DataFlux Data Management Studio

DataFlux
A sas COMPANY

YOUR DATA.
YOUR BUSINESS.
ONE SOLUTION.

This page is intentionally blank

DataFlux Data Management Studio

Installation and Configuration Guide

Version 2.1.1

September 17, 2010

This page is intentionally blank

Contact DataFlux

Corporate Headquarters

DataFlux Corporation

940 NW Cary Parkway, Suite 201
Cary, NC 27513-2792
Toll Free Phone: 877-846-FLUX (3589)
Toll Free Fax: 877-769-FLUX (3589)
Local Phone: 1-919-447-3000
Local Fax: 919-447-3100
Web: http://www.dataflux.com

DataFlux United Kingdom

Enterprise House
1-2 Hatfields
London
SE1 9PG
Phone: +44 (0) 20 3176 0025

DataFlux Germany

In der Neckarhelle 162
69118 Heidelberg
Germany
Phone: +49 (0) 6221 4150

DataFlux France

Immeuble Danica B
21, avenue Georges Pompidou
Lyon Cedex 03
69486 Lyon
France
Phone: +33 (0) 4 72 91 31 42

Technical Support

Phone: 1-919-531-9000
Email: techsupport@dataflux.com
Web: http://www.dataflux.com/MyDataFlux-Portal

Documentation Support

Email: docs@dataflux.com

Legal Information

Copyright © 1997 - 2010 DataFlux Corporation LLC, Cary, NC, USA. All Rights Reserved.

DataFlux and all other DataFlux Corporation LLC product or service names are registered trademarks or trademarks of, or licensed to, DataFlux Corporation LLC in the USA and other countries. ® indicates USA registration.

DataFlux Legal Statements

DataFlux Solutions and Accelerators Legal Statements

DataFlux Legal Statements

Apache Portable Runtime License Disclosure

Copyright © 2008 DataFlux Corporation LLC, Cary, NC USA.

Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

Apache/Xerces Copyright Disclosure

The Apache Software License, Version 1.1

Copyright © 1999-2003 The Apache Software Foundation. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. The end-user documentation included with the redistribution, if any, must include the following acknowledgment:

   "This product includes software developed by the Apache Software Foundation (http://www.apache.org)."

   Alternately, this acknowledgment may appear in the software itself, if and wherever such third-party acknowledgments normally appear.

4. The names "Xerces" and "Apache Software Foundation" must not be used to endorse or promote products derived from this software without prior written permission. For written permission, please contact apache@apache.org.

5. Products derived from this software may not be called "Apache", nor may "Apache" appear in their name, without prior written permission of the Apache Software Foundation.

THIS SOFTWARE IS PROVIDED "AS IS'' AND ANY EXPRESSED OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE APACHE SOFTWARE FOUNDATION OR ITS CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

This software consists of voluntary contributions made by many individuals on behalf of the Apache Software Foundation and was originally based on software copyright (c) 1999, International Business Machines, Inc., http://www.ibm.com. For more information on the Apache Software Foundation, please see http://www.apache.org.

DataDirect Copyright Disclosure

Portions of this software are copyrighted by DataDirect Technologies Corp., 1991 - 2008.

Expat Copyright Disclosure

Part of the software embedded in this product is Expat software.

Copyright © 1998, 1999, 2000 Thai Open Source Software Center Ltd.

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

gSOAP Copyright Disclosure

Part of the software embedded in this product is gSOAP software.

Portions created by gSOAP are Copyright © 2001-2004 Robert A. van Engelen, Genivia inc. All Rights Reserved.

THE SOFTWARE IN THIS PRODUCT WAS IN PART PROVIDED BY GENIVIA INC AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

IBM Copyright Disclosure

ICU License - ICU 1.8.1 and later [used in DataFlux Data Management Platform]

COPYRIGHT AND PERMISSION NOTICE

Copyright © 1995-2005 International Business Machines Corporation and others. All Rights Reserved.

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, provided that the above copyright notice(s) and this permission notice appear in all copies of the Software and that both the above copyright notice(s) and this permission notice appear in supporting documentation.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR HOLDERS INCLUDED IN THIS NOTICE BE LIABLE FOR ANY CLAIM, OR ANY SPECIAL INDIRECT OR CONSEQUENTIAL DAMAGES, OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

Except as contained in this notice, the name of a copyright holder shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization of the copyright holder.

Microsoft Copyright Disclosure

Microsoft®, Windows, NT, SQL Server, and Access, are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

Oracle Copyright Disclosure

Oracle, JD Edwards, PeopleSoft, and Siebel are registered trademarks of Oracle Corporation and/or its affiliates.

PCRE Copyright Disclosure

A modified version of the open source software PCRE library package, written by Philip Hazel and copyrighted by the University of Cambridge, England, has been used by DataFlux for regular expression support. More information on this library can be found at: ftp://ftp.csx.cam.ac.uk/pub/software/programming/pcre/.

Copyright © 1997-2005 University of Cambridge. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- Neither the name of the University of Cambridge nor the name of Google Inc. nor the names of their contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Red Hat Copyright Disclosure

Red Hat® Enterprise Linux®, and Red Hat Fedora™ are registered trademarks of Red Hat, Inc. in the United States and other countries.

SAS Copyright Disclosure

Portions of this software and documentation are copyrighted by SAS® Institute Inc., Cary, NC, USA, 2009. All Rights Reserved.

SQLite Copyright Disclosure

The original author of SQLite has dedicated the code to the public domain. Anyone is free to copy, modify, publish, use, compile, sell, or distribute the original SQLite code, either in source code form or as a compiled binary, for any purpose, commercial or non-commercial, and by any means.

Sun Microsystems Copyright Disclosure

Java™ is a trademark of Sun Microsystems, Inc. in the U.S. or other countries.

Tele Atlas North American Copyright Disclosure

Portions copyright © 2006 Tele Atlas North American, Inc. All rights reserved. This material is proprietary and the subject of copyright protection and other intellectual property rights owned by or licensed to Tele Atlas North America, Inc. The use of this material is subject to the terms of a license agreement. You will be held liable for any unauthorized copying or disclosure of this material.

USPS Copyright Disclosure

National ZIP®, ZIP+4®, Delivery Point Barcode Information, DPV, RDI. © United States Postal Service 2005. ZIP Code® and ZIP+4® are registered trademarks of the U.S. Postal Service.

DataFlux 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.

VMware

DataFlux Corporation LLC technical support service levels should not vary for products running in a VMware® virtual environment provided those products faithfully replicate the native hardware and provided the native hardware is one supported in the applicable DataFlux product documentation. All DataFlux technical support is provided under the terms of a written license agreement signed by the DataFlux customer.

The VMware virtual environment may affect certain functions in DataFlux products (for example, sizing and recommendations), and it may not be possible to fix all problems.

If DataFlux believes the virtualization layer is the root cause of an incident; the customer will be directed to contact the appropriate VMware support provider to resolve the VMware issue and DataFlux shall have no further obligation for the issue.

Solutions and Accelerators Legal Statements

Components of DataFlux Solutions and Accelerators may be licensed from other organizations or open source foundations.

Apache

This product may contain software technology licensed from Apache.

Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at: http://www.apache.org/licenses/LICENSE-2.0.

Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.

See the License for the specific language governing permissions and limitations under the License.

Creative Commons Attribution

This product may include icons created by Mark James http://www.famfamfam.com/lab/icons/silk/ and licensed under a Creative Commons Attribution 2.5 License: http://creativecommons.org/licenses/by/2.5/.

Degrafa

This product may include software technology from Degrafa (Declarative Graphics Framework) licensed under the MIT License a copy of which can be found here: http://www.opensource.org/licenses/mit-license.php.

Copyright © 2008-2010 Degrafa. All rights reserved.

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

Google Web Toolkit

This product may include Google Web Toolkit software developed by Google and licensed under the Apache License 2.0.

JDOM Project

This product may include software developed by the JDOM Project (http://www.jdom.org/).

OpenSymphony

This product may include software technology from OpenSymphony. A copy of this license can be found here: http://www.opensymphony.com/osworkflow/license.action. It is derived from and fully compatible with the Apache license that can be found here: http://www.apache.org/licenses/.

Sun Microsystems

This product may include software copyrighted by Sun Microsystems, jaxrpc.jar and saaj.jar, whose use and distribution is subject to the Sun Binary code license.

This product may include Java Software technologies developed by Sun Microsystems, Inc. and licensed to Doug Lea.

The Java Software technologies are copyright © 1994-2000 Sun Microsystems, Inc. All rights reserved.

This software is provided "AS IS," without a warranty of any kind. ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT, ARE HEREBY EXCLUDED. DATAFLUX CORPORATION LLC, SUN MICROSYSTEMS, INC. AND THEIR RESPECTIVE LICENSORS SHALL NOT BE LIABLE FOR ANY DAMAGES SUFFERED BY LICENSEE AS A RESULT OF USING, MODIFYING OR DISTRIBUTING THE SOFTWARE OR ITS DERIVATIVES. IN NO EVENT WILL SUN MICROSYSTEMS, INC. OR ITS LICENSORS BE LIABLE FOR ANY LOST REVENUE, PROFIT OR DATA, OR FOR DIRECT, INDIRECT, SPECIAL, CONSEQUENTIAL, INCIDENTAL OR PUNITIVE DAMAGES, HOWEVER CAUSED AND REGARDLESS OF THE THEORY OF LIABILITY, ARISING OUT OF THE USE OF OR INABILITY TO USE SOFTWARE, EVEN IF SUN MICROSYSTEMS, INC. HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Java Toolkit

This product includes the Web Services Description Language for Java Toolkit 1.5.1 (WSDL4J). The WSDL4J binary code is located in the file wsdl4j.jar.

Use of WSDL4J is governed by the terms and conditions of the Common Public License Version 1.0 (CPL). A copy of the CPL can be found here at http://www.opensource.org/licenses/cpl1.0.php.

Table of Contents

Introduction to the Documentation 1

Conventions Used in this Document 1

Reference Publications 2

Installation 3

Installing Data Management Studio 3

System Requirements 7

Data Storage 7

Repository Storage 9

Supported Platforms 10

Configuration 13

Configuration Files 13

Configuration Options 14

Data Job Configuration Directives 19

Data Access Component Directives 25

Add-On Products 28

Installing a Quality Knowledge Base 28

Installing Data Packs 28

Installing Supplemental Language Support 35

Technical Support 36

Frequently Asked Questions (FAQ) 36

Glossary 40

Introduction to the Documentation

- Conventions Used in this Document
- DataFlux References

Conventions Used in this Document

This document uses several conventions for special terms and actions.

