

# DataFlux Data Management Studio Installation and Configuration Guide



YOUR DATA.  
YOUR BUSINESS.  
ONE SOLUTION.



This page is intentionally blank



# DataFlux Data Management Studio Installation and Configuration Guide

---

Version 2.2

November 17, 2011

This page is intentionally blank

# Contact DataFlux

## DataFlux Corporate Headquarters

Toll Free: (877) 846-3589  
Tel: (919) 447-3000  
Fax: (919) 447-3100  
940 NW Cary Parkway, Suite 201  
Cary, NC 27513  
USA

## DataFlux West

Tel: (818) 906-7638  
Fax: (818) 907-6012  
15300 Ventura Boulevard, Suite 523  
Sherman Oaks, CA 91403  
USA

## DataFlux United Kingdom

Tel: +44 (0) 20 3176 0025  
Fax: +44 (0) 20 3411 8382  
Enterprise House  
1-2 Hatfields  
London  
SE1 9PG  
United Kingdom

## DataFlux France

Tel: +33 (0) 4 72 91 31 42  
Immeuble Danica B  
21, avenue Georges Pompidou  
69003 Lyon  
France

## DataFlux Germany

Tel: +49 (0) 69 66 55 42 04  
In der Neckarhelle 162  
69118 Heidelberg  
Germany

## DataFlux Australia

Tel: +61 2 9428 0553  
300 Burns Bay Road  
Lane Cove, NSW 2066  
Australia

## Technical Support

Phone: 1-919-531-9000  
Email: [techsupport@dataflux.com](mailto:techsupport@dataflux.com)  
Web: <http://dataflux.com/MyDataFlux-Portal.aspx>

## Documentation Support

Email: [docs@dataflux.com](mailto:docs@dataflux.com)

# Legal Information

Copyright © 1997 - 2011 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.

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

[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 3.1

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](mailto: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, and NCOA<sup>Link</sup>®. © United States Postal Service 2005. ZIP Code® and ZIP+4® are registered trademarks of the U.S. Postal Service.

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

## VMware

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

<b>Introduction</b> .....	<b>1</b>
Accessibility.....	1
Conventions Used in this Document .....	1
Reference Publications .....	2
<b>Installing Data Management Studio</b> .....	<b>3</b>
Installing Data Management Studio.....	3
System Requirements .....	7
Data Storage.....	7
Repository Storage.....	8
Supported Platforms.....	10
<b>Configuring Data Management Studio</b> .....	<b>13</b>
Configuration Files.....	13
Configuration Options .....	14
Data Access Component Directives .....	21
<b>Logging Options</b> .....	<b>29</b>
<b>Installing Add-On Products</b> .....	<b>31</b>
Installing a Quality Knowledge Base.....	31
Installing Data Packs .....	31
Installing Supplemental Language Support .....	32
Installing the Address Update Add-On .....	33
Installing Support for the SAP RFC Node.....	33
<b>Technical Support</b> .....	<b>34</b>
Frequently Asked Questions (FAQ).....	34
<b>Glossary</b> .....	<b>36</b>



# Introduction

- [Conventions Used in this Document](#)
- [DataFlux References](#)

## Accessibility

DataFlux Data Management Platform (Data Management Platform) includes features that improve usability of the product for users with disabilities. These features are related to accessibility standards for electronic information technology that were adopted by the United States (U.S.) Government under Section 508 of the U.S. Rehabilitation Act of 1973, as amended.

If you have questions or concerns about the accessibility of DataFlux products, send an e-mail to [techsupport@dataflux.com](mailto:techsupport@dataflux.com).

## 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
<b>Bold</b>	Text in bold signifies a button or action
<i>italic</i>	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 <b>Start &gt; Programs &gt; DataFlux Data Management Studio 2.2 &gt; Documentation.</b>

# 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 Online Help

*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

# Installing Data Management Studio

- [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:

- [Obtain the Microsoft .Net Framework](#)
- [License Data Management Studio](#)
- [Run the Data Management Studio Installation Wizard](#)
- [Install the Quality Knowledge Base](#)
- [Perform Reinstallations](#)

### Obtain the Microsoft .Net Framework

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

### License 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 **Imhostid** 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 2.2**.
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 2.2**.
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 2.2**. 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).

## Run 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® web site.
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.

## Install 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.

## Perform 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.



**Note:** A number of features in Data Management Studio 2.2 will not work unless you have Java 1.5 or later. Please verify that your Java version meets this requirement.

