



Informatica®
10.1.1

Analyst Tool Guide

This software and documentation are provided only under a separate license agreement containing restrictions on use and disclosure. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording or otherwise) without prior consent of Informatica LLC.

Informatica, and the Informatica logo are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners.

Portions of this software and/or documentation are subject to copyright held by third parties, including without limitation: Copyright DataDirect Technologies. All rights reserved. Copyright © Sun Microsystems. All rights reserved. Copyright © RSA Security Inc. All Rights Reserved. Copyright © Ordinal Technology Corp. All rights reserved. Copyright © Aandacht c.v. All rights reserved. Copyright Genivia, Inc. All rights reserved. Copyright Isomorphic Software. All rights reserved. Copyright © Meta Integration Technology, Inc. All rights reserved. Copyright © Intalio. All rights reserved. Copyright © Oracle. All rights reserved. Copyright © Adobe Systems Incorporated. All rights reserved. Copyright © DataArt, Inc. All rights reserved. Copyright © ComponentSource. All rights reserved. Copyright © Microsoft Corporation. All rights reserved. Copyright © Rogue Wave Software, Inc. All rights reserved. Copyright © Teradata Corporation. All rights reserved. Copyright © Yahoo! Inc. All rights reserved. Copyright © Glyph & Cog, LLC. All rights reserved. Copyright © Thinkmap, Inc. All rights reserved. Copyright © Clearpace Software Limited. All rights reserved. Copyright © Information Builders, Inc. All rights reserved. Copyright © OSS Nokalva, Inc. All rights reserved. Copyright Edifecs, Inc. All rights reserved. Copyright Cleo Communications, Inc. All rights reserved. Copyright © International Organization for Standardization 1986. All rights reserved. Copyright © ej-technologies GmbH. All rights reserved. Copyright © Jaspersoft Corporation. All rights reserved. Copyright © International Business Machines Corporation. All rights reserved. Copyright © yWorks GmbH. All rights reserved. Copyright © Lucent Technologies. All rights reserved. Copyright © University of Toronto. All rights reserved. Copyright © Daniel Veillard. All rights reserved. Copyright © Unicode, Inc. Copyright IBM Corp. All rights reserved. Copyright © MicroQuill Software Publishing, Inc. All rights reserved. Copyright © PassMark Software Pty Ltd. All rights reserved. Copyright © LogiXML, Inc. All rights reserved. Copyright © 2003-2010 Lorenzi Davide, All rights reserved. Copyright © Red Hat, Inc. All rights reserved. Copyright © The Board of Trustees of the Leland Stanford Junior University. All rights reserved. Copyright © EMC Corporation. All rights reserved. Copyright © Flexera Software. All rights reserved. Copyright © Jinfonet Software. All rights reserved. Copyright © Apple Inc. All rights reserved. Copyright © Telerik Inc. All rights reserved. Copyright © BEA Systems. All rights reserved. Copyright © PDFlib GmbH. All rights reserved. Copyright © Orientation in Objects GmbH. All rights reserved. Copyright © Tanuki Software, Ltd. All rights reserved. Copyright © Ricebridge. All rights reserved. Copyright © Sencha, Inc. All rights reserved. Copyright © Scalable Systems, Inc. All rights reserved. Copyright © jqWidgets. All rights reserved. Copyright © Tableau Software, Inc. All rights reserved. Copyright © MaxMind, Inc. All Rights Reserved. Copyright © TMatte Software s.r.o. All rights reserved. Copyright © MapR Technologies Inc. All rights reserved. Copyright © Amazon Corporate LLC. All rights reserved. Copyright © Highsoft. All rights reserved. Copyright © Python Software Foundation. All rights reserved. Copyright © BeOpen.com. All rights reserved. Copyright © CNRI. All rights reserved.

This product includes software developed by the Apache Software Foundation (<http://www.apache.org/>), and/or other software which is licensed under various versions of the Apache License (the "License"). You may obtain a copy of these Licenses at <http://www.apache.org/licenses/>. Unless required by applicable law or agreed to in writing, software distributed under these Licenses is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the Licenses for the specific language governing permissions and limitations under the Licenses.

This product includes software which was developed by Mozilla (<http://www.mozilla.org/>), software copyright The JBoss Group, LLC, all rights reserved; software copyright © 1999-2006 by Bruno Lowagie and Paulo Soares and other software which is licensed under various versions of the GNU Lesser General Public License Agreement, which may be found at <http://www.gnu.org/licenses/lgpl.html>. The materials are provided free of charge by Informatica, "as-is", without warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability and fitness for a particular purpose.

The product includes ACE(TM) and TAO(TM) software copyrighted by Douglas C. Schmidt and his research group at Washington University, University of California, Irvine, and Vanderbilt University, Copyright (©) 1993-2006, all rights reserved.

This product includes software developed by the OpenSSL Project for use in the OpenSSL Toolkit (copyright The OpenSSL Project. All Rights Reserved) and redistribution of this software is subject to terms available at <http://www.openssl.org> and <http://www.openssl.org/source/license.html>.

This product includes Curl software which is Copyright 1996-2013, Daniel Stenberg, <daniel@haxx.se>. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://curl.haxx.se/docs/copyright.html>. Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

The product includes software copyright 2001-2005 (©) MetaStuff, Ltd. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://www.dom4j.org/license.html>.

This product includes software copyright © 1996-2006 Per Bothner. All rights reserved. Your right to use such materials is set forth in the license which may be found at <http://www.gnu.org/software/kawa/Software-License.html>.

This product includes OOSP UUID software which is Copyright © 2002 Ralf S. Engelschall, Copyright © 2002 The OOSP Project Copyright © 2002 Cable & Wireless Deutschland. Permissions and limitations regarding this software are subject to terms available at <http://www.opensource.org/licenses/mit-license.php>.

This product includes software developed by Boost (<http://www.boost.org/>) or under the Boost software license. Permissions and limitations regarding this software are subject to terms available at http://www.boost.org/LICENSE_1_0.txt.

This product includes software copyright © 1997-2007 University of Cambridge. Permissions and limitations regarding this software are subject to terms available at <http://www.pcre.org/license.txt>.

This product includes software copyright © 2007 The Eclipse Foundation. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://www.eclipse.org/org/documents/epl-v10.php> and at <http://www.eclipse.org/org/documents/edl-v10.php>.

This product includes software licensed under the terms at <http://www.tcl.tk/software/tcltk/license.html>, <http://www.bosrup.com/web/overlib/?License>, <http://www.stlport.org/doc/license.html>, <http://asm.ow2.org/license.html>, <http://www.cryptix.org/LICENSE.TXT>, <http://hsqldb.org/web/hsqLicense.html>, <http://httpunit.sourceforge.net/doc/license.html>, <http://jung.sourceforge.net/license.txt>, http://www.gzip.org/zlib/zlib_license.html, <http://www.openldap.org/software/release/license.html>, <http://www.libssh2.org>, <http://slf4j.org/license.html>, <http://www.sente.ch/software/OpenSourceLicense.html>, <http://fusesource.com/downloads/license-agreements/fuse-message-broker-v-5-3-license-agreement>, <http://antlr.org/license.html>, <http://aopalliance.sourceforge.net/>, <http://www.bouncycastle.org/license.html>, <http://www.jgraph.com/jgraphdownload.html>, <http://www.jcraft.com/jsch/LICENSE.txt>, http://jotm.objectweb.org/bsd_license.html, <http://www.w3.org/Consortium/Legal/2002/copyright-software-20021231>, <http://www.slf4j.org/license.html>, <http://nanoxml.sourceforge.net/org/copyright.html>, <http://www.json.org/license.html>, <http://forge.ow2.org/projects/javaservice/>, <http://www.postgresql.org/about/license.html>, <http://www.sqlite.org/copyright.html>, <http://www.tcl.tk/software/tcltk/license.html>, <http://www.jaxen.org/faq.html>, <http://www.jdom.org/docs/faq.html>, <http://www.slf4j.org/license.html>, <http://www.iodbc.org/dataspace/iodbc/wiki/IODBC/License>, <http://www.keplerproject.org/md5/license.html>, <http://www.toedter.com/en/jcalendar/license.html>, <http://www.edankert.com/bounce/index.html>, <http://www.net-snmp.org/about/license.html>, <http://www.openmdx.org/#FAQ>, http://www.php.net/license/3_01.txt, <http://www.srp.stanford.edu/license.txt>, <http://www.schneider.com/blowfish.html>, <http://www.jmock.org/license.html>, <http://xsom.java.net/>, <http://benalman.com/about/license/>, <https://github.com/CreateJS/EaselJS/blob/master/src/easeljs/display/Bitmap.js>, <http://www.h2database.com/html/license.html#summary>, <http://jsoncpp.sourceforge.net/LICENSE>, <http://jdbc.postgresql.org/license.html>, <http://protobuf.googlecode.com/svn/trunk/src/google/protobuf/descriptor.proto>, <https://github.com/rantav/hector/blob/master/LICENSE>, <http://web.mit.edu/Kerberos/krb5-current/doc/mitK5license.html>, <http://jibx.sourceforge.net/jibx-license.html>, <https://github.com/lyokato/libgeohash/blob/master/LICENSE>, <https://github.com/hjiang/jsonxx/blob/master/LICENSE>, <https://code.google.com/p/lz4/>, <https://github.com/jedisct1/libsodium/blob/master/>

LICENSE; <http://one-jar.sourceforge.net/index.php?page=documents&file=license>; <https://github.com/EsotericSoftware/kryo/blob/master/license.txt>; <http://www.scala-lang.org/license.html>; <https://github.com/tinkerpop/blueprints/blob/master/LICENSE.txt>; <http://gee.cs.oswego.edu/dl/classes/EDU/oswego/cs/dl/util/concurrent/intro.html>; <https://aws.amazon.com/asl/>; <https://github.com/twbs/bootstrap/blob/master/LICENSE>; <https://sourceforge.net/p/xmlunit/code/HEAD/tree/trunk/LICENSE.txt>; <https://github.com/documentcloud/underscore-contrib/blob/master/LICENSE>, and <https://github.com/apache/hbase/blob/master/LICENSE.txt>.