Typographical Conventions

The following typographical conventions are used in this document:

Typeface	Description
Bold	Text in bold signifies a button or action
italic	Identifies document and topic titles
`monospace`	Typeface used to indicate filenames, directory paths, and examples of code

Syntax Conventions

The following syntax conventions are used in this document:

Syntax	Description
[]	Brackets [] are used to indicate variable text, such as version numbers
#	The pound # sign at the beginning of example code indicates a comment that is not part of the code
>	The greater than symbol is used to show a browse path, for example **Start** > **Programs** > **DataFlux Data Management Studio** 2.1 > **Documentation**.

Reference Publications

This document might reference other DataFlux® publications including:

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 Data Management Studio User's Guide

DataFlux Expression Language Reference Guide

DataFlux Federation Server Administrator's Guide

DataFlux Federation Server User's Guide

DataFlux Migration Guide

DataFlux Quality Knowledge Base Online Help

Installation

- Installing Data Management Studio
- System Requirements
- Data Storage
- Repository Storage
- Supported Platforms

Installing Data Management Studio

The installation process for DataFlux® Data Management Studio consists of the following stages:

- Obtaining the Microsoft .Net Framework
- Licensing Data Management Studio
- Running the Data Management Studio Installation Wizard
- Installing the Quality Knowledge Base
- Performing Reinstallations

Obtaining the Microsoft .Net Framework

Go to the download page for Microsoft® .NET Framework 3.5 Service Pack 1 and download the software.

Licensing Data Management Studio

Three licensing options are available for Data Management Studio:

- DataFlux License Server - Multiple computers can share a DataFlux license file on a DataFlux License Server.
- DataFlux License File - A single computer can access a DataFlux license file on a file system.
- SAS License File - A single computer can access a SAS license file on a file system.

DataFlux License Server

Perform the following steps to enable multiple computers to share a DataFlux license file on a DataFlux License Server:

1. Download the **License Manager** from the DataFlux MyPortal site http://www.dataflux.com/MyDataFlux-Portal.
2. Install the **License Manager** on the computer that will be your license server. Double-click the installation package and follow the instructions. (A licensing user guide is installed in the folder where the license server is installed. Later, you will use the instructions in this user guide to start the licensing server.)
3. Run the **lmhostid** command, which generates a host ID for your license server.
4. Email the host ID to your DataFlux representative and request a license file.
5. When you receive the license file, save it on the computer that will be your license server.
6. Start the license server according to the instructions in the licensing user guide.
7. On the computer where Data Management Studio is installed, select **Start** > **Programs** > **DataFlux** > **License Manager 2.1**.
8. In the License Manager dialog, select **DataFlux License** as the primary licensing method.
9. In the **DataFlux license file** section of the dialog, specify the location of the license server. This value will be something like @server.dataflux.com.
10. When finished, click **OK**.

DataFlux License File

Perform the following steps to enable a single computer to access a DataFlux license file on a file system:

1. On a computer where Data Management Studio is installed, select **Start** > **Programs** > **DataFlux**> **Show Host ID**. The host ID is displayed.
2. Email the host ID to your DataFlux representative and request a license file.
3. When you receive the license file, save it in the license folder where Data Management Studio is installed.
4. On the computer where Data Management Studio is installed, select **Start** > **Programs** > **DataFlux**> **License Manager 2.1**.
5. In the License Manager dialog, select **DataFlux License** as the primary licensing method.
6. In the **DataFlux license file** section of the dialog, specify the location of the license file.
7. When finished, click **OK**.

SAS License File

Perform the following steps to enable a single computer to access a SAS license file on a file system:

1. Obtain a SAS license file from your SAS representative.
2. Save the SAS license file to a location that is accessible to Data Management Studio.
3. On the computer where Data Management Studio is installed, select **Start** > **Programs** > **DataFlux**> **License Manager 2.1**. The License Manager dialog is displayed.
4. In the License Manager dialog, select **SAS License** as the primary licensing method.
5. In the **SAS license file** section of the dialog, specify the location of the license file.
6. When finished, click **OK**.

Annual Licensing Notification

For DataFlux licenses, thirty days prior to license expiration, you will receive a message that your license will expire in a certain number of days. For SAS licenses (setinits), the expiration notice is defined by the warning period. The warning period is configurable through SAS.

> **Note:** DataFlux licenses are not configurable. Contact your DataFlux sales representative to renew your DataFlux product license(s).

Running the Data Management Studio Installation Wizard

Perform the following steps to run the installation wizard:

1. Go to the downloads section of MyDataFlux Portal on the DataFlux® website.
2. Find the Data Management Studio section and download the needed version.
3. Once downloaded, run the installation wizard.
4. Review the welcome window for the wizard. Click **Next** to access the **Destination Location** window.
5. Review the location of the destination folder. You can either accept the default location or click **Browse** to select a new location. Click **Next** to access the **Select Components** window.
6. Review the available components and select the ones that you need. Click **Next** to access the Licensing window.
7. Select a licensing method and specify any information required for that method. Click **Next** to access the **Start Installation** window.
8. Click **Next** to access the **Installing** window and begin installing Data Management Studio files.
9. Review the installation summary. Click **Finish** to exit the wizard.

Installing the Quality Knowledge Base

Perform the following steps to install one or more Quality Knowledge Bases (QKBs):

1. Go to the downloads section of MyDataFlux Portal on the DataFlux website.
2. Find the QKB section and download the needed version.
3. Once downloaded, run the installation wizard.
4. Review the welcome window for the wizard. Click **Next** to access the **Quality Knowledge Base License** window.
5. Review the terms of the license. Click **Accept** to accept the terms and access the **Choose QKB Name** window.
6. Review the QKB name and location. Click **Next** to access the next window.
7. Select the locales that you need. Click **Next** to access the next window.
8. Decide whether this will be the active QKB. Select the appropriate option and click **Next** twice to install the QKB files.
9. Open Data Management Studio and click the **Administration** riser bar.
10. Right-click **Quality Knowledge Base** and click **New** in the pop-up menu.
11. Supply a name and the directory for the QKB that you installed.
12. Click **OK** to save the settings and exit the **Add Quality Knowledge Base Location** window.

Performing Reinstallations

If need to reinstall Data Management Studio, you should consider the following issues:

- You should always uninstall your existing Data Management Studio to remove any configuration file changes before you install a new image.
- You might need to upgrade your repositories. To perform the upgrade, click the **Administration** riser bar. Then, click Repository Definitions. Next, right-click a repository that needs the upgrade and click **Upgrade** in the pop-up menu. Right-click the upgraded repository and click **Connect** in the pop-up menu. Finally, close and restart the Data Management Studio application.

System Requirements

System requirements for Data Management Studio are as follows:

Requirement	Minimum	Recommended
Platforms	Microsoft® Windows XP® and Windows Vista®.	Microsoft Windows XP Professional
.Net Framework	Microsoft .NET Framework 3.5 Service Pack 1	Microsoft .NET Framework 3.5 Service Pack 1
Processor	Intel® Pentium® 4 - 1.2 GHz or higher	Intel Pentium 4 - 2.2 GHZ or higher
Memory (RAM)	512 MB	2+ GB
Disk Space	5 GB	10+ GB

See also the Supported Platforms for Data Management Studio.

Data Storage

DataFlux Data Management Studio can access data in the following databases:

Database	Driver	ODBC	TKTS Native	TKTS ODBC
Btrieve® 6.15	Btrieve	X		
Clipper™	dBASE File	X		
dBASE™ IV and V	dBASE	X		
FoxPro 2.5, 2.6, and 3.0	dBase	X		
FoxPro 6.0 (with 3.0 functionality only)	dBase	X		
FoxPro 3.0 Database Container	dBase	X		
Greenplum Database 3.1, 3.2, and 3.3	Greenplum Wire Protocol	X		
IBM® DB2® v9.1, v9.5, and v9.7 for Linux®, UNIX, and Windows®	DB2 Wire Protocol	X		
IBM DB2 Universal Database™ (UDB) v7.x, v8.x for Linux, UNIX, and Windows	DB2 Wire Protocol	X	Min 8.2 Fix pack 9	X
IBM DB2 v9.1 for z/OS	DB2 Wire Protocol	X		
IBM DB2 UDB v7.x and v8.1 for z/OS	DB2 Wire Protocol	X		
IBM DB2 UDB V5R1, V5R2, V5R3, V5R4, and V6R1 for iSeries®	DB2 Wire Protocol	X		
IBM Informix© Dynamic Server 11 and 11.5	Informix	X		
IBM Informix Dynamic Server 11 and 11.5	Informix Wire Protocol	X		

Database	Driver	ODBC	TKTS Native	TKTS ODBC
Informix© Dynamic Server 9.2, 9.3, 9.4, and 10	Informix	X		
Informix Dynamic Server 9.2, 9.3, 9.4, and 10	Informix Wire Protocol	X		
Microsoft Excel® Workbook 5.1, 7.0	Excel	X		
Microsoft® SQL Server® 7.0	SQL Server Classic Wire Protocol	X		X
Microsoft SQL Server 2000	SQL Server Classic Wire Protocol	X		X
Microsoft SQL Server 2005	SQL Server Classic Wire Protocol	X		
Microsoft SQL Server 2008	SQL Server Classic Wire Protocol	X		
Microsoft SQL Server 2000	SQL Server Native Wire Protocol	X		
Microsoft SQL Server 2005	SQL Server Native Wire Protocol	X		
Microsoft SQL Server 2008	SQL Server Native Wire Protocol	X		
MySQL™ 5.0 and 5.1	MySQL Wire Protocol	X		
Oracle® 8.0.5+	Oracle	X		X
Oracle 8i R2, R3 (8.1.6 and 8.1.7)	Oracle	X		X
Oracle 9i R1, R2 (9.0.1 and 9.2)	Oracle	X		X
Oracle 10g R1 and R2 (10.1 and 10.2)	Oracle	X		X
Oracle 11g R1 and R2 (11.1 and 11.2)	Oracle	X		X
Oracle 8i R2, R3 (8.1.6 and 8.1.7)	Oracle Wire Protocol	X		X
Oracle 9i R1 and R2 (9.0.1 and 9.2)	Oracle Wire Protocol	X		X
Oracle 10g R1 and R2 (10.1 and 10.2)	Oracle Wire Protocol	X	X	X
Oracle 11g R1 and R2 (11.1 and 11.2)	Oracle Wire Protocol	X		
Paradox® 4, 5, 7, 8, 9, and 10	ParadoxFile	X		
Pervasive.SQL™ 7.0 and 2000	Btrieve	X		
PostgreSQL® 8.2, 8.3, and 8.4	PostgreSQL Wire Protocol	X		
Sybase® Adaptive Server® 11.5 and 11.9	Sybase Wire Protocol	X		
Sybase Adaptive Server Enterprise® 12.0, 12.5x, and 15	Sybase Wire Protocol	X		
Teradata® 12.0	Teradata	X		
Teradata V2R6.0, V2R6.1, and V2R6.2	Teradata	X	X	X