## Data Storage

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

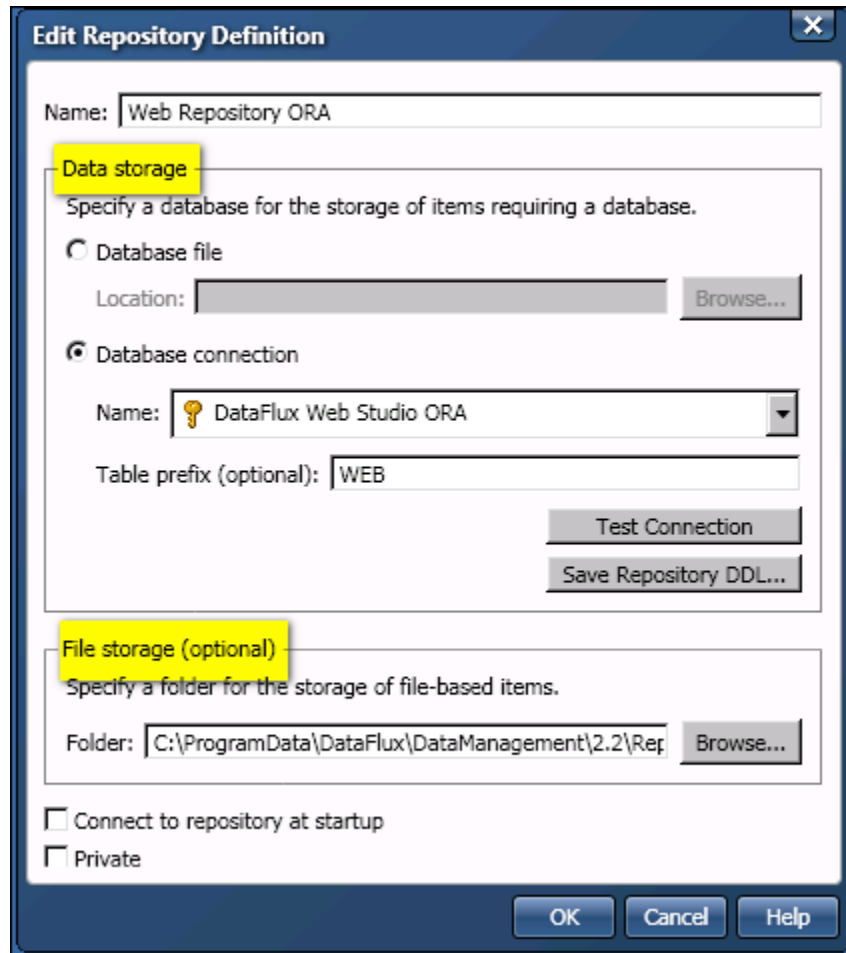
Database Version	Driver	ODBC
Btrieve® 6.15	Btrieve	X
Clipper™	dBASE File	X
dBASE™ IV and V	dBASE	X
FoxPro 2.5, 2.6, and 3.0	dBase	X
FoxPro 6.0 (with 3.0 functionality only)	dBase	X
FoxPro 3.0 Database Container (DBC)	dBase	X
Greenplum Database 3.0, 3.2, 3.3, and 4	Greenplum Wire Protocol	X
IBM® DB2® v8.x, v9.1, v9.5, and v9.7 for Linux®, UNIX, and Windows®	DB2 Wire Protocol	X
IBM DB2 Universal Database™ (UDB) v7.x and v8.x for Linux, UNIX, and Windows	DB2 Wire Protocol	X
IBM DB2 v9.1 and v10 for z/OS	DB2 Wire Protocol	X
IBM DB2 UDB v7.x, v8.1, and v9.1 for z/OS	DB2 Wire Protocol	X
IBM DB2 UDB V5R1, V5R2, V5R3, V5R4, V6R1, and V7R1 for iSeries®	DB2 Wire Protocol	X
Informix© Dynamic Server 9.2, 9.3, 9.4, 10.0, 11.0, and 11.5	Informix	X
Informix Dynamic Server 9.2, 9.3, 9.4, 10.0, 11.0, and 11.5	Informix Wire Protocol	X
Microsoft® SQL Server® 7.0	SQL Server Classic Wire Protocol	X