This product includes software licensed under the Academic Free License (<http://www.opensource.org/licenses/afl-3.0.php>), the Common Development and Distribution License (<http://www.opensource.org/licenses/cddl1.php>) the Common Public License (<http://www.opensource.org/licenses/cpl1.0.php>), the Sun Binary Code License Agreement Supplemental License Terms, the BSD License (<http://www.opensource.org/licenses/bsd-license.php>), the new BSD License (<http://opensource.org/licenses/BSD-3-Clause>), the MIT License (<http://www.opensource.org/licenses/mit-license.php>), the Artistic License (<http://www.opensource.org/licenses/artistic-license-1.0>) and the Initial Developer's Public License Version 1.0 (<http://www.firebirdsql.org/en/initial-developer-s-public-license-version-1-0/>).

This product includes software copyright © 2003-2006 Joe Walnes, 2006-2007 XStream Committers. All rights reserved. Permissions and limitations regarding this software are subject to terms available at <http://xstream.codehaus.org/license.html>. This product includes software developed by the Indiana University Extreme! Lab. For further information please visit <http://www.extreme.indiana.edu/>.

This product includes software Copyright (c) 2013 Frank Balluffi and Markus Moeller. All rights reserved. Permissions and limitations regarding this software are subject to terms of the MIT license.

See patents at <https://www.informatica.com/legal/patents.html>.

DISCLAIMER: Informatica LLC provides this documentation "as is" without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of noninfringement, merchantability, or use for a particular purpose. Informatica LLC does not warrant that this software or documentation is error free. The information provided in this software or documentation may include technical inaccuracies or typographical errors. The information in this software and documentation is subject to change at any time without notice.

NOTICES

This Informatica product (the "Software") includes certain drivers (the "DataDirect Drivers") from DataDirect Technologies, an operating company of Progress Software Corporation ("DataDirect") which are subject to the following terms and conditions:

1. THE DATADIRECT DRIVERS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT.
2. IN NO EVENT WILL DATADIRECT OR ITS THIRD PARTY SUPPLIERS BE LIABLE TO THE END-USER CUSTOMER FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL OR OTHER DAMAGES ARISING OUT OF THE USE OF THE ODBC DRIVERS, WHETHER OR NOT INFORMED OF THE POSSIBILITIES OF DAMAGES IN ADVANCE. THESE LIMITATIONS APPLY TO ALL CAUSES OF ACTION, INCLUDING, WITHOUT LIMITATION, BREACH OF CONTRACT, BREACH OF WARRANTY, NEGLIGENCE, STRICT LIABILITY, MISREPRESENTATION AND OTHER TORTS.

The information in this documentation is subject to change without notice. If you find any problems in this documentation, please report them to us in writing at Informatica LLC 2100 Seaport Blvd. Redwood City, CA 94063.

INFORMATICA LLC PROVIDES THE INFORMATION IN THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2018-10-31

Table of Contents

| | |
|--|---------------|
| Preface | 7 |
| Informatica Resources. | 7 |
| Informatica Network. | 7 |
| Informatica Knowledge Base. | 7 |
| Informatica Documentation. | 7 |
| Informatica Product Availability Matrixes. | 8 |
| Informatica Velocity. | 8 |
| Informatica Marketplace. | 8 |
| Informatica Global Customer Support. | 8 |
| Chapter 1: Introduction to Informatica Analyst..... | 9 |
| Informatica Analyst Overview. | 9 |
| Informatica Analyst Interface. | 10 |
| Informatica Analyst Header. | 10 |
| Informatica Analyst Workspaces. | 11 |
| Keyboard Shortcuts. | 12 |
| Informatica Analyst Assets. | 13 |
| Logging In to the Analyst Tool. | 14 |
| Chapter 2: Library Workspace..... | 15 |
| Library Workspace Overview. | 15 |
| Accessing the Library Workspace. | 16 |
| Library Tasks. | 16 |
| Creating a User-Defined Tag. | 17 |
| Assigning and Removing a Tag. | 17 |
| Chapter 3: Connections Workspace..... | 18 |
| Connections Workspace Overview. | 18 |
| IBM DB2 Connection Properties. | 19 |
| JDBC Connection Properties. | 21 |
| MS SQL Server Connection Properties. | 23 |
| ODBC Connection Properties. | 27 |
| Oracle Connection Properties. | 28 |
| Hive Connection Properties. | 30 |
| HDFS Connection Properties. | 36 |
| Identifier Properties in Database Connections. | 37 |
| Regular Identifiers. | 37 |
| Delimited Identifiers. | 38 |
| Identifier Properties. | 38 |
| Searching for a Database Connection. | 40 |

| | |
|---|-----------|
| Creating a Database Connection. | 40 |
| Editing a Database Connection. | 41 |
| Deleting a Database Connection. | 41 |
| Chapter 4: Job Status Workspace. | 42 |
| Job Status Workspace Overview. | 42 |
| Accessing the Job Status Workspace. | 42 |
| Job Properties. | 43 |
| Monitoring Jobs. | 44 |
| Chapter 5: Projects Workspace. | 45 |
| Projects Workspace Overview. | 45 |
| Accessing the Projects Workspace. | 45 |
| Manage Projects and Folders. | 46 |
| Project Security. | 47 |
| Project Permissions. | 47 |
| Assigning Direct Permissions on a Project. | 48 |
| Viewing Permissions on a Project. | 48 |
| Chapter 6: Model Repository. | 49 |
| Model Repository Overview. | 49 |
| Informatica Analyst Assets. | 49 |
| Repository Asset Locks. | 50 |
| Rules and Guidelines for Asset Lock Management. | 51 |
| Team-based Development with Versioned Objects. | 51 |
| Versioned Asset Management. | 51 |
| Chapter 7: Data Objects. | 53 |
| Data Objects Overview. | 53 |
| Flat File Data Objects. | 54 |
| Import Flat File Data Objects. | 54 |
| Flat File Options. | 54 |
| Flat File Data Types. | 55 |
| Datetime Data Types. | 55 |
| Adding a Delimited Flat File. | 57 |
| Adding a Fixed-width Flat File. | 57 |
| Rules and Guidelines for Flat Files. | 58 |
| Table Data Objects. | 59 |
| Adding a Table. | 59 |
| Rules and Guidelines for Tables. | 60 |
| Synchronize Data Objects. | 60 |
| Synchronizing a Flat File Data Object. | 61 |
| Synchronizing a Relational Data Object. | 61 |

| | |
|--|-----------|
| Viewing Data Objects. | 62 |
| Editing Data Objects. | 62 |
| Chapter 8: Search. | 63 |
| Search Overview. | 63 |
| Search Results. | 63 |
| Search Query. | 64 |
| Search Properties. | 64 |
| Appendix A: Configure the Web Browser | 65 |
| Configure the Web Browser | 65 |
| Index. | 66 |

Preface

The *Informatica Analyst Tool Guide* describes how to use Informatica Analyst (the Analyst tool) to discover and define business logic and collaborate on business projects in an enterprise. It is written for business users such as analysts and data stewards who collaborate on projects within an organization. This guide assumes that you have an understanding of flat file and relational database concepts, and the database engines in your environment.

Informatica Resources

Informatica Network

Informatica Network hosts Informatica Global Customer Support, the Informatica Knowledge Base, and other product resources. To access Informatica Network, visit <https://network.informatica.com>.

As a member, you can:

- Access all of your Informatica resources in one place.
- Search the Knowledge Base for product resources, including documentation, FAQs, and best practices.
- View product availability information.
- Review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

Informatica Knowledge Base

Use the Informatica Knowledge Base to search Informatica Network for product resources such as documentation, how-to articles, best practices, and PAMs.

To access the Knowledge Base, visit <https://kb.informatica.com>. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at KB_Feedback@informatica.com.

Informatica Documentation

To get the latest documentation for your product, browse the Informatica Knowledge Base at https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx.

If you have questions, comments, or ideas about this documentation, contact the Informatica Documentation team through email at infa_documentation@informatica.com.

Informatica Product Availability Matrixes

Product Availability Matrixes (PAMs) indicate the versions of operating systems, databases, and other types of data sources and targets that a product release supports. If you are an Informatica Network member, you can access PAMs at

<https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Velocity

Informatica Velocity is a collection of tips and best practices developed by Informatica Professional Services. Developed from the real-world experience of hundreds of data management projects, Informatica Velocity represents the collective knowledge of our consultants who have worked with organizations from around the world to plan, develop, deploy, and maintain successful data management solutions.

If you are an Informatica Network member, you can access Informatica Velocity resources at

<http://velocity.informatica.com>.

If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Marketplace

The Informatica Marketplace is a forum where you can find solutions that augment, extend, or enhance your Informatica implementations. By leveraging any of the hundreds of solutions from Informatica developers and partners, you can improve your productivity and speed up time to implementation on your projects. You can access Informatica Marketplace at <https://marketplace.informatica.com>.

Informatica Global Customer Support

You can contact a Global Support Center by telephone or through Online Support on Informatica Network.