Database	Driver	ODBC	TKTS Native	TKTS ODBC
Text Files	Text	X		
XML Documents (tabular and hierarchical formatted)	XML	X		

Repository Storage

- Database Storage for Repositories
- File Storage for Repositories

Database Storage for Repositories

A Data Management Studio repository supports two kinds of storage: database storage and file storage. The following databases can be used for the database storage section of a repository:

Database	ODBC Driver	Federation Server TKTS Native	Federation Server TKTS ODBC
DB2® Universal Database (UDB) v7.x, v8.1, and v8.2 for Linux®, UNIX, and Windows®	X	Min 8.2 FP 9	X
DB2 UDB v7.x and v8.1 for z/OS	X		
DB2 UDB V5R1, V5R2, and V5R3 for iSeries	X		
Microsoft® SQL Server® 6.5	X		X
Microsoft SQL Server 7.0	X		X
Microsoft SQL Server 2000 (including SP 1, 2, 3 and 3a)	X		X
Microsoft SQL Server 2000 Desktop Engine (MSDE 2000)	X		X
Microsoft SQL Server 2000 Enterprise (64-bit)	X		X
Microsoft SQL Server 2005	X		X
Oracle® 8.0.5+ Oracle 8i R1, R2, R3 (8.1.5, 8.1.6,8.1.7)	X		X
Oracle 9i R1, R2 (9.0.1, 9.2)	X		X
Oracle 10g R1 (10.1)	X	X	X
Sybase® Adaptive Server® 11.5 and higher	X		
Sybase Adaptive Server Enterprise® 12.0, 12.5, 12.5.1, 12.5.2 and 12.5.3	X		
Teradata® 12.0	X		

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

File Storage for Repositories

The following platforms can be used for the file storage section of a repository:

- 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
- Microsoft® Windows® (x86-32): Windows XP® Professional, Windows Server 2003®, Windows Vista®
- Microsoft Windows on x64 (EM64T/AMD64): Windows XP Professional for x64, Windows Vista for x64, Windows Server 2003 for x64
- Solaris™ on SPARC: Version 8, 9, 10
- Solaris on x64: Version 10

Supported Platforms

- Data Management Studio IDE Client
- DataFlux License Server
- Virtual Environments

Data Management Studio IDE Client

Operating System	Bit	Chip	SAS Platform
Windows® HPC Server 2008 Edition	32	x64	9.2
Windows Server 2003®, Data Center Edition (SP1 and SP2)	32	x64	9.2
Windows Server 2003, Data Center Edition (SP1 and SP2)		x86	9.1.2
Windows Server 2003, Data Center Edition - 32 bit compatibility mode (SP1 and SP2)	64	x64	
Windows Server 2003, Enterprise Edition (SP1 and SP2)	32	x64	9.2
Windows Server 2003, Enterprise Edition (SP1 and SP2)		x86	9.1.2
Windows Server 2003, Enterprise Edition - 32 bit compatibility mode (SP1 and SP2)	64	x64	
Windows Server 2003, Small Business Server (SP1 and SP2)	32	x64	9.2
Windows Server 2003, Small Business Server (SP1 and SP2)		x86	9.1.2
Windows Server 2003, Standard Edition (SP1 and SP2)	32	x64	9.2
Windows Server 2003, Standard Edition (SP1 and SP2)		x86	9.1.2

Operating System	Bit	Chip	SAS Platform
Windows Server 2003, Standard Edition - 32 bit compatibility mode (SP1 and SP2)	64	x64	
Windows Server 2003, Web Edition (SP1 and SP2)	32	x64	9.2
Windows Server 2003, Web Edition (SP1 and SP2)		x86	9.1.2
Windows Server 2008®, Data Center Edition	32	x64	9.2
Windows Server 2008, Data Center Edition		x86	9.1.2
Windows Server 2008, Data Center Edition - 32 bit compatibility mode	64	x64	
Windows Server 2008, Data Center without Hyper-V Edition	32	x64	9.2
Windows Server 2008, Data Center without Hyper-V Edition		x86	9.1.2
Windows Server 2008, Data Center without Hyper-V Edition - 32 bit compatibility mode	64	x64	
Windows Server 2008, Enterprise Edition	32	x64	9.2
Windows Server 2008, Enterprise Edition		x86	9.1.2
Windows Server 2008, Enterprise Edition - 32 bit compatibility mode	64	x64	
Windows Server 2008, Enterprise without Hyper-V Edition	32	x64	9.2
Windows Server 2008, Enterprise without Hyper-V Edition		x86	9.1.2
Windows Server 2008, Enterprise without Hyper-V Edition - 32 bit compatibility mode	64	x64	
Windows Server 2008, Foundation Edition	32	x64	9.2
Windows Server 2008, Foundation Edition		x86	9.1.2
Windows Server 2008, Foundation Edition - 32 bit compatibility mode	64	x64	
Windows Server 2008, Small Business Server	32	x64	9.2
Windows Server 2008, Small Business Server		x86	9.1.2
Windows Server 2008, Small Business Server- 32 bit compatibility mode	64	x64	
Windows Server 2008, Standard Edition	32	x64	9.2
Windows Server 2008, Standard Edition		x86	9.1.2
Windows Server 2008, Standard Edition - 32 bit compatibility mode	64	x64	
Windows Server 2008, Standard without Hyper-V Edition		x86	9.1.2
Windows Vista® Business	32	x64	9.2
Windows Vista Business- 32 bit compatibility mode	64	x64	
Windows Vista Enterprise	32	x64	9.2
Windows Vista Enterprise- 32 bit compatibility mode	64	x64	
Windows Vista Ultimate	32	x64	9.2
Windows Vista Ultimate- 32 bit compatibility mode	64	x64	
Windows XP® Professional (SP2)	32	x64	9.2

Operating System	Bit	Chip	SAS Platform
Windows XP Professional (SP2)		x86	9.1.2
Windows XP Professional (SP2)- 32 bit compatibility mode	64	x64	
Windows 7® Enterprise - 32 bit compatibility mode	64	x64	
Windows 7 Home Basic- 32 bit compatibility mode	64	x64	
Windows 7 Home Premium- 32 bit compatibility mode	64	x64	
Windows 7 Professional - 32 bit compatibility mode	64	x64	
Windows 7 Starter - 32 bit compatibility mode	64	x64	
Windows 7 Ultimate - 32 bit compatibility mode	64	x64	

DataFlux License Server

A DataFlux License Server can be installed on the following platforms:

Platform
AIX® 64-bit - Power PC™ RS/6000®
HP-UX® 64-bit - HP 64-bit
HP-UX 64-bit - Intel® Itanium®
Microsoft® Windows® 32-bit - x86
Red Hat® Enterprise Linux 32-bit - x86 / AMD Opteron™
Red Hat Enterprise Linux 64-bit - Intel Xeon© / AMD Opteron
Solaris™ 64-bit - SPARC© 64-bit
Solaris 64-bit - AMD Opteron
SUSE® Linux Enterprise Server 32-bit - x86 / AMD Opteron
SUSE Linux Enterprise Server 64-bit - Intel Xeon / AMD Opteron

Virtual Environments

Environments	Supported
Citrix®	Contact your DataFlux Sales Executive for more information.
Remote Desktop	DataFlux products supported in this environment.
Terminal Server	Contact your DataFlux Sales Executive for more information.
Virtual PC	DataFlux products supported in this environment.
VMWare®	DataFlux products supported in this environment.

Configuration

- Configuration Files
- Configuration Options
- Data Job Configuration Directives
- Data Access Component Directives

Configuration Files

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

Data Management Studio reads configuration settings in this order:

1. The app.cfg file in the etc folder where Studio 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 dis.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 above configuration files.
6. The macros folder in a user folder.
7. The environment variables.
8. The command-line options if applicable.

Configuration Options

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

Option	Purpose	In App.cfg By Default?	Source	Notes
General Application				
BASE/LIBRARY_PATH	Path for Java jar dependencies	No	Optional	Determined by startup code (DFEXEC_HOME/lib)
BASE/PLUGIN_PATH	Path used by all subsystems to find plugins	No	Optional	Determined by startup code
BASE/EXE_PATH	Path containing executables	No	Optional	Calculated
BASE/PRIMARY_ LICENSE	Primary licensing method	Yes	Required by base	Must be set in the DATAFLUX or SAS configuration file
BASE/MACROS_PATH	Path for system macros.cfg file	No	Optional	User to change the default location of the system macros.cfg file. By default this file is located in the **etc** subfolder of the installation folder.
BASE/PRIMARY_ LICENSE_LOC	Location of the primary license file or server	Yes	Required by base	Must be set in the configuration file
BASE/SECONDARY_ LICENSE	Secondary licensing method	Yes, but commented out	Required by base	Must be set in the DATAFLUX or SAS configuration file
BASE/SECONDARY_ LICENSE_LOC	Location of the secondary license file or server	Yes, but commented out	Required by base	Must be set in the configuration file
BASE/LOGCONFIG_ PATH	Full path to the log configuration file	No	Optional	Must be set in the configuration file (defaults to logging.xml)
BASE/MESSAGE_ PATH	Path to the message directory	No	Optional	Determined by startup code