Database Version	Driver	ODBC
Microsoft SQL Server 2000 - Enterprise Edition (64-bit) and Desktop Engine (MSDE 2000)	SQL Server Classic Wire Protocol	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 - Enterprise Edition (64-bit) and Desktop Engine (MSDE 2000)	SQL Server Native Wire Protocol	X
Microsoft SQL Server 2005	SQL Server Native Wire Protocol	X
Microsoft SQL Server 2008 R1, R2	SQL Server Native Wire Protocol	X
MySQL™ 5.0x, 5.1, and 5.5 server	MySQL Wire Protocol	X
Oracle® 8.0.5 and higher (32-bit only)	Oracle	X
Oracle 8i R1 and R2, R3 (8.1.5 and 8.1.6) (32-bit only)	Oracle	X
Oracle 8i R3 (8.1.7)	Oracle	X
Oracle 9i R1, R2 (9.0.1 and 9.2)	Oracle	X
Oracle 10g R1 and R2 (10.1 and 10.2)	Oracle	X
Oracle 11g R1 and R2 (11.1 and 11.2)	Oracle	X
Oracle 8i R2, R3 (8.1.6 and 8.1.7)	Oracle Wire Protocol	X
Oracle 9i R1 and R2 (9.0.1 and 9.2)	Oracle Wire Protocol	X
Oracle 10g R1 and R2 (10.1 and 10.2)	Oracle Wire Protocol	X
Oracle 11g R1 and R2 (11.1 and 11.2)	Oracle Wire Protocol	X
Pervasive.SQL™ 7.0, 8.5, and 2000	Btrieve	X
PostgreSQL® 8.2, 8.3, and 8.4	PostgreSQL Wire Protocol	X
Sybase® Adaptive Server® 11.5 and higher	Sybase Wire Protocol	X
Sybase Adaptive Server Enterprise® 12.0, 12.5x, 15.0, and 15.5	Sybase Wire Protocol	X
Teradata® 12.0, 13.0, and 13.1	Teradata	X
Teradata V2R6.0, V2R6.1, and V2R6.2	Teradata	X
Text Files (ASCII text files)	Text	X
XML Documents (tabular and hierarchical formatted) 32-bit and Windows only	XML	X

## Repository Storage

### Overview

To add or update a repository, you will access the Repository Definition dialog. This dialog is used to create a repository definition file (.RCF file). There are two main sections of the Repository Definition dialog, as shown in the next figure:



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


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

## 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
DB2® V8.x, V9.1, V9.5, and V9.7 for Linux®, UNIX, and Windows	X
DB2® Universal Database (UDB) v7.x for Linux®, UNIX, and Windows	X
Informix Server 10.0, 11.0, and 11.5	X
Microsoft® SQL Server® 2008 R1, R2	X
Microsoft SQL Server 2005	X

Database	ODBC Driver
Microsoft SQL Server 2000 Enterprise Edition (64-bit)	X
Microsoft SQL Server 2000 Desktop Engine (MSDE 2000)	X
Microsoft SQL Server 2000 Service Packs 1, 2, 3, 3a and 4)	X
Oracle® 11g R1, R2 (11.1, 11.2)	X
Oracle 10g R1, R2(10.1, 10.2)	X
Oracle 9i R1, R2 (9.0.1, 9.2)	X
Sybase® Adaptive Server® 12.0, 12.5x, 15, and 15.5	X
Teradata® 12.0 and 13.0	X

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

## File Storage for Repositories

The file storage location can be any physical path that is accessible to Data Management Studio, which is a Windows application. Any version of Windows that supports Data Management Studio can of course be used for file storage. For more information, see [Operating Systems](#).

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

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

## Supported Platforms

- [Operating Systems](#)
- [License Servers](#)
- [Virtual Environments](#)

## Operating Systems


Data Management Studio is a 32-bit application that runs on Microsoft Windows. It can be used on the following versions of Windows:

<b>Supported Operating Systems</b>	<b>32-bit</b>	<b>32-bit compatibility mode on 64 bit OS</b>	<b>64-bit</b>
Windows Server 2003® SP1 (Small Business Server, Web Edition, Datacenter, Standard)	X	X	n/a
Windows Server 2003 SP2 (Small Business Server, Web Edition, Datacenter, Standard) x86			
Windows Server 2003 SP2 (Small Business Server, Web Edition, Datacenter, Standard)	X	X	n/a
Windows Server 2003 SP1 (Enterprise) x86			
Windows Server 2003 SP1 (Enterprise) x64	X	X	n/a
Windows Server 2003 SP2 (Enterprise) x86			
Windows Server 2003 SP2 (Enterprise) x64	X	X	n/a
Windows XP Professional® SP2 on an x86 chip			
Windows XP Professional SP2 on an x64 chip	X	X	n/a
Windows XP Professional SP3 on an x86 chip			
Windows XP Professional SP3 on an x64 chip	X	X	n/a
Windows Vista® Business on an x64 chip	X	X	n/a
Windows Vista (Enterprise, Ultimate) on an x64 chip	X	X	n/a
Windows Server 2008® (Standard, HPC, Small Business, Datacenter, Datacenter without Hyper-V, Enterprise without Hyper-V, Foundation, Standard without Hyper-V)	X	X	n/a
Windows Server 2008 R2 (Standard, HPC, Small Business, Datacenter, Datacenter without Hyper-V, Enterprise without Hyper-V, Foundation, Standard without Hyper-V)	X	X	n/a
Windows Server 2008 (Enterprise)	X	X	n/a
Windows Server 2008 R2 (Enterprise)	X	X	n/a
Windows Server 2008 for Itanium-based Systems on an Itanium chip			
Windows Server 2008 R2 for Itanium-based Systems on an Itanium chip			
Windows 7® (the 64 bit OS) on an x64 chip (Starter, Home Basic, Home Premium, Professional, Ultimate)		X	n/a
Windows 7 (the 64 bit OS) on an x64 chip (Enterprise)		X	n/a