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:

<http://www.informatica.com/us/services-and-training/support-services/global-support-centers>.

If you are an Informatica Network member, you can use Online Support at <http://network.informatica.com>.

CHAPTER 1

Introduction to Informatica Analyst

This chapter includes the following topics:

- [Informatica Analyst Overview, 9](#)
- [Informatica Analyst Interface, 10](#)
- [Informatica Analyst Assets, 13](#)
- [Logging In to the Analyst Tool, 14](#)

Informatica Analyst Overview

Informatica Analyst (the Analyst tool) is a web-based client tool that is available to multiple Informatica products and is used by business users to collaborate on projects within an organization. For example, business analysts can use the Analyst tool to collaborate on data integration projects in an organization.

Use the Analyst tool to discover, define, and review the business logic for projects in an organization. The tasks that you can perform in the Analyst tool depend on the license for Informatica products and the privileges to perform tasks.

Based on the license that your organization has, you can use the Analyst tool to perform the following tasks:

- Define business glossaries, terms, and policies to maintain standardized definitions of data assets in the organization.
- Perform data discovery to find the content, quality, and structure of data sources, and monitor data quality trends.
- Define data integration logic and collaborate on projects to accelerate project delivery.
- Define and manage rules to verify data conformance to business policies.
- Review and resolve data quality issues to find and fix data quality issues in the organization.

The Analyst Service manages the Analyst tool. The Analyst tool stores projects, folders, and data objects in the Model repository. The Analyst tool connects to the Model repository database to create, update, and delete projects, folders, and data objects.

Informatica Analyst Interface

Use the web-based interface of the Analyst tool to collaborate on business projects within an organization.

The Analyst tool interface has headers and workspaces. A workspace is a web page where you perform tasks based on licensed functionality that you access through tabs in the Analyst tool. You must also have privileges to perform tasks in a workspace.

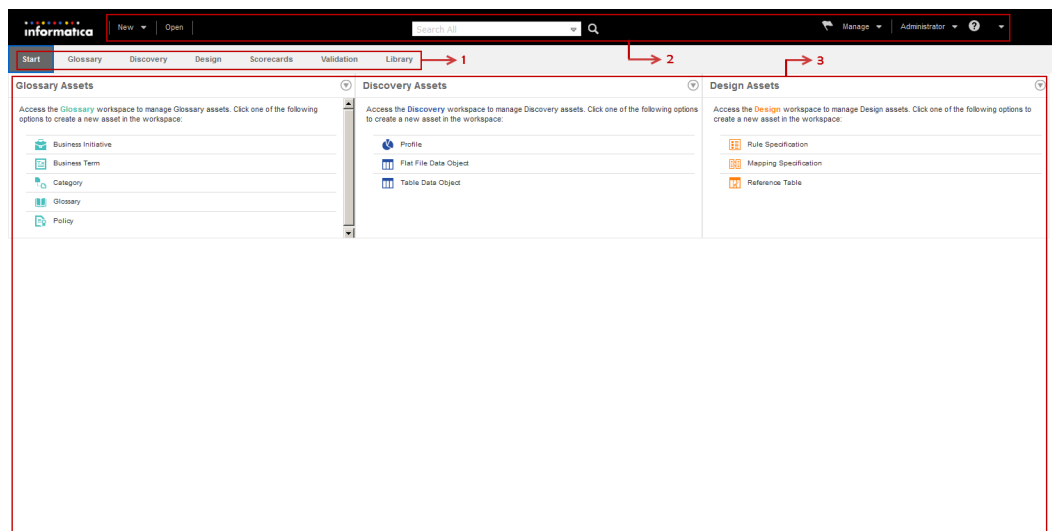
When you log in to the Analyst tool, the **Start** workspace appears. You can open multiple workspaces in the Analyst tool interface.

For example, use the **Discovery** workspace to analyze the quality of data and metadata in source systems. You can access a workspace through workspace tabs or through menus in the Analyst tool header.

You can use assets in some workspaces to perform tasks such as running profiles, creating business rules, or creating mapping specifications. An asset is a type of object in the Analyst tool that supports business operations within an organization.

If you have the license to use the business glossary, you can view notification alerts for business glossary assets. View notification alerts in the Analyst tool header.

The following figure shows the Analyst tool:



1. Workspace access panel
2. Header area
3. Workspace tabs

Informatica Analyst Header

The Analyst tool header appears at the top of the Analyst tool user interface.

The Analyst tool has the following header items:

New

Create assets in the **Glossary**, **Discover**, and **Design** workspaces.

Open

Open the **Library** workspace.

Notifications alert

View notifications for Glossary assets.

Manage

Open temporary workspaces and notifications. You can open the **Connections**, **Data Domains**, **Job Status**, **Projects**, and **Glossary Security** workspaces.

User name

Set user preferences to change the password and to log out of the Analyst tool.

Help

Access help in the current workspace.

Informatica Analyst Workspaces

A workspace is a web page that you can access based on license and privilege. You can perform tasks within a workspace. You can manage assets or use assets to perform tasks in some workspaces. The Analyst tool has permanent workspaces and temporary workspaces.

A permanent workspace is always available on the workspace tab. You can navigate to another workspace, but you cannot close a permanent workspace. A temporary workspace is available through a workspace tab. You can open a temporary workspace from the Analyst tool header or from access panels within a workspace. You can close the workspace from the tab when you do not need it.

Permanent Workspaces

The Analyst tool contains the following permanent workspaces:

Start

Access other workspaces that you have the license to access through access panels on the workspace. If you have the license to perform exception management, your tasks appear on the **My Tasks** panel of the workspace. If you have to cast your vote as an approver in the approval workflow process, your pending tasks appear in the **My Tasks** panel.

Glossary

Define and describe business concepts that are important to your organization. You can create and manage business terms, business initiatives, categories, glossaries, and policies.

Discovery

Analyze the quality of data and metadata in source systems. You can create and manage profiles, flat file data objects, and table data objects. You can view and manage Developer tool objects such as SAP and mainframe objects that are stored in projects in the Model repository.

Design

Design business logic that helps analysts and developers collaborate. You can create and manage mapping specifications, reference tables, and rule specifications.

Scorecards

Open, edit, and run scorecards that you created from profile results. You can add metrics, drill down on columns, add scorecard filters, and view trend charts for a scorecard.

Temporary Workspaces

The Analyst tool contains the following temporary workspaces:

Library

Search for assets in the Model repository. You can also view metadata in the **Library** workspace. When you open an asset, it opens in the workspace where it was created.

Exceptions

View and manage exception record data for a task. When you open a task from the **My Tasks** panel of the **Start** workspace, the Analyst tool opens a temporary workspace called the **Exceptions** workspace. View duplicate record clusters or exception records depending on the type of task you are working on. View an audit trail of the changes you make to records in a task.

Connections

Create and manage connections to import relational data objects, preview data, run a profile, and run mapping specifications.

Data Domains

Create, manage, and remove data domains and data domain groups. A data domain describes the semantics of column data such as Social Security number or phone number. You can categorize data domains into data domain groups such as Social Security number and phone number into the Personal Health Information (PHI) data domain group.

Job Status

Monitor the status of Analyst tool jobs such as data preview for all objects and drilldown operations on profiles.

Projects

Create and manage folders and projects and assign permissions on projects.

Glossary Security

Manage permissions, privileges, and roles for business glossary users.

Settings

Select an operating system profile to run mappings, profiles, and scorecards. An operating system profile is a type of security that the Data Integration Service uses to isolate run-time user environments. The **Settings** workspace is available if the Informatica administrator assigns more than one operating system profile to the user. To open the **Settings** workspace, go to the **User Name** Analyst header item, click **Settings**.

Keyboard Shortcuts

You can use keyboard shortcuts to navigate and work with the Analyst tool interface.

The navigation order of objects is from top to bottom and left to right.

You can perform the following tasks with keyboard shortcuts:

Navigate among elements and select an element.

Press Tab.

Navigate among portlets and panes in a workspace.

Press Alt+P.

Close a temporary workspace.

Press Ctrl+Shift+X.

Informatica Analyst Assets

You can manage assets within some workspaces. An asset is a type of object that you use to support business operations within the enterprise.

For example, a profile is an asset that an analyst can create to discover the content, quality, and structure of a data source.

You can create the following types of assets:

Glossary assets

Create Glossary assets in the **Glossary** workspace. You can create the following types of Glossary assets:

- Business term. A word or phrase that uses business language to define relevant concepts for business users in an organization.
- Business initiative. A business decision that results in bulk changes to a collection of Glossary assets.
- Category. A descriptive classification of business terms and policies.
- Glossary. A collection of categories, business terms, and policies.
- Policy. The business purpose, process, or protocol that governs business practices that are related to business terms.

Discovery assets

Create Discovery assets in the **Discovery** workspace. You can create the following types of Discovery assets:

- Profile. A single data object profile or enterprise discovery profile that discovers column characteristics and data domains in a single data source or multiple data sources.
- Flat file data object. A representation of data based on a flat file.
- Table data object. A representation of data based on a relational table.

Design assets

Create Design assets in the **Design** workspace. You can create the following types of Design assets:

- Mapping specification. A template that describes the movement and transformation of data from a source to a target.
- Reference table. A table that contains the standard version and alternative versions of a set of data values.
- Rule specification. An object that represents the logic in a business rule.

Scorecards assets

Open scorecard assets in the **Scorecards** workspace. A scorecard is a graphical representation of the quality measurements in a profile.

Logging In to the Analyst Tool

Use the Analyst tool URL to log in to the Analyst tool. When you log in to the Analyst tool, you specify an Informatica login name, a password, and the native domain or the LDAP security domain.