Option	Purpose	In App.cfg By Default?	Source	Notes
BASE/MESSAGE_ LOCALE	Error message locale	No	Optional	If not specified, determined from the system locale
BASE/MESSAGE_ LEVEL	Error level of messages	No	Optional	0 (or not specified) - normal messages; 1 - includes source file and line number in messages
BASE/USER_PATH	Path for user configuration files	No	Optional	Determined by dfcurver
BASE/REPOS_SYS_ PATH	System path for repository configuration files	No	Optional	Automatically determined
BASE/REPOS_USER_ PATH	User directory for repository configuration files	No	Optional	Determined by dfcurver
BASE/TEMP	Temporary directory	No	Optional	If not specified, inherits the value of the TEMP environment variable
BASE/DATE_FORMAT	Specific date formats	No	Optional	If specified, iso8601
BASE/APP_VER	Application version number	No	Optional	Defaults to 2.1
BASE/UPDATE_LEVEL	Application update level	No	Optional	Defaults to 0. Could be used as a minor revision number
PROC_TXT_MACRO_ TEST		No		
DAC Logging				
DAC/DFTKLOGFILE	DFTK logging	No	Optional	Filename
DAC Logging (New in Data Management Studio)				
DAC/TKTSLOGFILE	TKTS logging	Yes	Optional	Filename
DAC/DFTKDISABLEC EDA	Disables CEDA support	No	Optional	"Yes" turns it on
DAC/SAVEDCONNSY STEM	Location of system saved connections	No	Optional	Defaults to DFEXEC_HOME/etc/dsn
DAC/SAVEDCONNUS ER	Location of user saved connections	No	Optional	Defaults to your application directory/DataFlux/dac/9.0
DAC/DSN	DSN directory for TKTS dsns	No	Optional	Defaults to DFEXEC_HOME/etc/dftkdsn

Option	Purpose	In App.cfg By Default?	Source	Notes
DAC/DFTK_PROCESS	Run DFTK out of process	No	Optional	"Yes" turns it on; off by default
DAC/DFTK_PROCESS_TKPATH	TKTS path for DFTK out of process	No	Optional	Defaults to a core/sasext dir off the executable dir
Profile (New in Data Management Studio)				
PROF/DEBUG_MODE	Frequency distribution engine debug mode	Yes, but commented out	Optional	0 not debug mode, 1 debug mode: default is not debug mode. The log is located at C:\Documents and Settings\<USER ID>\Local Settings\Temp.
PROF/PER_TABLE_BYTES	Frequency distribution engine per table bytes	Yes, but commented out	Optional	default is -1 (frequency distribution engine default)
QKB				
QKB/PATH	Path to QKB	Yes, but commented out	Required by QKB products	Maintained by the QKB installation
QKB/SURFACEALL	Surfaces all parse definitions	Yes	Optional	Default is NO
QKB (New in Data Management Studio)				
QKB/COMPATVER		Yes	Optional	Possible values: dfpower82, unity21 Default: unity21
QKB/ALLOW_INCOMPAT		Yes	Optional	Default is NO
QKB/ON_DEMAND		Yes	Optional	Default is YES
Base				
BASE/SORTBYTES	Specifies the bytes used in sorting	Yes	Optional	
CLUSTER/BYTES	Specifies the bytes used in clustering	Yes, but commented out	Optional	
CLUSTER/LOG	Specifies whether clustering log is needed	Yes, but commented out	Optional	

Option	Purpose	In App.cfg By Default?	Source	Notes
FRED/LOG	Specifies whether FRED log is needed	No	Optional	
BASE/TEMP		Yes, but commented out		
BASE/EMAILCMD	Specifies the command used to send email	Yes, but commented out	Required	Can include %T and %B where %T will be replaced with the recipient and %B will be a file containing the body of the message; also used by monitor event
MONITOR/REPOSFILE		N/A		
Base (New in Data Management Studio)				
BASE/SORTMERGES	Enables merge during sort	No	Optional	
BASE/SORTTEMP	Specifies the temporary path for sorts	No	Optional	
BASE/SORTTHREADS	Specifies the number of sort threads	No	Optional	
ARCHITECT/AutoPass Thru	Client option to set mappings	No	Optional	Maintained by client; choices are 0 (target), 1 (Source and Target), and 2 (All)
Verify				
VERIFY/CACHESIZE	Specifies a percentage value	Yes	Optional	
VERIFY/CANADA	Specifies the path to Canadian data	Yes, but commented out	Required by SERP nodes	Maintained by Canada installation
VERIFY/GEO	Specifies the geo/phone path	Yes, but commented out	Required by Geo	Maintained by Geo installation
VERIFY/PRELOAD	Specifies the preload string for verify	Yes, but blank	Optional	Valid values are ALL or empty string
VERIFY/USPS	Specifies the USPS data path	Yes	Required by USPS	Maintained by USPS installation

Option	Purpose	In App.cfg By Default?	Source	Notes
VERIFY/UPSPINST	Determines whether the USPS data is installed	Yes	Required	Maintained by USPS installation
VERIFYWORLD/DB	Specifies the Platon data path	Yes, but commented out	Required for Platon	Path maintained by component installation
VERIFYWORLD/UNLK	Specifies the Platon library universal unlock code	Yes, but commented out	Required for Platon	Path maintained by component installation
Verify (New in Data Management Studio)				
CLUSTER/TEMP	Specifies the cluster temporary path	No	Optional	
BASE/FTPGETCMD	Specifies the command used for Ftp Get Functionality	Yes, but commented out	Required	Should default in the install, as follows: • %U: Replace with username • %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	Yes, but commented out	Required	
dfIntelliServer				
DFCLIENT/CFG	Used for dfIntelliServer	No	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

Option	Purpose	In App.cfg By Default?	Source	Notes
Other				
EXPRESS_MAX_STRING_LENGTH	Specifies the Expression node	No	Optional	Default maximum length of any string in this node is 32k. This enables specifying a larger value in bytes

Data Job Configuration Directives

Each configuration file in Data Management Studio uses the following rules:

1. Comment lines start with #
2. Config directive lines use the key = value form, where each key needs to have '=' after it to be recognized
3. On a config directive line, everything after the first '=' is treated as a single value of the key

The following table lists the configuration settings for Data Management Studio data jobs:

Setting	Description

arch config	This path indicates the location of the macro definitions file. If not set, this value defaults to \etc\macros.cfg (batch jobs and real-time services).
	
	`# Windows Example`
	`arch config = C:\Program Files\DataFlux\Data Management Studio\[version]\etc\macros.cfg`
	
	`# UNIX Example`
	`arch config = /opt/dataflux/aix/[version]/dfpower/etc/macros.cfg`

canada post db	This setting indicates the path to the Canada Post database for Canadian address verification (batch jobs and real-time services).
	
	`# Windows Example`
	`canada post db = C:\Program Files\DataFlux\Data Management Studio\[version]\mgmtrsrc\RefSrc\SERPData`
	
	`# UNIX Example`
	`canada post db = /opt/dataflux/aix/dfpower/[version]/mgmtrsrc/refsrc/serpdata`

checkpoint	Sets the minimum time between log checkpoints, allowing control of how often the log file is updated. Add one of the following to indicate the unit of time: h, min, s (batch jobs and Profile jobs).
	
	`# Windows or UNIX Example`
	`checkpoint = 15min`

Setting	Description

cluster memory	Cluster memory is the amount of memory to use per cluster of match-coded data. Use this setting if you are using clustering nodes in Data Management Studio (batch jobs and real-time services). This setting may affect memory allocation.
	
	Note: This setting must be entered in megabytes, for example, 1 GB should be set to 1024 MB.
	
	`# Windows or UNIX Example`
	`cluster memory = 64MB`

copy qas files	If you plan to run jobs using the Address Verification - QAS node from the command line on UNIX you will need to add the following entry to the app.cfg file:
	
	`copy_qas_files = yes`
	
	If this setting is not defined, jobs run with dmpexec may not complete successfully.
	
	`# Windows or UNIX Example`
	`copy qas files = yes`

datalib path	This is the path to the verify data libraries (batch jobs and real-time services), excluding USPS data. All values containing special characters or spaces must be enclosed in single quotes.
	
	`# Windows Example`
	`datalib path = 'C:\Program Files\DataFlux\DIS\[version]\data'`
	
	`# UNIX Example`
	`datalib path = '/opt/dataflux/hpux/dis/[version]/data'`

dfclient config	Sets the path for the dfIntelliServer® client configuration file, if using dfIntelliServer software. The client can be local or loaded on another machine (Integration Server, dfIntelliServer). This setting is necessary if using distributed nodes in a data job.
	
	`# Windows Example`
	`dfclient config = C:\Program Files\DataFlux\dfIntelliServer\etc\dfclient.cfg`
	
	`# UNIX Example`
	`dfclient config = /opt/dataflux/solaris/dfintelliserver/etc/dfclient.cfg`

enable dpv	To enable Delivery Point Validation (DPV[1]) processing for US Address Verification, set to yes. It is disabled by default (batch jobs and real-time services).
	
	`# Windows or UNIX Example`
	`enable dpv = yes`

[1]Delivery Point Validation (DPV) specifies if the given address is a confirmed delivery point as opposed to being within a valid range of house numbers on the street.

Setting	Description

enable elot	To enable USPS eLOT processing for US Address Verification, set to yes. It is disabled by default (batch jobs and real-time services).
	
	`# Windows or UNIX Example`
	`enable elot = yes`

enable lacs	To enable Locatable Address Conversion System (LACS[2]) processing, set to yes. It is disabled by default (batch jobs and real-time services).
	
	`# Windows or UNIX Example`
	`enable lacs = yes`

enable rdi	Enables Residential Delivery Indicator (RDI[3]) processing for US Address Verification. The default is no (batch jobs and real-time services).
	
	`# Windows or UNIX Example`
	`enable rdi = yes`

fd table memory	Sets the memory size for calculating frequency distribution. If this is not set, a default value of 262,144 bytes will be used on 32-bit systems and 524,288 on 64-bit systems. This memory refers to the number of bytes used per field while processing a table. When processing tables with many fields, this number may be reduced to alleviate memory issues. The larger the value, the more efficient the calculation will be. A minimum value of 4096 bytes exists (8192 on 64 bit systems).
	
	 Note: This is a separate parameter from the frequency distribution memory cache size that is specified on a per job basis.
	