## License Servers

A DataFlux License Server 2.2 can be installed on the following server types:

Supported Server Types (Minimum Versions)	32-bit	32-bit compatibility mode on 64 bit OS	64-bit
AIX® 5.3			
HP/UX® 11.23 on a PA-RISC® chip (11i v2.0)			
HP/UX 11.23 on an Itanium chip (11i v2.0)			
Red Hat® Enterprise Linux LSB 3.0	X		X
Solaris™ 9 on a SPARC® chip			
Solaris 10 on an x86 chip			
Windows Server 2003 (NT 5.2)	X		X

 **Note:** The Linux server is supported on any Linux system that is compliant with Linux Server Base (LSB) 3.0 or greater. You can determine the level of LSB compliance by running the command `lsb_release` on the Linux system.

## 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.



# Configuring Data Management Studio

- [Configuration Files](#)
- [Configuration Options](#)
- [Data Access Component Directives](#)
- [Logging Options](#)

## 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 dmserver.cfg.
4. The application-specific configuration files in a user folder.
5. The macros folder in the etc folder. The default path to the macros folder can be overridden with BASE/MACROS\_PATH setting in the 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	Source	Notes
<b>General Application</b>			
ODBC_INI	Overrides the location of the odbc.ini file		UNIX only
BASE/LIBRARY_PATH	Path for Java jar dependencies	Optional	Determined by startup code (DFEXEC_HOME/lib)
BASE/PLUGIN_PATH	Path used by all subsystems to find plugins	Optional	Determined by startup code
BASE/EXE_PATH	Path containing executables	Optional	Calculated
BASE/PRIMARY_LICENSE	Primary licensing method	Required by base	Must be set in the configuration file as DATAFLUX or SAS
BASE/MACROS_PATH	Path for system macros.cfg file	Optional	If not specified, this file is located in the <b>etc</b> subfolder of the installation folder.
BASE/PRIMARY_LICENSE_LOC	Location of the primary license file or server	Required by base	Must be set in the configuration file
BASE/SECONDARY_LICENSE	Secondary licensing method	Required by base	Must be set in the configuration file as DATAFLUX or SAS
BASE/SECONDARY_LICENSE_LOC	Location of the secondary license file or server	Required by base	Must be set in the configuration file
BASE/LOGCONFIG_PATH	Full path to the log configuration file	Optional	Must be set in the configuration file or it defaults to logging.xml in the <b>etc</b> directory.
BASE/MESSAGE_PATH	Path to the message directory	Optional	Determined by startup code
BASE/MESSAGE_LOCALE	Error message locale	Optional	If not specified, it is determined from the system locale

BASE/MESSAGE_LEVEL	Error level of messages	Optional	0 (or not specified) - normal messages; 1 - includes source file and line number in messages
BASE/USER_PATH	Path for user configuration files	Optional	Automatically determined by dfcurver
BASE/REPOS_SYS_PATH	System path for repository configuration files	Optional	Automatically determined
BASE/REPOS_USER_PATH	User directory for repository configuration files	Optional	Automatically determined by dfcurver
BASE/REPOS_FILE_ROOT	Overrides the root of the repository for URI lookups	Optional	If specified this is used as the root for repository when resolving the URI. In other words, the path in the URI will be concatenated to this path to give the actual filename of a URI
BASE/REPOS_EVENT_WAIT_QUERYMS	Repository event that processes the wait time between processing queries.	Optional	Specifies how frequently in milliseconds to query the repository for changes in the event table. This might need to be changed due to slow servers or IT issues. This is an overriding value and the default is used if no value is set by the user. A setting of -1 disables events from client.
BASE/REPOS_EVENT_CLEAN_TIMEMIN	Repository event processor that removes all events older than X minutes before startup.	Optional	
BASE/TEMP	Temporary directory	Optional	If not specified, it inherits the value of the TEMP environment variable