1. Start a Microsoft Internet Explorer or Google Chrome browser.
2. In the Address field, enter the URL for the Analyst tool:
`http[s]://<fully qualified host name>:<port number>/analyst/`
3. If the domain uses LDAP or native authentication, enter a login name and a password on the login page.
4. Select **Native** or the name of a specific security domain.
The Security Domain field appears when the Informatica domain uses LDAP or Kerberos authentication. If you do not know the security domain that your user account belongs to, contact the Informatica domain administrator.
5. Click **Log In**.
The Analyst tool opens on the **Start** workspace.

CHAPTER 2

Library Workspace

This chapter includes the following topics:

- [Library Workspace Overview, 15](#)
- [Accessing the Library Workspace, 16](#)
- [Library Tasks, 16](#)

Library Workspace Overview

Use the **Library** workspace to browse, search, or filter a collection of assets that you have the privilege to access. You can browse and view assets, glossaries, projects, and tags.

Begin browsing Model repository content with the Library Navigator. The Library Navigator contains sections based on the different ways that you might want to look up Model repository content. For example, if you want to find all Glossary assets associated with one glossary, you can start browsing from the Glossary section.

You can look up Model repository content in the following ways through the Library Navigator:

- By asset
- By project
- By glossary
- By tags

When you open a section, you can select a type of asset, and view the list of assets in the asset list. You can sort or group the asset list by asset property to organize your assets. You cannot sort the asset list by description.

Use the Library Navigator to perform a Discovery Search. A discovery search finds assets and their relationships with other assets in an organization. You can add filters including search filters to narrow the asset list. You can sort by description of an asset in the search results.

You can open an asset from the asset list. When you click an asset, the asset opens in its corresponding workspace. You can edit the asset, view history, add comments, and view related assets.

Accessing the Library Workspace

Access the **Library** workspace to view and manage the collection of assets that you have the privilege to view or manage.

- From the Analyst tool header, click **Open**.
The **Library** workspace opens.

Library Tasks

You can manage the collection of assets that you have the privilege to access and perform library tasks.

You can perform the following library tasks:

Perform a discovery search.

A discovery search finds assets and their relationships with other assets in an organization. For example, you need to find where all the customer information exists in a financial organization. Perform discovery search to find data objects that meet the search criteria for a customer search string. The discovery search results include assets related to the data object that you searched for in the Analyst tool. The related assets include profiles run against the data object, associated scorecards, and business terms.

For more information, see the *Data Discovery Guide*.

View assets.

When you look up Model repository content from a section in the Library Navigator, the Analyst tool displays the list of assets in an asset list. For example, when you select **Data Objects**, the Analyst tool displays a list of data objects that you have the privilege to access.

From the Library Navigator, click the **Assets** section and select an asset. You can view the list of assets that belong to the asset list.

View glossaries.

View glossaries that you have the privilege to access. When you select a glossary from the **Glossary** section, the Analyst tool lists the contents of the glossary such as business terms, categories, or policies in the asset list.

From the Library Navigator, click the **Glossaries** section and select a glossary. You can view glossary assets in asset list.

View projects.

View projects and folders and their contents. When you select a project or folder, the Analyst tool lists the contents of the project or folder in the asset list.

From the Library Navigator, click the **Projects** section and select a project or folder. You can view the project or folder contents in the **Assets** panel.

View, add, or remove tags.

View tagged business terms by system-defined tags or view assets by user-defined tags. System-defined tags group business terms according to their usage. You can create tags from the **Tags** section. You can assign tags or remove tags from assets from the **Projects** section.

Search for assets.

Search for assets by a search string or apply filters to search for assets. Enter the filter properties to filter the search results.

From the **Filter** panel, enter a search string in the search box or add filter properties to filter the search results.

Creating a User-Defined Tag

Create a user-defined tag to group an asset according to business usage.

1. From the **Tags** section in the Library Navigator, right-click **User-Defined** and choose **New Tag**.
The **New Tag** dialog box appears.
2. Enter a name and an optional description.
3. Click **OK**.

Assigning and Removing a Tag

Assign a tag to an asset to group an asset according to business usage. You can also remove a tag from an asset when it is no longer needed.

1. From the **Projects** section in the Library Navigator, select a project.
2. From the asset list, right-click an asset and select **Tag**.
The **Tag** dialog box appears.
3. Choose to add or remove a tag.
 - To add a tag, enter a user-defined tag name in the **New Tag** panel, and click **Add**.
 - To remove a tag, select a tag from the **Tags** panel and click **Remove**.
4. Click **OK**.

CHAPTER 3

Connections Workspace

This chapter includes the following topics:

- [Connections Workspace Overview, 18](#)
- [IBM DB2 Connection Properties, 19](#)
- [JDBC Connection Properties, 21](#)
- [MS SQL Server Connection Properties, 23](#)
- [ODBC Connection Properties, 27](#)
- [Oracle Connection Properties, 28](#)
- [Hive Connection Properties, 30](#)
- [HDFS Connection Properties, 36](#)
- [Identifier Properties in Database Connections, 37](#)
- [Searching for a Database Connection, 40](#)
- [Creating a Database Connection, 40](#)
- [Editing a Database Connection, 41](#)
- [Deleting a Database Connection, 41](#)

Connections Workspace Overview

Use the **Connections** workspace to view, create, and manage connections. A connection is a repository object that defines a connection in the domain configuration repository.

Create a connection to import data objects, preview data, and run profiles or mappings. The Analyst tool uses the connection when you import a data object. The Data Integration Service uses the connection when you preview data, run a profile, or run a mapping.

The Analyst tool stores connections in the domain configuration repository. Any connection that you create in the Analyst tool is available in the Developer tool or the Administrator tool.

You can perform the following tasks in the **Connections** workspace:

- Search for a connection.
- Create a connection.
- Test a connection.
- Edit a connection.
- Delete a connection.

You can create the following types of connections in the Analyst tool:

- IBM DB2
- Microsoft SQL Server
- ODBC
- Oracle
- Hive
- Hadoop File System

You can browse and import tables from IBM DB2/zOS connections. However, you must create IBM DB2/zOS connections in the Administrator tool or the Developer tool.

IBM DB2 Connection Properties

Use an IBM DB2 connection to access IBM DB2. An IBM DB2 connection is a relational database connection. You can create and manage an IBM DB2 connection in the Administrator tool, the Developer tool, or the Analyst tool.

Note: The order of the connection properties might vary depending on the tool where you view them.

The following table describes DB2 connection properties:

| Property | Description |
|-----------------------------------|--|
| Database Type | The database type. |
| Name | Name of the connection. The name is not case sensitive and must be unique within the domain. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * () - + = { [] } \ : ; " ' < , > . ? / |
| ID | String that the Data Integration Service uses to identify the connection. The ID is not case sensitive. It must be 255 characters or less and must be unique in the domain. You cannot change this property after you create the connection. Default value is the connection name. |
| Description | The description of the connection. The description cannot exceed 765 characters. |
| User Name | The database user name. |
| Password | The password for the database user name. |
| Pass-through security enabled | Enables pass-through security for the connection. When you enable pass-through security for a connection, the domain uses the client user name and password to log into the corresponding database, instead of the credentials defined in the connection object. |
| Connection String for data access | The DB2 connection URL used to access metadata from the database. dbname Where <code>dbname</code> is the alias configured in the DB2 client. |

| Property | Description |
|---|--|
| Metadata Access Properties: Connection String | <p>Use the following metadata connection string URL:</p> <pre>jdbc:informatica:db2://<host name>:<port>;DatabaseName=<database name></pre> <p>When you import a table, by default, all tables are displayed under the default schema name. To view tables under a specific schema instead of the default schema, you can specify the schema name from which you want to import the table. Include the ischemaname parameter in the URL to specify the schema name. For example, use the following syntax to import a table from a specific schema:</p> <pre>jdbc:informatica:db2://<host name>:<port>;DatabaseName=<database name>;ischemaname=<schema_name></pre> <p>To search for a table in multiple schemas and import it, you can specify multiple schema names in the ischemaname parameter. The schema name is case sensitive. You cannot use special characters when you specify multiple schema names. Use the pipe () character to separate multiple schema names. For example, use the following syntax to search for a table in three schemas and import it:</p> <pre>jdbc:informatica:db2://<host name>:<port>;DatabaseName=<database name>;ischemaname=<schema_name1> <schema_name2> <schema_name3></pre> <p>When you specify multiple schema names, you must clear the Show Default Schema Only option to view the tables under the specified schema names.</p> |
| AdvancedJDBCSecurityOptions | <p>Database parameters for metadata access to a secure database. Informatica treats the value of the AdvancedJDBCSecurityOptions field as sensitive data and stores the parameter string encrypted.</p> <p>To connect to a secure database, include the following parameters:</p> <ul style="list-style-type: none"> - EncryptionMethod. Required. Indicates whether data is encrypted when transmitted over the network. This parameter must be set to SSL. - ValidateServerCertificate. Optional. Indicates whether Informatica validates the certificate that is sent by the database server. <p>If this parameter is set to True, Informatica validates the certificate that is sent by the database server. If you specify the HostNameInCertificate parameter, Informatica also validates the host name in the certificate.</p> <p>If this parameter is set to false, Informatica does not validate the certificate that is sent by the database server. Informatica ignores any truststore information that you specify.</p> <ul style="list-style-type: none"> - HostNameInCertificate. Optional. Host name of the machine that hosts the secure database. If you specify a host name, Informatica validates the host name included in the connection string against the host name in the SSL certificate. - cryptoProtocolVersion. Optional. If you enable TLS for the IBM DB2 instance, set the cryptoProtocolVersion parameter as follows: cryptoProtocolVersion=TLSv<version number>. For example, cryptoProtocolVersion=TLSv1.2 <p>Note: The version number must be the same as the TLS version you configured for the server.</p> <ul style="list-style-type: none"> - TrustStore. Required. Path and file name of the truststore file. - TrustStorePassword. Required. Password for the truststore file for the secure database. <p>Note: Informatica appends the secure JDBC parameters to the connection string. If you include the secure JDBC parameters directly to the connection string, do not enter any parameters in the AdvancedJDBCSecurityOptions field.</p> |
| Data Access Properties: Connection String | <p>The connection string used to access data from the database.</p> <p>For IBM DB2 this is <database name></p> |