	`# Windows or UNIX Example`
	`fd table memory = 65536`

ftp get command	Used to receive files by FTP. During the DIS installation, the operating system is scanned for the following FTP utilities: NcFTP, Perl LWP Modules, cURL, and Wget. If multiple utilities are found, NcFTP and Perl LWP Modules are given precedence and FTP get/put commands are written to the dfexec.cfg file.
	
	`# Windows or UNIX Example`
	`ftp get command = '"C:\Program Files\NcFTP\ncftpget.exe" -d %L -u %U -p %P %S %T %F'`

ftp put command	Used to send files by FTP. During the DIS installation, the operating system is scanned for the following FTP utilities: NcFTP, Perl LWP Modules, cURL, and Wget. If multiple utilities are found, NcFTP and Perl LWP Modules are given precedence and FTP get/put commands are written to the dfexec.cfg file.
	
	`# Windows or UNIX Example`
	`ftp put command = '"C:\Program Files\NcFTP\ncftpput.exe" -d %L -u %U -p %P %S %T %F'`

[2]US Locatable Address Conversion Service (LACS) is a product/system in a different USPS product line that allows mailers to identify and convert a rural route address to a "city-style" address.
[3]Residential Delivery Indicator (RDI)

Setting	Description

geo db	Sets the path to the database used for geocoding and coding telephone information (batch jobs and real-time services).
	
	`# Windows Example`
	`geo db = C:\Program Files\DataFlux\Data Management Studio\[version]\mgmtrsrc\RefSrc\GeoPhoneData`
	
	`# UNIX Example`
	`geo db = /opt/dataflux/hpux/dfpower/[version]/mgmtrsrc/fresrc/geophonedata`

java classpath	Setting used for the Java Plugin to indicate the location of compiled Java code.
	
	`# Windows Example`
	`java classpath = \usr\java14_64\jre\bin`
	
	`# UNIX Example`
	`java classpath = /usr/java14_64/jre/bin`

java debug	Optional Java Plugin setting that enables debugging in the Java Virtual Machine (JVM™) used by Data Management Studio or Integration Server. The default setting is no.
	
	`# Windows or UNIX Example`
	`java debug = yes`

java debug port	Optional Java Plugin setting that indicates the port number where the JVM listens for debugger connect requests. This can be any free port on the machine.
	
	`# Windows or UNIX Example`
	`java debug port = 23017`

java vm	This Java Plugin setting references the location of the JVM DLL (or shared library on UNIX variants).
	
	`# Windows Example`
	`java vm = [JRE install directory]\bin\server\jvm.dll`
	
	`# UNIX Example`
	`java vm = /[JRE install directory]/bin/server/jvm.dll`

license location	This is the license directory containing the license file (batch jobs, real-time services, and Profile jobs). It was labeled **license dir** in previous versions. All values containing special characters or spaces must be enclosed in single quotes.
	
	Caution: License location is only valid for UNIX. In Windows, set or change the license location using the License Manager. To access the License Manager application click **Start** > **Programs** > **DataFlux Integration Server** > **License Manager**.
	
	`# UNIX Example`
	`license location = '/opt/dataflux/dis/[version]/etc'`

Setting	Description

mail command	This command is used for sending alerts by email (Profile jobs). The command may contain the substitutions %T (To) and %B (Body).
	
	%T will be replaced with the destination email address and %B with the path of a temporary file containing the message body. If %T and %B are left blank, these fields default to what was specified in the job.
	
	The -s mail server parameter specifies the mail server and is not necessary on UNIX systems.
	
	All values containing special characters or spaces must be enclosed in single quotes.
	
	Sendmail is the open source program in UNIX used for sending mail. In Windows, mail is sent by the vbscript mail.vbs.
	
	`# Windows Example`
	`(where mail server is named mailhost)`
	`mail command = 'cscript -nologo "%DFEXEC_HOME%\bin\mail.vbs" -s mailhost "%T" < "%B"'`
	
	`# UNIX Example`
	`mail command = '/usr/lib/sendmail %T < %B'`

odbc ini	Where the odbc.ini file is stored (batch jobs, Profile jobs, and Integration Server).
	
	`# Windows Example`
	`odbc ini = C:\Windows`
	
	`# UNIX Example`
	`odbc ini = /opt/dataflux/solaris`

plugin dir	Where plug-ins are located (batch jobs and real-time services and Profile jobs).
	
	`# Windows Example`
	`plugin dir = C:\Program Files\DataFlux\dis\[version]\bin`
	
	`# UNIX Example`
	`plugin dir = /opt/dataflux/aix/dis/[version]/bin`

qkb root	Location of the Quality Knowledge Base (QKB) files. This location must be set if using steps that depend on algorithms and reference data in the QKB, such as matching or parsing (batch jobs and real-time services and Profile jobs).
	
	 Note: If changes are made to the QKB make sure the server copy is updated as well.
	
	`# Windows Example`
	`qkb root = C:\Program Files\DataFlux\qkb`
	
	`# UNIX Example`
	`qkb root = /opt/dataflux/qkb`

Setting	Description

repository config	Location of the Profile repository config file (Profile jobs and Integration Server). All values containing special characters or spaces must be enclosed in single quotes.
	
	`# Windows Example`
	`repository config = 'C:\Program Files\DataFlux\DIS\[version]\etc\profrepos.cfg'`
	
	`# UNIX Example`
	`repository config = '/opt/dataflux/linux/dis/[version]/etc/profrepos.cfg'`

sort chunk	Allows you to specify the amount of memory to use while performing sorting operations. The amount may be given in KB or MB, but not GB (batch jobs and real-time services).
	
	`# Windows or UNIX Example`
	`sort chunk = 128MB`

usps db	This is the path to the USPS database required for US address verification (batch jobs and real-time services).
	
	`# Windows Example`
	`usps db = C:\Program Files\DataFlux\verify\uspsdata`
	
	`# UNIX Example`
	`usps db = /opt/dataflux/aix/verify/uspsdata`

verify cache	Indicates an approximated percentage (0 - 100) of the USPS reference data set that will be cached in memory prior to an address verification procedure (batch jobs and real-time services). This setting can affect memory allocation.
	
	`# Windows or UNIX Example`
	`verify cache = 30`

verify preload	Allows you to specify a list of states whose address data will be preloaded. Preloading increases memory usage, but significantly decreases the time required to verify addresses in a state (batch jobs and real-time services).
	
	`# Windows or UNIX Examples`
	`verify preload = NY TX CA FL  or verify preload = ALL`

world address db	Sets the path where AddressDoctor data is stored.
	
	`# Windows Example`
	`world address db= 'C:\world_data\'`
	
	`# UNIX Example`
	`world address db= '/opt/dataflux/linux/worlddata'`

world address license	The license key provided by DataFlux used to unlock AddressDoctor country data. The value must be enclosed in single quotes (batch jobs and real-time services).
	
	`# Windows or UNIX Example`
	`world address license = 'abcdefghijklmnop123456789'`

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. TKTS DSNs, however, are managed by the DAC, and TKTS connections are stored in a TKTS DSN directory.

Both DataFlux Data Management Studio and the DataFlux Data Management Server can use the DAC. The default DAC directives for Data Management Studio are specified in its app.cfg file. You can also specify DAC directives in Studio's macros.cfg file. These settings apply when you use Studio to access data via a TKTS connection without using a DataFlux Federation Server. For information about Studio configuration files, see Data Management Studio Configuration Files.

DAC directives can also be specified for a DataFlux Data Management Server if one is installed at your site. For more information, see the *DataFlux Data Management Server Administrator's Guide*.

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

Setting	Description
Command file execution	Specifies a text file with SQL commands (one per line). These commands will run in turn, on any new connection that is made. For example, they can be used to set session settings. Note that this is only implemented for the ODBC driver. The DAC/SAVEDCONNSYSTEM configuration value may specify the path to the saved connections. The DAC checks for files with the same filename as the DSN and a .sql extension.
DFTK log file	Specifies the log file that interacts with the DFTKSRV layer and is only useful for debugging issues specific to dftksrv. This setting is only applicable if you are running DFTK out of process. The DAC/DFTKLOGFILE configuration value specifies the path to the DFTK log file.
Disable CEDA	Specifies whether to disable CEDA. This setting is only applicable to tkts connections. The DAC/DFTKDISABLECEDA configuration value, which should specify any non-null value, for example, yes.

Setting	Description
Run DFTK out of process	Specifies whether to run TKTS out of process, allowing you to perform troubleshooting. The DAC/DFTK_PROCESS configuration value should specify any non-null value, for example, yes.
System saved connection	Specifies where to find system saved connections. The DAC/SAVEDCONNSYSTEM configuration value may specify the path. If it does not, the DAC checks the following values and locations, based on your operating system: Windows - The \etc\dsn subdirectory, which is in the installation directory. UNIX - The \etc\dsn subdirectory, which is in the installation directory.
TK Path	Specifies where TK files are located. This setting is only applicable if you are running Data Factory Took Kit (DFTK) out of process. The dftksrv path and core directory should be specified. The DAC/DFTK_PROCESS_PATH configuration value may specify the TK path. If it does not, the DAC checks the following locations, based on your operating system: Windows - $DFEXEC_HOME\bin;$DFEXEC_HOME\bin\core\sasext UNIX - $DFEXEC_HOME/lib/tkts
TKTS DSN directory	Specifies the path where TKTS DSNs are stored in XML files. The DAC/DSN configuration value should specify the directory. If it does not, the DAC checks the following locations, based on your operating system: Windows - The \etc\dsn subdirectory, which is in the installation directory. UNIX - The \etc\dsn subdirectory, which is in the installation directory.
TKTS log file	Specifies the log file that is produced by the TKTS layer and is useful for debugging tkts issues. The DAC/TKTSLOGFILE configuration value specifies the path to the TKTS log file.

Setting	Description
TKTS startup sleep	Specifies how much time in seconds to delay between the start of the dfktsrv program and the booting of TK. This setting is only applicable if you are running DFTK out of process. The DAC checks the following values and locations, based on your operating system: Windows - The registry for a tktssleep value. UNIX - This setting is not supported.
User saved connection	Specifies where to find user-saved connections. The DAC/SAVEDCONNUSER configuration value may specify the path. If it does not, the DAC checks the following values and locations, based on your operating system: Windows - The application settings directory for the user, which is usually 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. UNIX - The $HOME/.dfpower/dsn directory.

Note: Environment variables are specified as $*variable_name*. Typically, Data Management Studio will set environment variables to appropriate locations. For example, $DFEXEC_HOME is set to the Data Management Studio home directory.

