folder for Master Data Management within a repository. This option allows the foundations/master_data to be overwritten by the end user when putting the contents of [INSTALL_ROOT]/share/mdm into a repository.</p>
MONITOR/ BULK_ROW_SIZE	Specifies the number of rows in a bulk load.		<p>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.</p>

Option	Purpose	Source	Notes
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.
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 49 .

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	<p>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.</p> <p>To turn on the 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:</p> <p>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).</p>

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	<p>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.</p> <p>To turn on the 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:</p> <p>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).</p>

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 12 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/SAVEDCONNSYST EM 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.
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.

Setting	Description	Windows	UNIX
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/DFTKDISABLECE DA configuration value, which should specify any non-null value (for example, yes).	The same as Windows.
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 \oratum38real for dword value of 1. If it is not found in USER, look in SYSTEM \oratum38real.	Look in \$HOME/.dfpower/dsn.cfg for a line dsn_name = oratum38real where dsn_name is the name of the DSN.

Setting	Description	Windows	UNIX
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.
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/SAVEDCONNSYSTEM 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/etc/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 dfktsrv path and core directory should be specified.	Look for config value DAC/DFTK_PROCESS_PATH. 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.

Setting	Description	Windows	UNIX
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\dfkdsn\.	Look in the same config value. If it is not found, look in \$DFEXEC_HOME/etc/dfkdsn/.
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 tktsleep 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. 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.
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 12.

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.

Name	Scope	Level Thresholds	Configuration Location
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

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 “[Configuration Options](#)” on page 12 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>
```


Chapter 3

Installing Add-On Products

Installing QKBs and Data Packs	43
Installing Supplemental Language Support	44
Setup for the Address Update Add-On	44
Overview	44
Obtain a License for DataFlux Products and USPS Components	44
Install DataFlux Software	44
Install Address Update Data from USPS	45
Update the ncoa.cfg File	47
Update the app.cfg File	48
Setup Continues in DataFlux Data Management Studio	49
USPS Copyright Disclosure	49
Installing Support for the SAP RFC Node	49

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.

QKB: [DataFlux QKBs](#) and [SAS QKBs](#)

DataPacks: [DataFlux Data Updates](#)

You can also download these and other products from [SAS DataFlux Software Products](#) 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 will 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.ft.1 clk.ft.2 and so on	
C:\NCOA Data\DVD 2\	clk.a.hf0 clk.b.hf0 and so on	
C:\NCOA Data\DVD 2\ CLK \	business.lst clk.ft.3 and so on	Folder unzipped from CLK.ZIP
C:\NCOA Data\DVD 2\ RV9\	cln.dat cln.txt crd.dat 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	

Subfolder	Files	Source
C:\NCOA Data\ DVD 2\CLK\		Folder unzipped from CLK.ZIP
	business.lst	
	clk.ftt.3	
	and so on	
C:\NCOA Data\ DVD 2\RV9\		Folder unzipped from RV9.ZIP
	cln.dat	
	cln.txt	
	crd.dat	
	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 CASS and LACS configuration. Set this option in the ncoa.cfg file.

NCOA/DFAV_CACHE_SIZE

Range: 0 through 100 and indicates how much 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. 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 Phone: (919) 677-8000 Fax: (919) 677-4444	100 SAS Campus Drive Cary, NC 27513-2414 USA
Technical Support	Phone: 919-677-8008 E-mail: techsupport@sas.com Web: http:// support.sas.com/ techsup/contact/	
Documentation Support	E-mail: yourturn@sas.com	