| Property | Description |
|--------------------------------|--|
| Code Page | The code page used to read from a source database or to write to a target database or file. |
| Environment SQL | SQL commands to set the database environment when you connect to the database. The Data Integration Service runs the connection environment SQL each time it connects to the database. |
| Transaction SQL | SQL commands to set the database environment when you connect to the database. The Data Integration Service runs the transaction environment SQL at the beginning of each transaction. |
| Retry Period | This property is reserved for future use. |
| Tablespace | The tablespace name of the database. |
| SQL Identifier Character | Type of character that the database uses to enclose delimited identifiers in SQL queries. The available characters depend on the database type. Select (None) if the database uses regular identifiers. When the Data Integration Service generates SQL queries, the service does not place delimited characters around any identifiers. Select a character if the database uses delimited identifiers. When the Data Integration Service generates SQL queries, the service encloses delimited identifiers within this character. |
| Support Mixed-case Identifiers | Enable if the database uses case-sensitive identifiers. When enabled, the Data Integration Service encloses all identifiers within the character selected for the SQL Identifier Character property. When the SQL Identifier Character property is set to none, the Support Mixed-case Identifiers property is disabled. |
| ODBC Provider | ODBC. The type of database to which ODBC connects. For pushdown optimization, specify the database type to enable the Data Integration Service to generate native database SQL. The options are: <ul style="list-style-type: none"> - Other - Sybase - Microsoft_SQL_Server Default is Other. |

JDBC Connection Properties

You can use a JDBC connection to access tables in a database. You can create and manage a JDBC connection in the Administrator tool, the Developer tool, or the Analyst tool.

Note: The order of the connection properties might vary depending on the tool where you view them.

The following table describes JDBC connection properties:

| Property | Description |
|--------------------------|---|
| Database Type | The database type. |
| Name | Name of the connection. The name is not case sensitive and must be unique within the domain. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * () - + = { [] } \ : ; " ' < , > . ? / |
| ID | String that the Data Integration Service uses to identify the connection. The ID is not case sensitive. It must be 255 characters or less and must be unique in the domain. You cannot change this property after you create the connection. Default value is the connection name. |
| Description | The description of the connection. The description cannot exceed 765 characters. |
| User Name | The database user name. |
| Password | The password for the database user name. |
| JDBC Driver Class Name | <p>Name of the JDBC driver class.</p> <p>The following list provides the driver class name that you can enter for the applicable database type:</p> <ul style="list-style-type: none"> - DataDirect JDBC driver class name for Oracle: <code>com.informatica.jdbc.oracle.OracleDriver</code> - DataDirect JDBC driver class name for IBM DB2: <code>com.informatica.jdbc.db2.DB2Driver</code> - DataDirect JDBC driver class name for Microsoft SQL Server: <code>com.informatica.jdbc.sqlserver.SQLServerDriver</code> - DataDirect JDBC driver class name for Sybase ASE: <code>com.informatica.jdbc.sybase.SybaseDriver</code> - DataDirect JDBC driver class name for Informix: <code>com.informatica.jdbc.informix.InformixDriver</code> - DataDirect JDBC driver class name for MySQL: <code>com.informatica.jdbc.mysql.MySQLDriver</code> <p>For more information about which driver class to use with specific databases, see the vendor documentation.</p> |
| Connection String | <p>Connection string to connect to the database. Use the following connection string:</p> <pre>jdbc:<subprotocol>:<subname></pre> |
| Environment SQL | <p>Optional. Enter SQL commands to set the database environment when you connect to the database. The Data Integration Service executes the connection environment SQL each time it connects to the database.</p> <p>Note: If you configure Sqoop, the Sqoop program does not honor this property.</p> |
| Transaction SQL | <p>Optional. Enter SQL commands to set the database environment when you connect to the database. The Data Integration Service executes the transaction environment SQL at the beginning of each transaction.</p> <p>Note: If you configure Sqoop, the Sqoop program does not honor this property.</p> |
| SQL Identifier Character | <p>Type of character that the database uses to enclose delimited identifiers in SQL queries. The available characters depend on the database type.</p> <p>Select (None) if the database uses regular identifiers. When the Data Integration Service generates SQL queries, the service does not place delimited characters around any identifiers.</p> <p>Select a character if the database uses delimited identifiers. When the Data Integration Service generates SQL queries, the service encloses delimited identifiers within this character.</p> <p>Note: If you configure Sqoop, the Sqoop program does not honor this property.</p> |

| Property | Description |
|--------------------------------|--|
| Support Mixed-case Identifiers | <p>Enable if the database uses case-sensitive identifiers. When enabled, the Data Integration Service encloses all identifiers within the character selected for the SQL Identifier Character property.</p> <p>When the SQL Identifier Character property is set to none, the Support Mixed-case Identifiers property is disabled.</p> <p>Note: If you configure Sqoop, the Sqoop program honors this property when you generate and execute a DDL script to create or replace a target at run time. In all other scenarios, the Sqoop program ignores this property.</p> |
| Use Sqoop Connector | <p>Enables Sqoop connectivity for the data object that uses the JDBC connection. The Data Integration Service runs the mapping in the Hadoop run-time environment through Sqoop.</p> <p>You can configure Sqoop connectivity for relational data objects, customized data objects, and logical data objects that are based on a JDBC-compliant database.</p> <p>Select Sqoop v1.x to enable Sqoop connectivity.</p> <p>Default is None.</p> |
| Sqoop Arguments | <p>Enter the arguments that the Sqoop program must use to connect to the database. Separate multiple arguments with a space.</p> <p>To read data from or write data to Teradata through Teradata Connector for Hadoop (TDCH) specialized connectors for Sqoop, define the TDCH connection factory class in the Sqoop arguments. The connection factory class varies based on the TDCH Sqoop Connector that you want to use.</p> <ul style="list-style-type: none"> - To use the Cloudera Connector Powered by Teradata, configure the following Sqoop argument: <pre>- Dsqaop.connection.factories=com.cloudera.connector.teradata.TeradataManagerFactory</pre> - To use the Hortonworks Connector for Teradata (powered by the Teradata Connector for Hadoop), configure the following Sqoop argument: <pre>- Dsqaop.connection.factories=org.apache.sqaop.teradata.TeradataManagerFactory</pre> <p>If you do not enter Sqoop arguments, the Data Integration Service constructs the Sqoop command based on the JDBC connection properties.</p> |

MS SQL Server Connection Properties

Use a Microsoft SQL Server connection to access Microsoft SQL Server. A Microsoft SQL Server connection is a connection to a Microsoft SQL Server relational database. You can create and manage a Microsoft SQL Server connection in the Administrator tool or the Developer tool.

Note: The order of the connection properties might vary depending on the tool where you view them.

The following table describes MS SQL Server connection properties:

| Property | Description |
|---|--|
| Database Type | The database type. |
| Name | Name of the connection. The name is not case sensitive and must be unique within the domain. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * () - + = { [] } \ : ; " ' < , > . ? / |
| ID | String that the Data Integration Service uses to identify the connection. The ID is not case sensitive. It must be 255 characters or less and must be unique in the domain. You cannot change this property after you create the connection. Default value is the connection name. |
| Description | The description of the connection. The description cannot exceed 765 characters. |
| Use trusted connection | Enables the application service to use Windows authentication to access the database. The user name that starts the application service must be a valid Windows user with access to the database. By default, this option is cleared. |
| User Name | The database user name. Required if Microsoft SQL Server uses NTLMv1 or NTLMv2 authentication. |
| Password | The password for the database user name. Required if Microsoft SQL Server uses NTLMv1 or NTLMv2 authentication. |
| Pass-through security enabled | Enables pass-through security for the connection. When you enable pass-through security for a connection, the domain uses the client user name and password to log into the corresponding database, instead of the credentials defined in the connection object. |
| Metadata Access Properties: Connection String | <p>Connection string used to access metadata from the database.</p> <p>Use the following connection string:</p> <pre>jdbc:informatica:sqlserver://<host name>:<port>;DatabaseName=<database name></pre> <p>To test the connection with NTLM authentication, include the following parameters in the connection string:</p> <ul style="list-style-type: none"> - AuthenticationMethod. The NTLM authentication version to use. <p>Note: UNIX supports NTLMv1 and NTLMv2 but not NTLM.</p> <ul style="list-style-type: none"> - Domain. The domain that the SQL server belongs to. <p>The following example shows the connection string for a SQL server that uses NTLMv2 authentication in a NT domain named Informatica.com:</p> <pre>jdbc:informatica:sqlserver://host01:1433;DatabaseName=SQL1;AuthenticationMethod=ntlm2java;Domain=Informatica.com</pre> <p>If you connect with NTLM authentication, you can enable the Use trusted connection option in the MS SQL Server connection properties. If you connect with NTLMv1 or NTLMv2 authentication, you must provide the user name and password in the connection properties.</p> |