Add-On Products

- Installing a Quality Knowledge Base
- Installing Data Packs
- Installing Supplemental Language Support

Installing a Quality Knowledge Base

The Quality Knowledge Base (QKB) is a collection of files that store data and logic that define data management operations. DataFlux® software product reference the QKB when performing data management operations on your data.

Microsoft Windows

1. Insert the Quality Knowledge Base CD-ROM into the CD-ROM drive.
2. From the Microsoft® Windows® taskbar, click **Start** > **Run**.
3. Type [your_drive]:\QKB_[version].exe, where [your_drive] is replaced by the letter corresponding to your CD-ROM drive and where [version] is replaced by the QKB version you are installing (for example, QKB_CI_2009A).
4. Follow the instructions on the installation setup Wizard.
5. After you install the QKB, restart Data Management Studio.

Note: If you downloaded the QKB installation file from the DataFlux FTP site, then double-click on the name of the installation file in Windows Explorer.

For more information about the DataFlux Quality Knowledge Base products, refer to the DataFlux Web site or refer to the QKB online documentation.

Installing Data Packs

If you are using external data, install USPS, Software Evaluation and Recognition Program (SERP), Geocode/Phone, QuickAddress Software (QAS), World, or other enrichment data. Make a note of the path to each data source. You will need this information to update the dfwproc.cfg configuration file.

Downloading and Installing Data Packs

If your Data Management Studio installation includes a Verify license, you need to install the proper USPS, Canada Post, and Geocode databases to do address verification. If you are licensed to use QAS, you must acquire the postal reference databases directly from QAS for the countries they support. For more information, contact your DataFlux® representative.

Data Packs for data enrichment are available for download on the MyDataFlux Portal at http://www.dataflux.com/MyDataFlux-Portal. To download data packs, follow these steps:

1. Obtain a user name and password from your DataFlux representative.
2. Log in to the MyDataFlux Portal.

   Note: You may also retrieve the data pack installation files through FTP. Please contact DataFlux Technical Support for more information regarding downloading through FTP.

3. Click **Downloads** > **Data Updates**.
4. Select the installation file corresponding to your data pack and operating system to download.

Close all other applications and follow the procedure that is appropriate for your operating system.

Windows

Browse to and double-click the installation file to begin the installation wizard. If you are installing QAS data, you must enter a license key. When the wizard prompts you for a license key, enter your key for the locale you are installing.

UNIX

Installation notes accompany the download for each of the UNIX® data packs from DataFlux. For Platon and USPS data, check with the vendor for more information.

Notes:

1. Be sure to select a location to which you have write access and which has at least 430 MB of available space.
2. Download links are also available from the MyDataFlux Portal link at http://www.dataflux.com/MyDataFlux-Portal.

Configuring Enrichment Data

If you are using external data, install USPS, SERP, Geocode/Phone, QAS, World, or other enrichment data. You will need to specify the path to each data source in your configuration file.

Configuring USPS

Windows

Download **Windows Verify Data Setup** from the MyDataFlux Portal and run the installation file.

UNIX

Download **UNIX Verify Data Setup** from the MyDataFlux Portal and install the file on your Data Management Studio machine.

Setting	Description

usps db	This is the path to the USPS database, which is required for US address verification (batch jobs and real-time services).
	
	`# Windows Example`
	`usps db = C:\Program Files\DataFlux\verify\uspsdata`
	`# UNIX Example`
	`usps db = /opt/dataflux/verify/uspsdata`

Configuring DPV

Windows

Download **Windows Verify DPV Data Setup** from the MyDataFlux Portal, and run the installation file. Enable DPV by changing the **enable dpv** setting in the dfwproc.cfg file.

UNIX

Download **UNIX Verify DPV Data Setup**, under USPS in the Data Updates section of the MyDataFlux Portal. Enable DPV by changing the **enable dpv** setting in the dfwproc.cfg file.

Setting	Description

enable dpv	To enable Delivery Point Validation (DPV) processing (for US Address Verification), set to *yes*. It is disabled by default (batch jobs and real-time services).
	
	`# Windows or UNIX Example`
	`enable dpv = yes`

Configuring USPS eLOT

Windows

Download **Windows Verify eLOT Data Setup** from the MyDataFlux Portal, and run the installation file. Enable eLOT by changing the **enable elot** setting in the dfwproc.cfg file.

UNIX

Download **UNIX Verify eLOT Data Setup**, under in the Data Updates section of the MyDataFlux Portal. Enable eLOT by changing the **enable elot** setting in the dfwproc.cfg file.

Setting	Description

enable elot	To enable USPS eLOT processing (for US Address Verification), set to *yes*. It is disabled by default (batch jobs and real-time services).
	
	`# Windows or UNIX Example`
	`enable elot = yes`

Configuring Canada Post (SERP)

Windows

Download the Microsoft Windows **SERP** data update from the MyDataFlux Portal and install the file on your Data Management Studio machine.

UNIX

Download the **SERP** data update that corresponds to your operating system from the MyDataFlux Portal and install the file on your Data Management Studio machine.

Setting	Description

canada post db	This setting indicates the path to the Canada Post database for Canadian address verification (batch jobs and real-time services).
	
	`# Windows Example`
	`canada post db = C:\Program Files\DataFlux\Data Management Studio\`*`version`*`\mgmtrsrc\RefSrc\SERPData`
	`# UNIX Example`
	`canada post db = /opt/dataflux/aix/dfpower/`*`version`*`/mgmtrsrc/refsrc/serpdata`

Configuring Geocode / Phone

Windows

Download the Windows **Geocode Data Pack** from the MyDataFlux Portal and install the file on your Data Management Studio machine.

UNIX

Download the UNIX **Geocode Data Pack** from the MyDataFlux Portal and install the file on your Data Management Studio machine.

Setting	Description

geo db	This sets the path to the database for geocoding and coding telephone information (batch jobs and real-time services).
	
	`# Windows Example`
	`geo db = C:\Program Files\DataFlux\Data Management Studio\`*`version`*`\mgmtrsrc\RefSrc\GeoPhoneData`
	`# UNIX Example`
	`geo db =`
	`/opt/dataflux/hpux/dfpower/`*`version`*`/mgmtrsrc/fresrc/geophonedata`

Configuring QAS Data

Windows

Contact QAS to download the latest data files for the countries you are interested in. Once you have downloaded the data sets, run the installation file and follow the instructions provided by the installation wizard.

UNIX

Run the installation file on a Windows machine to get the .dts, .tpx, and .zls files, then transfer all of these to your UNIX environment.

> **Note:** If you plan to run jobs using the Address Verification - QAS node from the command line on UNIX you will need to add the following entry to the app.cfg file:
>
> ```
> copy_qas_files = yes
> ```
>
> If this setting is not defined, jobs run with dmpexec may not complete successfully.

Configure the following QAS files located in the /etc subdirectory of your Data Management Studio directory:

- In the qalicn.ini file, copy your license key for the specific country. Each license key must be entered on a separate line.
- In the qaworld.ini file, you must specify the following information:

    1. Set the value of the **CountryBase** parameter equal to one or more country prefixes for the countries you have installed. For example, to search using Australian mappings, add the following line to your qaworld.ini file:

        ```
        CountryBase=AUS
        ```

        Additional country prefixes can be added to the CountryBase parameter. Separate each prefix by a space. For a complete list of supported countries, see the International Address Data lists at the QAS web site.

    2. Set the value of the **InputLineCount** parameter. Add the country prefix to the parameter name and set the count equal to the number of lines your input addresses contain. For example, to define four lines for Australia:

        ```
        AUSInputLineCount=4
        ```

    3. Set the value of the **AddressLineCount** parameter. Add the country prefix to the parameter name and set the count equal to the total number of lines. Then, specify which address element will appear on which line in the input address by setting the value of the **AddressLine** parameter equal to a comma-separated list of element codes. For example:

        ```
        AUSAddressLineCount=4
        AUSAddressLine1=W60
        AUSAddressLine2=W60
        AUSAddressLine3=W60
        AUSAddressLine4=W60,L21
        ```

    For more information on address elements and configuring the qaworld.ini file, see *QuickAddress Batch API Guide* and the country-specific data guides.

- In the qawserve.ini file, you must specify the following information for each parameter. If more than one country prefix is added to the parameter, each subsequent country prefix should be typed on a new line and preceded by a + (plus sign). For a complete list of supported countries, see the International Address Data lists at the QAS web site.

    1. Set the value of the **DataMappings** parameter equal to the country prefix, country name, and country prefix. Separate each value by a comma. For example:

        ```
        DataMappings=AUS,Australia,AUS
        ```

    2. Set the value of the **InstalledData** parameter equal to the country prefix and installation path. Separate each value by a comma. For example:

        ```
        InstalledData=AUS,C:\Program Files\QAS\Aus\
        ```

For more information on configuring the qawserve.ini file, see *QuickAddress Batch API Guide* and the country-specific data guides.

Note: If you have existing jobs that include the Address Verification (QAS) node, your jobs will not work. You must reconfigure your existing jobs to work with the new QAS 6.x engine.

Configuring AddressDoctor Data

Windows and UNIX

If you are using AddressDoctor data for address verification, download the address files for the countries you are interested in from the MyDataFlux Portal at http://www.dataflux.com/MyDataFlux-Portal. You will also need the addressformat.cfg file included with the data files. The addressformat.cfg file must be installed in the directory where the address data files reside.

Change the world address license and world address database settings in the dfwproc.cfg file:

Setting	Description

world address license	This is the license key provided by DataFlux that is used to unlock the AddressDoctor country data. The value must be enclosed in single quotes (batch jobs and real-time services).
	
	`# Example (same for Windows and Unix)`
	`world address license = 'abcdefghijklmnop123456789'`

world address db	This sets the path to where the AddressDoctor data is stored.
	
	`# Windows Example`
	`world address db= 'C:\world_data\'`
	`# UNIX Example`
	`world address db= '/opt/dataflux/linux/worlddata'`

Configuring LACS and RDI Data

Windows and UNIX

Residential Delivery Indicator (RDI) and Locatable Address Conversion System (LACS) are provided by the United States Postal Service®. If you are using these products, simply download the data with your USPS data, and set the applicable settings in the dfwproc.cfg file:

Setting	Description

enable lacs	To enable LACS processing, set to *yes*. It is disabled by default (batch jobs and real-time services).
	