BASE/DATE_FORMAT	Specific date formats	Optional	If specified, it is iso8601
BASE/APP_VER	Application version number	Optional	Defaults to 2.2
BASE/UPDATE_LEVEL	Application update level	Optional	Defaults to 0. Could be used as a minor revision number
BASE/AUTH_SERVER_LOC	Location of DataFlux Authentication Server	Optional	If specified, contains the IOM URI to an Authentication Server. The basic format of the Authentication Server IOM URI is <code>iom://&lt;host&gt;:&lt;port&gt;</code> . Where <host> is the name of the computer where the Authentication Server is executing and <port> is the port in which the Authentication Server may be contacted. The port should be specified as 21030 unless the default port for the Authentication Server has been changed. For more information, see the topic "Run a Job with Domain-Enabled ODBC Connections" in the Data Jobs chapter of the <i>DataFlux Data Management Studio User's Guide</i> .
<b>Data Access Component Logging</b>			
DAC/DFTKLOGFILE	DFTK logging	Optional	Filename
DAC/TKTSLOGFILE	TKTS logging	Optional	Filename
DAC/DFTKDISABLECEDA	Disables CEDA support	Optional	"Yes" turns it on
DAC/SAVEDCONNSYSTEM	Location of system saved connections	Optional	Defaults to DFEXEC_HOME/etc/dsn
DAC/SAVEDCONNUSER	Location of user saved connections	Optional	Defaults to the application directory/DataFlux/dac/9.1
DAC/DSN	DSN directory for TKTS dsns	Optional	Path that defaults to DFEXEC_HOME/etc/dftkdsn
DAC/DFTK_PROCESS	Run DFTK out of process	Optional	"Yes" turns it on; off by default
DAC/DFTK_PROCESS_TKPATH	TKTS path for DFTK out of process	Optional	Path that defaults to a core/sasext dir off the executable dir
<b>Profile</b>			
PROF/DEBUG_MODE	Frequency distribution engine debug mode	Optional	Possible values include 0 not debug mode or 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	Optional	Any numeric value. Default is -1 (frequency distribution engine default)
<b>QKB</b>			
QKB/PATH	Path to QKB	Required by QKB products	Path is set to the default QKB defined in application.
QKB/SURFACEALL	Surfaces all parse definitions	Optional	Default is NO. Note that the application start up creates a blue fusion pool which sets the option for all consumers (Profile, Explorer and Nodes) with the exception of the Expression Engine which continue to have its own initialization.
QKB/COMPATVER	Denotes which compatible version to use when interacting with QKB	Optional	Possible values: dfpower82, unity22 Default: unity21. Note that the application start up creates a blue fusion pool which sets the option for all consumers (Profile, Explorer and Nodes) with the exception of the Expression Engine which continue to have its own initialization.
QKB/ALLOW_INCOMPAT	Allow use of incompatible Quality Knowledge Base definitions	Optional	Default is NO. Note that the application start up creates a blue fusion pool which sets the option for all consumers (Profile, Explorer and Nodes) with the exception of the Expression Engine which continue to have its own initialization.
QKB/ON_DEMAND	Loads QKB definitions on demand	Optional	Default is YES. Note that the application start up creates a blue fusion pool which sets the option for all consumers (Profile, Explorer and Nodes) with the exception of the Expression Engine which continue to have its own initialization.

<b>Base</b>			
BASE/SORTBYTES	Specifies the bytes to use when sorting	Optional	
CLUSTER/BYTES	Specifies the bytes use when clustering	Optional	
CLUSTER/LOG	Specifies whether a clustering log is needed	Optional	
FRED/LOG	Specifies whether a FRED log is needed	Optional	
BASE/EMAILCMD	Specifies the command used to send email	Required	Can include %T and %B where %T is replaced with the recipient and %B is a file containing the body of the message; also used by monitor event as well as architect nodes.
BASE/SORTMERGES	Enables merge during sort	Optional	
BASE/SORTTEMP	Specifies the temporary path for sorts	Optional	
BASE/SORTTHREADS	Specifies the number of sort threads	Optional	
ARCHITECT/AutoPassThru	Client option to set mappings	Optional	Maintained by client; choices are 0 (target), 1 (Source and Target), and 2 (All)
<b>Verify</b>			
VERIFY/CACHESIZE	Specifies a percentage value	Optional	
VERIFY/CANADA	Specifies the path to Canadian data	Required by SERP nodes	Maintained by Canada installation

VERIFY/GEO	Specifies the geo/phone path	Required by Geo	Maintained by Geo installation
VERIFY/PRELOAD	Specifies the preload string for verify	Optional	Valid values are ALL or empty string
VERIFY/USEDPV	Enables or disables the PDV processing	Optional	
VERIFY/USERDI	Enables or disables the RDI processing	Optional	Residential Delivery Indicator (RDI)
VERIFY/USELACS	Enables or disables the LACSLink processing	Optional	Locatable Address Conversion System (LACS)
VERIFY/USEELOT	Enables or disables the eLOT processing	Optional	
VERIFY/USPS	Specifies the USPS data path	Required by USPS	Maintained by USPS installation
VERIFY/UPSPINST	Determines whether the USPS data is installed or if sample data is being used	Required	Maintained by USPS installation
VERIFYINTL/CFG	Verifies the international addresses	Required by international verification	Path maintained by component installation
VERIFYWORLD/DB	Specifies the Platon data path	Required for Platon	Path maintained by component installation
VERIFYWORLD/UNLK	Specifies the Platon library universal unlock code	Required for Platon	Path maintained by component installation