| Property | Description |
|---------------------------------------|---|
| AdvancedJDBCSecurityOptions | <p>Database parameters for metadata access to a secure database. Informatica treats the value of the AdvancedJDBCSecurityOptions field as sensitive data and stores the parameter string encrypted.</p> <p>To connect to a secure database, include the following parameters:</p> <ul style="list-style-type: none"> - EncryptionMethod. Required. Indicates whether data is encrypted when transmitted over the network. This parameter must be set to SSL. - ValidateServerCertificate. Optional. Indicates whether Informatica validates the certificate that is sent by the database server. <p>If this parameter is set to True, Informatica validates the certificate that is sent by the database server. If you specify the HostNameInCertificate parameter, Informatica also validates the host name in the certificate.</p> <p>If this parameter is set to false, Informatica does not validate the certificate that is sent by the database server. Informatica ignores any truststore information that you specify.</p> <ul style="list-style-type: none"> - HostNameInCertificate. Optional. Host name of the machine that hosts the secure database. If you specify a host name, Informatica validates the host name included in the connection string against the host name in the SSL certificate. - cryptoProtocolVersion. Optional. If you enable TLS for the Microsoft SQL Server instance, set the cryptoProtocolVersion parameter as follows: cryptoProtocolVersion=TLSv<version number>. For example, cryptoProtocolVersion=TLSv1.2 <p>Note: The version number must be the same as the TLS version you configured for the server.</p> <ul style="list-style-type: none"> - TrustStore. Required. Path and file name of the truststore file. - TrustStorePassword. Required. Password for the truststore file for the secure database. <p>Not applicable for ODBC.</p> <p>Note: Informatica appends the secure JDBC parameters to the connection string. If you include the secure JDBC parameters directly to the connection string, do not enter any parameters in the AdvancedJDBCSecurityOptions field.</p> |
| Data Access Properties: Provider Type | <p>The connection provider that you want to use to connect to the Microsoft SQL Server database.</p> <p>You can select the following provider types:</p> <ul style="list-style-type: none"> - ODBC - Oldeb(Deprecated) <p>Default is ODBC.</p> <p>OleDb is a deprecated provider type. Support for the OleDb provider type will be dropped in a future release.</p> |
| Use DSN | <p>Enables the Data Integration Service to use the Data Source Name for the connection.</p> <p>If you select the Use DSN option, the Data Integration Service retrieves the database and server names from the DSN.</p> <p>If you do not select the Use DSN option, you must provide the database and server names.</p> |
| Connection String | <p>Use the following connection string if you do not enable DSN mode:</p> <pre><server name>@<database name></pre> <p>If you enable DSN mode, use the following connection strings:</p> <pre><DSN Name></pre> |

| Property | Description |
|--------------------------------|---|
| Code Page | The code page used to read from a source database or to write to a target database or file. |
| Domain Name | The name of the domain. |
| Packet Size | The packet size used to transmit data. Used to optimize the native drivers for Microsoft SQL Server. |
| Owner Name | The name of the owner of the schema. Note: When you generate a table DDL through a dynamic mapping or through the Generate and Execute DDL option, the DDL metadata does not include schema name and owner name properties. |
| Schema Name | The name of the schema in the database. You must specify the schema name for the Profiling Warehouse if the schema name is different from the database user name. You must specify the schema name for the data object cache database if the schema name is different from the database user name and if you configure user-managed cache tables. Note: When you generate a table DDL through a dynamic mapping or through the Generate and Execute DDL option, the DDL metadata does not include schema name and owner name properties. |
| Environment SQL | SQL commands to set the database environment when you connect to the database. The Data Integration Service runs the connection environment SQL each time it connects to the database. |
| Transaction SQL | SQL commands to set the database environment when you connect to the database. The Data Integration Service runs the transaction environment SQL at the beginning of each transaction. |
| Retry Period | This property is reserved for future use. |
| SQL Identifier Character | Type of character that the database uses to enclose delimited identifiers in SQL queries. The available characters depend on the database type. Select (None) if the database uses regular identifiers. When the Data Integration Service generates SQL queries, the service does not place delimited characters around any identifiers. Select a character if the database uses delimited identifiers. When the Data Integration Service generates SQL queries, the service encloses delimited identifiers within this character. |
| Support Mixed-case Identifiers | Enable if the database uses case-sensitive identifiers. When enabled, the Data Integration Service encloses all identifiers within the character selected for the SQL Identifier Character property. When the SQL Identifier Character property is set to none, the Support Mixed-case Identifiers property is disabled. |
| ODBC Provider | ODBC. The type of database to which ODBC connects. For pushdown optimization, specify the database type to enable the Data Integration Service to generate native database SQL. The options are: <ul style="list-style-type: none"> - Other - Sybase - Microsoft_SQL_Server Default is Other. |

ODBC Connection Properties

Use an ODBC connection to access ODBC data. An ODBC connection is a relational database connection. You can create and manage an ODBC connection in the Administrator tool, the Developer tool, or the Analyst tool.

Note: The order of the connection properties might vary depending on the tool where you view them.

The following table describes ODBC connection properties:

| Property | Description |
|---|--|
| Database Type | The database type. |
| Name | Name of the connection. The name is not case sensitive and must be unique within the domain. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * () - + = { [] } \ : ; " ' < , > . ? / |
| ID | String that the Data Integration Service uses to identify the connection. The ID is not case sensitive. It must be 255 characters or less and must be unique in the domain. You cannot change this property after you create the connection. Default value is the connection name. |
| Description | The description of the connection. The description cannot exceed 765 characters. |
| User Name | The database user name. |
| Password | The password for the database user name. |
| Pass-through security enabled | Enables pass-through security for the connection. When you enable pass-through security for a connection, the domain uses the client user name and password to log into the corresponding database, instead of the credentials defined in the connection object. |
| Data Access Properties: Connection String | The ODBC connection URL used to access metadata from the database. <data source name> |
| Code Page | The code page used to read from a source database or to write to a target database or file. |
| Environment SQL | SQL commands to set the database environment when you connect to the database. The Data Integration Service runs the connection environment SQL each time it connects to the database. |
| Transaction SQL | SQL commands to set the database environment when you connect to the database. The Data Integration Service runs the transaction environment SQL at the beginning of each transaction. |
| Retry Period | This property is reserved for future use. |

| Property | Description |
|--------------------------------|---|
| SQL Identifier Character | <p>Type of character that the database uses to enclose delimited identifiers in SQL queries. The available characters depend on the database type.</p> <p>Select (None) if the database uses regular identifiers. When the Data Integration Service generates SQL queries, the service does not place delimited characters around any identifiers.</p> <p>Select a character if the database uses delimited identifiers. When the Data Integration Service generates SQL queries, the service encloses delimited identifiers within this character.</p> |
| Support Mixed-case Identifiers | <p>Enable if the database uses case-sensitive identifiers. When enabled, the Data Integration Service encloses all identifiers within the character selected for the SQL Identifier Character property.</p> <p>When the SQL Identifier Character property is set to none, the Support Mixed-case Identifiers property is disabled.</p> |
| ODBC Provider | <p>The type of database to which ODBC connects. For pushdown optimization, specify the database type to enable the Data Integration Service to generate native database SQL. The options are:</p> <ul style="list-style-type: none"> - Other - Sybase - Microsoft_SQL_Server <p>Default is Other.</p> |

Note: Use an ODBC connection to connect to Microsoft SQL Server when the Data Integration Service runs on UNIX or Linux. Use a native connection to Microsoft SQL Server when the Data Integration Service runs on Windows.

Oracle Connection Properties

Use an Oracle connection to connect to an Oracle database. The Oracle connection is a relational connection type. You can create and manage an Oracle connection in the Administrator tool, the Developer tool, or the Analyst tool.

Note: The order of the connection properties might vary depending on the tool where you view them.

The following table describes Oracle connection properties:

| Property | Description |
|---------------|---|
| Database Type | The database type. |
| Name | <p>Name of the connection. The name is not case sensitive and must be unique within the domain. The name cannot exceed 128 characters, contain spaces, or contain the following special characters:</p> <p>~ ` ! \$ % ^ & * () - + = { [] } \ : ; " ' < , > . ? /</p> |
| ID | <p>String that the Data Integration Service uses to identify the connection. The ID is not case sensitive. It must be 255 characters or less and must be unique in the domain. You cannot change this property after you create the connection. Default value is the connection name.</p> |