	`# Windows or UNIX Example`
	`enable lacs = yes`

enable rdi	This option enables or disables RDI processing (for US Address Verification). By default, it is set to *no* (batch jobs and real-time services).
	
	`# Windows or UNIX Example`
	`enable rdi = yes`

Installing Supplemental Language Support

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

1. Click **Start** > **Settings** > **Control Panel**.
2. Double-click **Regional and Language Options**.
3. In the Regional and Language Options dialog, select the **Languages** tab.
4. Under **Supplemental Language Support**, select the check boxes marked, **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.

Technical Support

- Frequently Asked Questions

Frequently Asked Questions (FAQ)

The following questions and answers are designed to assist you when working with Data Management Studio. If you do not find your answer, please contact DataFlux Technical Support.

- Data Connections
- General
- Jobs, Profiles, Data Explorations
- Repositories

Data Connections

Are there any special considerations for ODBC drivers using the wire protocol?

DataDirect provides a number of wire protocol ODBC drivers that communicate directly with a database server, without having to communicate through a client library. If these drivers are available at your site, they are available from the **Drivers** tab of the ODBC Data Source Administrator dialog.

If you use a wire protocol driver to create an ODBC connection, the following special considerations apply:

- Verify that the following value is set in the registry. It is needed in order for Unicode characters to display properly. For each wire driver, go into the registry under HKEY_LOCAL_MACHINE\SOFTWARE\ODBC\ODBC.INI\<DSN NAME>, create a string value called ColumnSizeAsCharacter, and set it to 1.
- After you set up an ODBC connection with a wire protocol driver, verify that the appropriate options have been set on the **Advanced** tab of the connection properties dialog. The Advanced tab can be displayed as follows: from the ODBC Data Source Administrator dialog, click the **System DSN** tab. Select the ODBC connection with a wire protocol driver, then click the **Configure** button. The properties dialog for the connection displays. Click the **Advanced** tab and review the options on this tab. To turn on an option, click the checkbox beside the option.

How Can I Read an XML File in a Data Job?

Due to the limitations of the ODBC 32-bit XML driver, we recommend that you use the XML Input node to read an XML file in a data job.

How Can I Read an XML File In a Profile?

We recommend that you extract the data from the XML file to a text file or to a database table, then profile the text file or table. To extract the data, create a data job in which the XML Input node is used to read the XML file, then use other nodes to output a text file or a database table.

General

Why can't I save global options for DataFlux Data Management Studio under Microsoft Vista or Windows Server 2008?

DataFlux Data Management Studio saves global options to a user.config file that is hidden by default under Microsoft Vista and Windows Server 2008. You must un-hide this file in order to save global options. The physical path to the file is as follows:

```
C:\Documents and Settings\<UserID>\Local Settings\Application
Data\DataFlux\ProDesigner.vshost.exe_Url_<some_hash_code>
```

Why doesn't my screen repaint when I'm prompted to log into a table with ODBC?

If you access a table with ODBC, you might be prompted to log in to the database. If your screen does not repaint properly, try setting the **Show window contents while dragging option** for Microsoft Windows. Consult the documentation for your version of Windows for details about setting this option.

Jobs, Profiles, Data Explorations

What is the maximum length for character variables (such as column names) in DataFlux Data Management Studio?

In data jobs, Data Input nodes and Data Output nodes support very long character fields for SAS data. They successfully work with 32K (32767 bytes) fields, which is the maximum length for character fields in SAS data sets. QKB-related nodes only process the first 256 characters and ignore the rest. Expression node string functions should work, including the mid() and len() functions. The 256 character limitation applies to regular expressions or QKB-related-functions.

In profiles, report metrics such as Data Type, Data Length, Unique Count, Frequency Distribution are correct for strings up to 32K in length. Pattern Frequency Distribution only uses the first 254 characters instead of 256.

How can I specify Quality Knowledge Base options for profiles and data explorations?

Configuration options for the QKB are set in the Quality Knowledge Base Engine section of the app.cfg file. For example, the QKB\PATH option enables you to specify the path to the QKB. The QKB/ON_DEMAND option determines whether the QKB is loaded on demand or all at once. By default, the option is set to YES. The QKB/ALLOW_INCOMPAT specifies how newer QKB definitions are handled. By default, this option is set to NO. You might want to change this option to YES if a profile or data exploration fails due to an incompatible (newer) QKB definition. The QKB\COMPATVER option enables you to specify the QKB version. Finally, the QKB\SURFACEALL determines whether all parse definitions are surfaced.

You can use Data Management Studio to change the QKB/ALLOW_INCOMPAT option. Click **Tools** in the main menu and select **Options** to display the Data Management Studio Options dialog. Click the **General** section of the dialog and update the checkbox for **Allow use of incompatible Quality Knowledge Base definitions**.

To change other QKB options, you must edit the app.cfg file. See the "Configuration" section of the *Data Management Studio Installation and Configuration Guide*.

How are SAS Data Types Converted When DataFlux Software Reads or Writes SAS Data?

DataFlux software and SAS data sets support different data types. Accordingly, automatic data-type conversions will take place when Data Management Studio software reads or writes SAS data sets. Also, nulls and missing values will be converted to other values. These changes can impact features that depend on particular data types. For example, when a profile reads a SAS data set, SAS fields with a format that applies to **datetime** values will be reported as **datetime**. SAS fields with a format that applies to time values will be reported as a time and SAS fields with a format that applies to date values will be reported as a date. As a result, the profile will not calculate some metrics such as Blank Count or Maximum Length for those fields.

The following data-type conversions are made automatically when DataFlux software, such as a data job or a profile, reads SAS data.

- For jobs: SAS numeric columns with a format that applies to date, time or datetime values will be converted to a DataFlux field of type date.
- For profiles: SAS fields with a format that applies to datetime values will be reported as datetime. SAS fields with a format that applies to time values will be reported as a time and SAS fields with a format that applies to date values will be reported as a date. Other SAS numeric columns will be converted to a DataFlux field of type real.
- SAS character columns will be converted to a DataFlux field of type string with the same length as the SAS character column.

Nulls and missing values will be converted to other values, as follows.

- SAS missing values will be converted to DataFlux null values. SAS special numeric missing values, whether they are specified with the MISSING statement in a SAS DATA step or with a dot followed by a letter or underscore, are also converted to null values.
- DataFlux null values will be converted to SAS missing values.
- A DataFlux field of type string that contains a blank will be converted to a SAS character field containing a blank. This blank will be interpreted by SAS as a missing value.

The following data-type conversions are made automatically when a Data Management Studio data job writes SAS data.

DataFlux Input Data Type	SAS Output Data Type		
	Type	Length	Format
boolean	num	8	
date	num	8	datetime19.2
integer	num	8	
real	num	8	
string	char	255	

Why are my fields not propagating automatically even though I set the Tools > Options > Job > Output Fields to "All"?

Some data job nodes require you to manually add all of the output fields before they can be propagated. These nodes include, Data Sorting, Data Union, and Cluster analysis, among others. Manually add all available fields by selecting the double-right arrow in the Output fields section of the node's **Properties** window. Once you manually add the fields, the fields will propagate correctly.

Repositories

What does "The repository is newer than this client" mean?

If you get a message that says something like, "The version of repository "<ReposName>" is newer than this client, then someone at your site has a newer version of Data Management Studio than you do and has upgraded the repository. Contact your site administrator about upgrading your Data Management Studio software.

Glossary

A

Access Control Entry

An Access Control Entry (ACE) is an entry of user information made to the Access Control Lists (ACLs) which is used to secure access to individual DataFlux Integration Server (DIS) objects.

Access Control Lists

Access Control Lists (ACLs) are used to secure access to individual DataFlux Integration Server (DIS) objects.

address verification

Address verification (validation) is the process of comparing a physical address to a reference database of known physical addresses so the original address can be standardized and corrected according to postal authority standards.

AIC

Analyze, Improve, Control (AIC) - DataFlux enables organizations to analyze, improve, and control their data from a single data quality integration platform. DataFlux tools and approaches can help you build a comprehensive set of business rules that can create a unified view of your enterprise data and enhance the effectiveness of CDI, CRM, ERP, legacy data migration, or compliance initiatives.

AMAS

Address Matching Approval System (AMAS) is the program the Australia Post administers to certify address verification software.

API

Application Programming Interface (API) is a set of software protocols, routines, and/or tools used when building software applications.

APO

Army/Air Force post office (APO) is an indication for the USPS.

Architect Job Templates

dfPower Studio can be used to modify and build work flows called jobs. These jobs can be delivered as templates that can be fleshed out by consultants or other IT professionals. Many job templates will be designed and delivered with the solution to accommodate such things as address verification, merging, assigning IDs, standardizing data, and so on.

ASCII

ASCII (American Standard Code for Information Interchange) is a character set based on the English alphabet

B

basic category

A basic category is a category that represents a single word. Basic categories are the basic building blocks of Grammar rules. Every basic category in a Grammar corresponds to a category in an ordered word list. For this reason, you should design Grammar rules in parallel with word-analysis logic.

batch processing

The application of data management routines to data source records in what are often very large groups, usually in processes that require no manual user intervention. Contrast with real-time processing.

business functions

These are expressions which are written in a generic manner so they can be reused from multiple rules or applications.

business rule

A conditional statement that tells a system running a business process how to react to a particular situation.

C

case definition

A set of logic used to accurately change the case of an input value, accounting for unique values that need to be case sensitive, such as abbreviations and business names.

CASS

Coding Accuracy Support System (CASS) is the program the United States Postal Service (USPS) administers to certify address verification software.

CBSA

Census Bureau Statistical Areas (CBSA)

CEDA

Cross-Environment Data Access (CEDA)

census string

The census string is a US Census Bureau designation for the boundary area in which the centroid exists. The census string contains state, county, and other census-type information.

centroid

A centroid is the approximate mathematical center of the ZIP or ZIP+4 boundary.

checks

These are built-in checks (expressions) that provide a template to the user to build common standard expressions.

chop table

A proprietary file type used by DataFlux as a lex table to separate characters in a subject value into more usable segments.

CMRA

US Commercial Mail Receiving Agency (CMRA)

CMSA

Consolidated Metropolitan Statistical Areas (CMSA)

Comments

Comments are text within a code segment that are not executed. Comments can be either C-style (starts with /* and ends with */) or C++ style (starts with // and continues to the end of a line).