CLUSTER/TEMP	Specifies the cluster temporary path	Optional	
BASE/FTPGETCMD	Specifies the command used for Ftp Get Functionality	Required	Should default in the install, as follows: <ul style="list-style-type: none"> <li>• %U: Replace with username</li> <li>• %P: Replace with password</li> <li>• %S: Replace with server</li> <li>• %T: Replace with local directory</li> <li>• %F: Replace with Files to download, multiple separated by spaces</li> <li>• %L: Replace with the log file to pipe the output</li> </ul>
BASE/FTPPUTCMD	Specifies the command used for Ftp Put Functionality	Required	
JAVA/CLASSPATH	Specifies the Java classpath	Optional	
JAVA/VMLIB	Specifies the path of the jvm.dll	Optional	This path should be something like C:\Program Files\Java\jre6\bin\client\jvm.dll.
JAVA/DEBUG		Optional	
JAVA/DEBUGPORT		Optional	
JAVA/VMHEAP_INIT	Specifies the initial JVM heap size in MB	Optional	This should be used if you are using JVM 1.6. Start at 16 and decrease if you keep getting errors.
JAVA/VMHEAP_MAX	Specifies the maximum JVM heap size in MB	Optional	This should be used if you are using JVM 1.6. You can decrease this along with JAVA/VMHEAP_INIT.
<b>dfIntelliServer</b>			
DFCLIENT/CFG	Used for dfIntelliServer	Required	Maintained by dfIntelliServer installation; typical location is 'C:\Program Files\DataFlux\dfIntelliServer\etc\dfclient.cfg; modify the dfclient.cfg file to point to the server and port



Other			
EXPRESS_MAX_STRING_LENGTH	Specifies the maximum size of strings declared in expression nodes	Optional	Default maximum length of any string in this node is 5,242,880 bytes (5MB). This enables specifying a larger value in bytes. If performance issues arise, the suggested setting is 65536 bytes.
EXPRESSION/UDFDIR	Specifies where to look for UDF files	Optional	If not specified, UDF files will be sought in installationdir/etc/udf
REPOS/CREATE_SPEC_PATH	Specifies how to create the repository table or index	Optional	This specification provides a means of configuring the commands to create tables and indexes in the repository.
REPOS/LOCK_RETRIES	Specifies the number of attempts to connect to a SQLite repository	Optional	Number of times to retry SQLite repository connection when a connect attempt times out or -1 to retry until a connection is established.
SAP_LIBPATH	Specifies the location of SAP RFC libraries on UNIX only.	Optional	These shared libraries are installed to support the SAP Remote Function Call node, a data job node in Data Management Studio. For more information, see <a href="#">Installing Support for the SAP RFC Node</a> .
WFE/ENGINE_THREAD_LIMIT	Specifies the thread pool limits for the work flow engine.	Optional	Use this setting to limit the number of engine threads. The default is 0 meaning unbounded, which will defer to the system for the thread pool limits. The optimal setting is the number of processors + 1.

## 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.

The DAC reads a list of settings to determine what actions to take, where to find various things it needs, and how to make the proper connections.


There are three types of DAC Configuration Directives:

**Configuration options:** Most of the DAC's settings come from configuration values as specified in app.cfg or other config files which are read in a specific sequence. It is

recommended to put these settings in the app.cfg file, but they can be put in macro.cfg or another application-specific configuration file. See [Configuration Options](#) for a more comprehensive list of possible configuration values.

**Registry:** The DAC only looks in the registry for DSN-specific items. These are found in either or both HKEY\_CURRENT\_USER\Software\DataFlux\dac\{version}\{dsn} or HKEY\_LOCAL\_MACHINE\Software\DataFlux\dac\{version}\{dsn}. "dsn" in this case is a string with the same name as the DSN to be affected.

**Environment variables:** You can identify the environment variables with \$NAME in the following table. Typically, the application using the DAC will set environment variables to appropriate locations. On Windows, \$DFEXEC\_HOME is set to the application's root directory.