| Property | Description |
|---|--|
| Description | The description of the connection. The description cannot exceed 765 characters. |
| User Name | The database user name. |
| Password | The password for the database user name. |
| Pass-through security enabled | Enables pass-through security for the connection. When you enable pass-through security for a connection, the domain uses the client user name and password to log into the corresponding database, instead of the credentials defined in the connection object. |
| Metadata Access Properties: Connection String | <p>Connection string used to access metadata from the database.</p> <p>Use the following connection string:</p> <pre>jdbc:informatica:oracle://<host_name>:<port>;SID=<database name></pre> |
| AdvancedJDBCSecurityOptions | <p>Database parameters for metadata access to a secure database. Informatica treats the value of the AdvancedJDBCSecurityOptions field as sensitive data and stores the parameter string encrypted.</p> <p>To connect to a secure database, include the following parameters:</p> <ul style="list-style-type: none"> - EncryptionMethod. Required. Indicates whether data is encrypted when transmitted over the network. This parameter must be set to SSL. - ValidateServerCertificate. Optional. Indicates whether Informatica validates the certificate that is sent by the database server. <p>If this parameter is set to True, Informatica validates the certificate that is sent by the database server. If you specify the HostNameInCertificate parameter, Informatica also validates the host name in the certificate.</p> <p>If this parameter is set to false, Informatica does not validate the certificate that is sent by the database server. Informatica ignores any truststore information that you specify.</p> <ul style="list-style-type: none"> - HostNameInCertificate. Optional. Host name of the machine that hosts the secure database. If you specify a host name, Informatica validates the host name included in the connection string against the host name in the SSL certificate. - cryptoProtocolVersion. Optional. If you enable TLS for the Oracle instance, set the cryptoProtocolVersion parameter as follows: cryptoProtocolVersion=TLSv<version number>. For example, cryptoProtocolVersion=TLSv1.2 <p>Note: The version number must be the same as the TLS version you configured for the server.</p> <ul style="list-style-type: none"> - TrustStore. Required. Path and file name of the truststore file. - TrustStorePassword. Required. Password for the truststore file for the secure database. - KeyStore. Required. Path and file name of the keystore file. - KeyStorePassword. Required. Password for the keystore file for the secure database. <p>Note: Informatica appends the secure JDBC parameters to the connection string. If you include the secure JDBC parameters directly to the connection string, do not enter any parameters in the AdvancedJDBCSecurityOptions field.</p> |
| Data Access Properties: Connection String | <p>Use the following connection string:</p> <pre><database name>.world</pre> |

| Property | Description |
|--------------------------------|--|
| Code Page | The code page used to read from a source database or to write to a target database or file. |
| Environment SQL | SQL commands to set the database environment when you connect to the database. The Data Integration Service runs the connection environment SQL each time it connects to the database. |
| Transaction SQL | SQL commands to set the database environment when you connect to the database. The Data Integration Service runs the transaction environment SQL at the beginning of each transaction. |
| Retry Period | This property is reserved for future use. |
| Enable Parallel Mode | Enables parallel processing when loading data into a table in bulk mode. By default, this option is cleared. |
| SQL Identifier Character | Type of character that the database uses to enclose delimited identifiers in SQL queries. The available characters depend on the database type. Select (None) if the database uses regular identifiers. When the Data Integration Service generates SQL queries, the service does not place delimited characters around any identifiers. Select a character if the database uses delimited identifiers. When the Data Integration Service generates SQL queries, the service encloses delimited identifiers within this character. |
| Support Mixed-case Identifiers | Enable if the database uses case-sensitive identifiers. When enabled, the Data Integration Service encloses all identifiers within the character selected for the SQL Identifier Character property. When the SQL Identifier Character property is set to none, the Support Mixed-case Identifiers property is disabled. |

Hive Connection Properties

Use the Hive connection to access Hive data. A Hive connection is a database type connection. You can create and manage a Hive connection in the Administrator tool, Analyst tool, or the Developer tool. Hive connection properties are case sensitive unless otherwise noted.

Note: The order of the connection properties might vary depending on the tool where you view them.

The following table describes Hive connection properties:

| Property | Description |
|------------------|--|
| Name | The name of the connection. The name is not case sensitive and must be unique within the domain. You can change this property after you create the connection. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * () - + = { [] } \ : ; " ' < , > . ? / |
| ID | String that the Data Integration Service uses to identify the connection. The ID is not case sensitive. It must be 255 characters or less and must be unique in the domain. You cannot change this property after you create the connection. Default value is the connection name. |
| Description | The description of the connection. The description cannot exceed 4000 characters. |
| Location | The domain where you want to create the connection. Not valid for the Analyst tool. |
| Type | The connection type. Select Hive. |
| Connection Modes | Hive connection mode. Select at least one of the following options: <ul style="list-style-type: none"> - Access Hive as a source or target. Select this option if you want to use Hive as a source or a target. - Use Hive to run mappings in Hadoop cluster. Select this option if you want to use the Hive driver to run mappings in the Hadoop cluster. |

| Property | Description |
|--|---|
| User Name | <p>User name of the user that the Data Integration Service impersonates to run mappings on a Hadoop cluster. The user name depends on the JDBC connection string that you specify in the Metadata Connection String or Data Access Connection String for the native environment.</p> <p>If the Hadoop cluster runs Hortonworks HDP, you must provide a user name. If you use Tez to run mappings, you must provide the user account for the Data Integration Service. If you do not use Tez to run mappings, you can use an impersonation user account.</p> <p>If the Hadoop cluster uses Kerberos authentication, the principal name for the JDBC connection string and the user name must be the same. Otherwise, the user name depends on the behavior of the JDBC driver. With Hive JDBC driver, you can specify a user name in many ways and the user name can become a part of the JDBC URL.</p> <p>If the Hadoop cluster does not use Kerberos authentication, the user name depends on the behavior of the JDBC driver.</p> <p>If you do not specify a user name, the Hadoop cluster authenticates jobs based on the following criteria:</p> <ul style="list-style-type: none"> - The Hadoop cluster does not use Kerberos authentication. It authenticates jobs based on the operating system profile user name of the machine that runs the Data Integration Service. - The Hadoop cluster uses Kerberos authentication. It authenticates jobs based on the SPN of the Data Integration Service. |
| Common Attributes to Both the Modes: Environment SQL | <p>SQL commands to set the Hadoop environment. In native environment type, the Data Integration Service executes the environment SQL each time it creates a connection to a Hive metastore. If you use the Hive connection to run profiles in the Hadoop cluster, the Data Integration Service executes the environment SQL at the beginning of each Hive session.</p> <p>The following rules and guidelines apply to the usage of environment SQL in both connection modes:</p> <ul style="list-style-type: none"> - Use the environment SQL to specify Hive queries. - Use the environment SQL to set the classpath for Hive user-defined functions and then use environment SQL or PreSQL to specify the Hive user-defined functions. You cannot use PreSQL in the data object properties to specify the classpath. The path must be the fully qualified path to the JAR files used for user-defined functions. Set the parameter hive.aux.jars.path with all the entries in infapdo.aux.jars.path and the path to the JAR files for user-defined functions. - You can use environment SQL to define Hadoop or Hive parameters that you want to use in the PreSQL commands or in custom queries. - If you use multiple values for the Environment SQL property, ensure that there is no space between the values. The following sample text shows two values that can be used for the Environment SQL: <pre>set hive.execution.engine='tez';set hive.exec.dynamic.partition.mode='nonstrict';</pre> <p>If you use the Hive connection to run profiles in the Hadoop cluster, the Data Integration service executes only the environment SQL of the Hive connection. If the Hive sources and targets are on different clusters, the Data Integration Service does not execute the different environment SQL commands for the connections of the Hive source or target.</p> |

Properties to Access Hive as Source or Target

The following table describes the connection properties that you configure to access Hive as a source or target:

| Property | Description |
|--|---|
| Metadata Connection String | <p>The JDBC connection URI used to access the metadata from the Hadoop server.</p> <p>You can use PowerExchange for Hive to communicate with a HiveServer service or HiveServer2 service.</p> <p>To connect to HiveServer, specify the connection string in the following format:</p> <pre>jdbc:hive2://<hostname>:<port>/<db></pre> <p>Where</p> <ul style="list-style-type: none">- <hostname> is name or IP address of the machine on which HiveServer2 runs.- <port> is the port number on which HiveServer2 listens.- <db> is the database name to which you want to connect. If you do not provide the database name, the Data Integration Service uses the default database details. <p>To connect to HiveServer 2, use the connection string format that Apache Hive implements for that specific Hadoop Distribution. For more information about Apache Hive connection string formats, see the Apache Hive documentation.</p> |
| Bypass Hive JDBC Server | <p>JDBC driver mode. Select the check box to use the embedded JDBC driver mode.</p> <p>To use the JDBC embedded mode, perform the following tasks:</p> <ul style="list-style-type: none">- Verify that Hive client and Informatica services are installed on the same machine.- Configure the Hive connection properties to run mappings in the Hadoop cluster. <p>If you choose the non-embedded mode, you must configure the Data Access Connection String. Informatica recommends that you use the JDBC embedded mode.</p> |
| Observe Fine Grained SQL Authorization | <p>When you select the option to observe fine-grained SQL authentication in a Hive source, the mapping observes row and column-level restrictions on data access. If you do not select the option, the Blaze run-time engine ignores the restrictions, and results include restricted data.</p> |
| Data Access Connection String | <p>The connection string to access data from the Hadoop data store.</p> <p>To connect to HiveServer, specify the non-embedded JDBC mode connection string in the following format:</p> <pre>jdbc:hive2://<hostname>:<port>/<db></pre> <p>Where</p> <ul style="list-style-type: none">- <hostname> is name or IP address of the machine on which HiveServer2 runs.- <port> is the port number on which HiveServer2 listens.- <db> is the database to which you want to connect. If you do not provide the database name, the Data Integration Service uses the default database details. <p>To connect to HiveServer 2, use the connection string format that Apache Hive implements for the specific Hadoop Distribution. For more information about Apache Hive connection string formats, see the Apache Hive documentation.</p> |