Core Fields

Default logic to handle data such as name and address, which inform the identity management process.

CPC

Canadian Post Certification (CPC) is the SERP program administered by the Canadian Post. This is similar to the CASS certification administered by the USPS.

CRM

Customer Relationship Management (CRM)

custom metrics

Custom metrics may be used when the standard metrics do not contain the rules you need to accomplish the desired results.

D

dashboard

The dashboard is a Web-based view of the task grid and graphs in the Monitor Viewer.

data profiling

A discovery process that uncovers potential problem areas in large amounts of structured data.

data type

Not used in the sense of a database data type ("varchar" for instance) but used to describe sets of data values that follow certain rules and conventions. "Name" and "Address" are two examples of data types.

database

A collection of tables containing data that can be accessed easily by a computer system.

definition

An algorithm available to a DataFlux application.

derived category

A derived category is a category composed of one or more other categories. The makeup of a derived category is described using rules.

dfIntelliServer

dfIntelliServer provides a real-time or transactional mechanism for communicating with the MCRD through the Architect API. dfIntelliServer has several client libraries (including a Web services client) that can be called from a number of different applications in many different computing environments. dfIntelliServer allows one at a time queries and modifications to the MCRD. dfIntelliServer allows organizations to access Architect jobs through an API that can accept one group of data elements at a time rather than a complete table. This functionality takes advantage of the power of encapsulation of discreet chunks of work in Architect, so a programmer need only make one call to the client API to perform a related set of activities.

DPV

Delivery Point Validation (DPV) specifies if the given address is a confirmed delivery point as opposed to being within a valid range of house numbers on the street.

DSN

Data Source Name (DSN)

E

EEL

Expression Engine Language (EEL)

ERP

Enterprise Resource Planning (ERP)

ETL

Extraction, Transformation, and Loading

event

An event represents an action which should be taken when a rule fails. Actions can include sending email messages, storing the offending row in the repository, or executing an external process.

Expression

This is the DataFlux syntax used in the Business Rule Manager to build business rules.

F

field

Also known as a "variable" or a "column," a single piece of data in a database table. Database tables can have many fields. The user defines the fields. Each field has a unique identifier in the repository. From a data monitoring standpoint, the fields are not tied to any specific database or table but are bound at the time of execution to the current data set or row.

field set

A field set is a collection of fields that belong together. These usually represent a table of data and are used to aid in building rules and viewing results.

FIPS

Federal Information Processing Standards (FIPS) - A 5-digit number assigned to each county in the U.S. by the Census Bureau. The first 2 digits are the state code, and the last 3 digits are the county number.

FPO

Fleet post office (FPO) indication for USPS used for military personnel.

G

gender analysis

An algorithm that can determine the gender of persons by their names.

gender definition

A set of logic used to determine the probable gender of a name or identity-type input string.

grammar

A proprietary file type used to store hierarchical patterns pertinent to a specific subject area.

group rule

A group rule evaluates and applies all rules to groups of data (for example, data grouped by state and the rules evaluated for each state).

H

historical metrics

A historical metric is available when a business rule is run a second time under the same report name. You can view and compare the last two reports.

I

identification analysis

An algorithm that can determine from a known set of options what type of data is represented by a particular subject value.

identification definition

A set of logic used to identify an input string as a member of a redefined or user-defined value group or category.

inputs

Input fields are the fields where you apply the checks specified in the Rule Manager.This list includes all the fields you have defined in the Business Rule Manager, including the Output fields from custom metrics and any grouped by field.

J

job

The saved configuration settings for a particular task in a dfPower Studio application. You can run jobs interactively or combine them with other jobs and schedule the set of jobs to run on a particular date or time.

L

LACS

US Locatable Address Conversion Service (LACS) is a product/system in a different USPS product line that allows mailers to identify and convert a rural route address to a "city-style" address.

locale

The country of origin based on an address or country code.

locale guessing

A process that attempts to identify the country of origin of a particular piece of data based on an address, country code, or other field.

M

match

The process of identifying data strings that can be different representations of the same semantic information. For example, the strings Mr. Bob Brauer, Robert J., and Brauer can be considered to match each other.

match cluster

A set of records grouped together based on some commonality. Cluster IDs are numeric values used to refer to these clusters. You can append cluster IDs to records in a database to document matches.

match codes

The end result of passing data through a match definition. A normalized, encrypted string that represents portions of a data string that are considered to be significant with regard to the semantic identity of the data. Two data strings are said to "match" if the same match code is generated for each.

match definition

A set of logic used to generate a match code for a data string of a specific data type.

match value

A string representing the value of a single token after match processing.

MCD

Minor Civil Division (MCD)

MDM

Master Data Management (MDM) focuses on master data shared by several different systems and groups.

merge

The process of joining records and eliminating duplicate records from a table based on user-specified conditions and rules.

metadata

Information that describes the properties of data , for example when was last accessed or the size of the data value.

micropolitan

This term is used in US Census data and refers to a population area including a city with 10,000 to 50,000 residents and surrounding areas.

MSA

Metropolitan Statistical Areas (MSA) - The MSA code assigned by the Office of Management and Budget. Use this code as an index key in the MSA file.

N

namespace

A namespace is a unique container created to hold a logical grouping of identifiers.

O

Object

An object is anything that can be stored in the dfPower Studio Navigator and accessed by the dfPower Studio applications.

objects

Objects are individual jobs and services.

ODBC

Open Database Connectivity (ODBC) - an open standard application programming interface (API) for accessing databases.

OFAC

Office of Foreign Assets Control (OFAC) - Federal regulations related to the Patriot Act.

OLAP

Online Analytical Processing (OLAP)

organization

A company, university, or other type of institution. For example: IBM Corporation, University of Connecticut, or St. Joseph's Hospital

outputs

The output field is the field(s) used to apply the rule in the custom metric. Set your output field to serve as the field where the results from your custom metric are collected.

P

parse

The process of dividing a data string into a set of token values. For example: Mr. Bob Brauer, Mr. = Prefix, Bob = Given, Brauer = Family

parse definition

A name for a context-specific parsing algorithm. A parse definition determines the names and contents of the sub-strings that will hold the results of a parse operation.

pattern analysis definition

A regular expression library that forms the basis of a pattern recognition algorithm.

phonetics

An algorithm applied to a data string to reduce it to a value that will match other data strings with similar pronunciations.

PMB

A private mailbox (PMB) is categorized as a mailbox located at a mail center other than the post office or home.

PMSA

Principal Metropolitan Statistical Areas (PMSA)

Primary Key

Primary key is a unique identifier assigned to a database field. Social Security Numbers or a ISBNs are examples of possible primary keys.

Q

QAS

QuickAddress Software (QAS)

QKB

The Quality Knowledge Base (QKB) is a collection of files and configuration settings that contain all DataFlux data management algorithms. The QKB is directly editable using dfPower Studio.

Quality Knowledge Base Locales

The Quality Knowledge Base (QKB) locales contain the files, file relationships, and metadata needed to correctly parse, match, standardize, and otherwise process data.

R

RDBMS

Relational Database Management System (RDBMS) allows you to access data in a database in unique ways, such as adding tables and records, and joining tables.

RDI

Residential Delivery Indicator (RDI)

real-time processing

Processing a record or data one piece at a time as it enters a computer system, for financial transactions, for example. Contrast with batch processing.

record

Also called a "row" or "observation," one complete set of fields in a database table.

regular expression

A mini-language composed of symbols and operators that enables you to express how a computer application should search for a specified pattern in text. A pattern may then be replaced with another pattern, also described using the regular expression language.

repository

A dfPower repository is a hierarchical data storage mechanism.

row rule

A row rule evaluates every row of data passed into the Monitoring node.

RP

Software Evaluation and Recognition Program is a program the Canada Post administers to certify address verification software.

rule

A single rule can be either a row level rule or a data set level rule. A row level rule is applied to each row which enters the system while a data set level rule is applied to an entire data set or a portion of a data set.

rule set

A rule set is a set of one or more rules which are applied together as a group. Use a rule set when you find you are using a few rules together frequently.

S

SDK

Software Development Kit (SDK)

sensitivity

Regarding matching procedures, sensitivity refers to the relative tightness or looseness of the expected match results. A higher sensitivity indicates you want the values in your match results to be very similar to each other. A lower sensitivity setting indicates that you would like the match results to be "fuzzier" in nature.

SERP

The Software Evaluation and Recognition Program (SERP) is a program the Canadian Post administers to certify address verification software.

Service Oriented Architecture

Service Oriented Architecture (SOA) - All of the interaction with the master customer reference database is through a service-oriented architecture that enables any system to talk to the customer database and request or update information.

set rule

A set rule evaluates and applies rules to all of the input data completely (for example, it will evaluate all 1000 rows of data as a set).

SQL

Structured Query Language (SQL) is a language used to request information from database systems.

standard metrics

Standard metrics are pre-defined rules (expressions) set in dfPower. Most of the time, this is enough to achieve the results for your job.

standardization definition

A set of logic used to standardize a string.

standardization scheme

A collection of transformation rules that typically apply to one subject area, like company name standardization or province code standardization.

standardize

The process of transforming a data string so each of the string's token values conforms to a preferred standard representation: IBM Corporation = IBM CORP; Mister Bob Brauer, Junior = MR BOB BRAUER JR.

Statement of Accuracy

Statement of Accuracy (SoA) is the form used for Canadian Post Certification (CPC) standards.

T

table

A table is a collection of records in a database.

tasks

Tasks contain the rules and the events that go with your individual rule. Tasks associate alert events with a rule that are triggered after a rule fails.

token

Used by DataFlux to designate the output strings of a parse process. The output string of a parse process. A word or atomic group of words with semantic meaning in a data string. A set of expected tokens is defined for each data type.

U

Unicode

An industry standard used to allow text and symbols from languages around the world.

unified

This is the version of the repository you are using. The term "unified" means the repository contains data for dfPower Profile reports, Business Rules, and Data Monitoring results.

URI

Uniform Resource Identifier (URI) is a string of characters identifying a resource or file path.

USPS

United States Postal Service (USPS) provides postal services in the United States. The USPS offers address verification and standardization tools.

V

vocabulary

A proprietary file type used for categorizing data look-ups pertinent to a specific subject area.