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

Setting	Description	Windows	UNIX
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. 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.	The same as Windows.
Connection and login timeout	Allows you to specify the time in seconds for a connection timeout and a login timeout.	The string type value is located in USER\dsn_name\logintimeout (or connecttimeout) where dsn_name is the name of the dsn. The string value is the number of seconds for the time out. If it does not find it, the DAC will look in the SYSTEM\dsn_name\logintimeout.	Not supported.
DAC logging	Determines whether to create a log file for DAC operations.	This is controlled through the effective logging configuration files, which uses log4sas. The logger name "DAC" should be used.	The same as Windows.
DFTK log file	Specifies the log file that interacts with the DFTKSRV layer and is only useful for debugging issues specific to dftksrv. This setting is only applicable if you are	Look for config value DAC/DFTKLOGFILE.	The same as Windows.

<b>Setting</b>	<b>Description</b>	<b>Windows</b>	<b>UNIX</b>
	running DFTK out of process.		
Disable CEDA	Specifies whether to disable CEDA. This setting is only applicable to tkts connections.	Look for DAC/DFTKDISABLECEDA configuration value, which should specify any non-null value, for example, yes.	The same as Windows.

Multiple active results sets (MARS)	Specifies whether to allow a connection to open multiple result sets at the same time. This is only applicable when using the SQL Server Native Client driver on Windows against an SQL Server 2005 (or later).	Look in USER\ <i>dsn_name</i> \mars for dword value of 1 where <i>dsn_name</i> is the name of the dsn. If it does not find it, the DAC will look in the SYSTEM\ <i>dsn_name</i> \mars.	Not applicable.
Oracle NUMBER(38) handling	If connected to Oracle (only), NUMBER(38) columns is treated as INTEGER by default. To override that functionality, treat them as REAL.	Look in USER\or anum38real for dword value of 1. If it is not found in USER, look in SYSTEM\or anum38real.	Look in \$HOME/.dfpower/dsn.cfg for a line dsn_name = or anum38real where dsn_name is the name of the DSN.
Run DFTK out of process	Specifies whether to run TKTS out of process, allowing you to perform troubleshooting.	Look for config value DAC/DFTK_PROCESS. This should specify any non-null value, for example, yes.	The same as Windows.

Suffix for CREATE TABLE statements	This allows you to provide a string that is appended to every CREATE TABLE statement. If you include %t in this string, it is substituted with the table name.	Look in USER\ <i>dsn_name</i> \postcreate for a string where <i>dsn_name</i> is the name of the dsn. If it is not found in USER, look in SYSTEM.	Not supported.
System saved connection	Specifies where to find system saved connections.	Look in DAC/SAVEDCONNSYSTEM config value for a path. If it is not found, the DAC checks for the DFEXEC_HOME environment variable in the \$DFEXEC_HOME\etc\dsn subdirectory, which is in the installation directory.	Look for the same config value. If not found, look in the \$DFEXEC_HOME/etc/dsn subdirectory, which is in the installation directory.

TK Path	Specifies where 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.	Look for config value DAC/DFTK_PROCESS_PATH. If still not found, set it to \$DFEXEC_HOME\bin;\$DFEXEC_HOME\bin\core\sasext.	Same as for Windows, but will set it to \$DFEXEC_HOME/lib/tkts if it is not found.
TKTS DSN directory	Specifies the path where TKTS DSNs are stored in XML files.	Look in config value DAC/DSN. If it is not found, look in \$DFEXEC_HOME/etc\dftkdsn\.	Look in the same config value. If it is not found, look in \$DFEXEC_HOME/etc/dftkdsn/.
TKTS log file	Specifies the log file that is produced by the TKTS layer and is useful for debugging.	Look for config value DAC/TKTSLOGFILE.	The same as Windows.

	g tkts issues.		
--	-------------------	--	--

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 registry for a tktssleep value.	Not supported.
User saved connection	Specifies where to find user-saved connections.	Look in DAC/SAVEDCONNUSER config value for a path. If not found, look in the user's application settings directory, which is usually in the %APPDATA% directory, in the DataFlux\dac\version subdirectory.  The %APPDATA% location is C:\Users\userid\AppData\ for Windows 7 and C:\Documents and Settings\userid\Application Data\ for Windows XP.	Look for the same config value. If not found, look in the \$HOME/.dfpower/dsn directory.
Use braces	Specifies whether to enclose DSN items with braces when they contain reserved characters. Most drivers do not support use of braces; however, this might be an option if problems occur with saving credentials.	The DAC looks in USER\usebraces for dword value of 1. If it is not found there, the DAC looks in SYSTEM\usebraces.	Look in \$HOME/.dfpower/dsn.cfg for a line dsn_name = usebraces where dsn_name is the name of the DSN.



# Logging Options

The following log files are provided for Data Management Studio:

- Studio log
- Platform log
- DAC log
- TKTS log

The Studio log, the Platform log, and the DAC log are enabled by default. The Studio, Platform, and DAC logs use the SAS 9.2 Logging Facility. The SAS 9.2 Logging Facility logging facility is a flexible and configurable framework that you can use to collect, categorize, and filter events. Then you can write them to a variety of output devices. The logging facility supports problem diagnosis and resolution, performance and capacity management, and auditing and regulatory compliance.

The logging facility framework categorizes and filters log messages in SAS server and SAS programming environments. It also writes log messages to various output devices. In the server environment, the logging facility logs messages based on predefined message categories such as the following:

- Admin for administrative messages
- App for application messages
- Perf for performance messages

Messages for a category can be written to files, consoles, and other system destinations simultaneously. The logging facility also enables messages to be filtered based on the following thresholds: TRACE, DEBUG, INFO, WARN, ERROR, and FATAL.

You can also configure the application to provide the TKTS log, which does not use the SAS 9.2 Logging Facility.

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

Name	Scope	Level Thresholds	Configuration Location
Studio	Data Management Studio events	TRACE, DEBUG, INFO, WARN, ERROR, and FATAL	By default these are enabled. If you want to turn them off, comment out the BASE/LOGCONFIGPATH line in ui.cfg and in batch.cfg and restart Data Management Studio.  For configuration, go to <i>drive:\Program Files\DataFlux\DMStudio\[version]\ etc\platform.log.xml</i>
Platform	Data Management Studio	TRACE, DEBUG,	

Name	Scope	Level Thresholds	Configuration Location
	engine events	INFO, WARN, ERROR, and FATAL	and <i>drive:</i> ]\Program Files\DataFlux\DMStudio\[ <i>version</i> ]\etc\dfwfproc.log.xml
DAC	Data access events	TRACE, DEBUG, INFO, WARN, ERROR, and FATAL	
TKTS	TKTS events such as BASE data sets and data sent across the wire to Federation Server	Not configurable	<i>drive:</i> \Program Files\DataFlux\DMStudio\[ <i>version</i> ]\etc\app.cfg

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

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

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

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

# Installing Add-On Products

- [Installing a Quality Knowledge Base](#)
- [Installing Data Packs](#)
- [Installing Supplemental Language Support](#)
- [Installing the Address Update Add-On](#)
- [Installing Support for the SAP RFC Node](#)

## 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, 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, installing 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. For path information on all configuration options, see the section on [Configuration Options](#).

## 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.

## Installing the Address Update Add-On

The DataFlux® Address Update add-on enables you to use the United States Postal Service (USPS) NCOALink® system to identify and update customer address information. To install the required resources for this add-on, see the *Address Update Add-On to Data Management Studio 2.2 QuickStart Guide*.

## Installing Support for the SAP RFC Node

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

- read information from SAP that is not readily available in tables
- use the functionality that SAP exposes as BAPIs
- load data to SAP by using SAP-provided functions that are a standard way of updating objects
- trigger an upload to SAP by using functions that take, for example, an external file input.

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

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

Data Management Servers can be either 32-bit or 64-bit. See the SAP documentation for details about installing the appropriate libraries.

Data Management Studio is a Windows application, but Data Management Servers can run on UNIX. For UNIX systems only, you can set the SAP\_LIBPATH option in the app.cfg file to point to the SAP DLLs.

# 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](#).

- [System Requirements](#)
- [Supported Platforms](#)
- [Data Storage](#)
- [Repository Storage](#)
- [Logging Options](#)
- [Configuration Files](#)

## System Requirements

### What are the recommended system requirements for Data Management Studio?

The recommended system requirements for Data Management Studio can be found in [System Requirements](#).

## Supported Platforms

### On what platforms can I install Data Management Studio?

Data Management Studio can be installed on many different platforms. For the full list of supported platforms, see the section on [Supported Platforms](#).

## Data Storage

### From what databases can DataFlux Data Management Studio access data?

Data Management Studio can access data from any of the databases listed in the [Data Storage](#) chapter.

## Repository Storage

### What databases are supported for Data Management Platform repositories?

Data Management Studio can access data from any of the databases listed in the [Database Storage for Repositories](#) chapter.

## Logging Options

### What options do I have for logging?

Data Management Studio provides logging files for Studio, Platform, and DAC. An additional TKTS log file is also available. See [Logging Options](#) section for the scope, level thresholds, and configuration location for each of these logs.

## Configuration Files

### In a previous installation of Data Management Studio, I had a config file named `dis.cfg`. I don't see that file. Where is it?

The `dis.cfg` file used in previous versions of Data Management Studio was renamed `dmserver.cfg`.

# 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**

Data Management 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 is 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 Data Management 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 Data Management Studio Navigator and accessed by the Data Management 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 Data Management 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 Data Management Studio 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 Data Management Studio. 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 Data Management Studio 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.