Properties to Run Mappings in Hadoop Cluster

The following table describes the Hive connection properties that you configure when you want to use the Hive connection to run Informatica mappings in the Hadoop cluster:

| Property | Description |
|--------------------------------------|--|
| Database Name | Namespace for tables. Use the name <code>default</code> for tables that do not have a specified database name. |
| Default FS URI | <p>The URI to access the default Hadoop Distributed File System.</p> <p>Use the following connection URI:</p> <pre>hdfs://<node name>:<port></pre> <p>Where</p> <ul style="list-style-type: none">- <code><node name></code> is the host name or IP address of the NameNode.- <code><port></code> is the port on which the NameNode listens for remote procedure calls (RPC). <p>If the Hadoop cluster runs MapR, use the following URI to access the MapR File system: <code>maprfs:///</code>.</p> |
| JobTracker/Yarn Resource Manager URI | <p>The service within Hadoop that submits the MapReduce tasks to specific nodes in the cluster.</p> <p>Use the following format:</p> <pre><hostname>:<port></pre> <p>Where</p> <ul style="list-style-type: none">- <code><hostname></code> is the host name or IP address of the JobTracker or Yarn resource manager.- <code><port></code> is the port on which the JobTracker or Yarn resource manager listens for remote procedure calls (RPC). <p>If the cluster uses MapR with YARN, use the value specified in the <code>yarn.resourcemanager.address</code> property in <code>yarn-site.xml</code>. You can find <code>yarn-site.xml</code> in the following directory on the NameNode of the cluster: <code>/opt/mapr/hadoop/hadoop-2.5.1/etc/hadoop</code>.</p> <p>MapR with MapReduce 1 supports a highly available JobTracker. If you are using MapR distribution, define the JobTracker URI in the following format: <code>maprfs:///</code></p> |
| Hive Warehouse Directory on HDFS | <p>The absolute HDFS file path of the default database for the warehouse that is local to the cluster. For example, the following file path specifies a local warehouse:</p> <pre>/user/hive/warehouse</pre> <p>For Cloudera CDH, if the Metastore Execution Mode is remote, then the file path must match the file path specified by the Hive Metastore Service on the Hadoop cluster.</p> <p>For MapR, use the value specified for the <code>hive.metastore.warehouse.dir</code> property in <code>hive-site.xml</code>. You can find <code>hive-site.xml</code> in the following directory on the node that runs HiveServer2: <code>/opt/mapr/hive/hive-0.13/conf</code>.</p> |

| Property | Description |
|-----------------------------------|--|
| Advanced Hive/Hadoop Properties | <p>Configures or overrides Hive or Hadoop cluster properties in hive-site.xml on the machine on which the Data Integration Service runs. You can specify multiple properties.</p> <p>Select Edit to specify the name and value for the property. The property appears in the following format:</p> <pre><property1>=<value></pre> <p>Where</p> <ul style="list-style-type: none"> - <property1> is a Hive or Hadoop property in hive-site.xml. - <value> is the value of the Hive or Hadoop property. <p>When you specify multiple properties, &: appears as the property separator.</p> <p>The maximum length for the format is 1 MB.</p> <p>If you enter a required property for a Hive connection, it overrides the property that you configure in the Advanced Hive/Hadoop Properties.</p> <p>The Data Integration Service adds or sets these properties for each map-reduce job. You can verify these properties in the JobConf of each mapper and reducer job. Access the JobConf of each job from the Jobtracker URL under each map-reduce job.</p> <p>The Data Integration Service writes messages for these properties to the Data Integration Service logs. The Data Integration Service must have the log tracing level set to log each row or have the log tracing level set to verbose initialization tracing.</p> <p>For example, specify the following properties to control and limit the number of reducers to run a mapping job:</p> <pre>mapred.reduce.tasks=2&hive.exec.reducers.max=10</pre> |
| Temporary Table Compression Codec | Hadoop compression library for a compression codec class name. |
| Codec Class Name | Codec class name that enables data compression and improves performance on temporary staging tables. |
| Metastore Execution Mode | Controls whether to connect to a remote metastore or a local metastore. By default, local is selected. For a local metastore, you must specify the Metastore Database URI, Driver, Username, and Password. For a remote metastore, you must specify only the Remote Metastore URI. |
| Metastore Database URI | <p>The JDBC connection URI used to access the data store in a local metastore setup. Use the following connection URI:</p> <pre>jdbc:<datastore type>://<node name>:<port>/<database name></pre> <p>where</p> <ul style="list-style-type: none"> - <node name> is the host name or IP address of the data store. - <data store type> is the type of the data store. - <port> is the port on which the data store listens for remote procedure calls (RPC). - <database name> is the name of the database. <p>For example, the following URI specifies a local metastore that uses MySQL as a data store:</p> <pre>jdbc:mysql://hostname23:3306/metastore</pre> <p>For MapR, use the value specified for the <code>javax.jdo.option.ConnectionURL</code> property in <code>hive-site.xml</code>. You can find <code>hive-site.xml</code> in the following directory on the node where HiveServer 2 runs: <code>/opt/mapr/hive/hive-0.13/conf</code>.</p> |

| Property | Description |
|-----------------------------|---|
| Metastore Database Driver | <p>Driver class name for the JDBC data store. For example, the following class name specifies a MySQL driver:</p> <pre>com.mysql.jdbc.Driver</pre> <p>For MapR, use the value specified for the <code>javax.jdo.option.ConnectionDriverName</code> property in <code>hive-site.xml</code>. You can find <code>hive-site.xml</code> in the following directory on the node where HiveServer 2 runs: <code>/opt/mapr/hive/hive-0.13/conf</code>.</p> |
| Metastore Database Username | <p>The metastore database user name.</p> <p>For MapR, use the value specified for the <code>javax.jdo.option.ConnectionUserName</code> property in <code>hive-site.xml</code>. You can find <code>hive-site.xml</code> in the following directory on the node where HiveServer 2 runs: <code>/opt/mapr/hive/hive-0.13/conf</code>.</p> |
| Metastore Database Password | <p>The password for the metastore user name.</p> <p>For MapR, use the value specified for the <code>javax.jdo.option.ConnectionPassword</code> property in <code>hive-site.xml</code>. You can find <code>hive-site.xml</code> in the following directory on the node where HiveServer 2 runs: <code>/opt/mapr/hive/hive-0.13/conf</code>.</p> |
| Remote Metastore URI | <p>The metastore URI used to access metadata in a remote metastore setup. For a remote metastore, you must specify the Thrift server details.</p> <p>Use the following connection URI:</p> <pre>thrift://<hostname>:<port></pre> <p>Where</p> <ul style="list-style-type: none"> - <code><hostname></code> is name or IP address of the Thrift metastore server. - <code><port></code> is the port on which the Thrift server is listening. <p>For MapR, use the value specified for the <code>hive.metastore.uris</code> property in <code>hive-site.xml</code>. You can find <code>hive-site.xml</code> in the following directory on the node where HiveServer 2 runs: <code>/opt/mapr/hive/hive-0.13/conf</code>.</p> |

HDFS Connection Properties

Use a Hadoop File System (HDFS) connection to access data in the Hadoop cluster. The HDFS connection is a file system type connection. You can create and manage an HDFS connection in the Administrator tool, Analyst tool, or the Developer tool. HDFS connection properties are case sensitive unless otherwise noted.

Note: The order of the connection properties might vary depending on the tool where you view them.

The following table describes HDFS connection properties:

| Property | Description |
|--------------|---|
| Name | Name of the connection. The name is not case sensitive and must be unique within the domain. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * () - + = { [] } \ : ; " ' < , > . ? / |
| ID | String that the Data Integration Service uses to identify the connection. The ID is not case sensitive. It must be 255 characters or less and must be unique in the domain. You cannot change this property after you create the connection. Default value is the connection name. |
| Description | The description of the connection. The description cannot exceed 765 characters. |
| Location | The domain where you want to create the connection. Not valid for the Analyst tool. |
| Type | The connection type. Default is Hadoop File System. |
| User Name | User name to access HDFS. |
| NameNode URI | <p>The URI to access HDFS.</p> <p>Use the following format to specify the NameNode URI in Cloudera and Hortonworks distributions:</p> <pre>hdfs://<namenode>:<port></pre> <p>Where</p> <ul style="list-style-type: none">- <namenode> is the host name or IP address of the NameNode.- <port> is the port that the NameNode listens for remote procedure calls (RPC). <p>Use one of the following formats to specify the NameNode URI in MapR distribution:</p> <ul style="list-style-type: none">- maprfs:///- maprfs:///mapr/my.cluster.com/ <p>Where <code>my.cluster.com</code> is the cluster name that you specify in the <code>mapr-clusters.conf</code> file.</p> |

Identifier Properties in Database Connections

When you create most relational database connections, you must configure database identifier properties. The identifier properties determine whether the Data Integration Service encloses identifiers within delimited characters when the service generates SQL queries to access the database.

A database identifier is a database object name. Tables, views, columns, indexes, triggers, procedures, constraints, and rules can have identifiers. You use the identifier to reference the object in SQL queries. A database can have regular identifiers or delimited identifiers that must be enclosed within delimited characters.

Regular Identifiers

Regular identifiers comply with the format rules for identifiers. Regular identifiers do not require delimited characters when they are used in SQL queries.

For example, the following SQL statement uses the regular identifiers *MYTABLE* and *MYCOLUMN*:

```
SELECT * FROM MYTABLE
WHERE MYCOLUMN = 10
```

Delimited Identifiers

Delimited identifiers must be enclosed within delimited characters because they do not comply with the format rules for identifiers.

Databases can use the following types of delimited identifiers:

Identifiers that use reserved keywords

If an identifier uses a reserved keyword, you must enclose the identifier within delimited characters in an SQL query. For example, the following SQL statement accesses a table named *ORDER*:

```
SELECT * FROM "ORDER"  
WHERE MYCOLUMN = 10
```

Identifiers that use special characters

If an identifier uses special characters, you must enclose the identifier within delimited characters in an SQL query. For example, the following SQL statement accesses a table named *MYTABLE\$@*:

```
SELECT * FROM "MYTABLE$@"  
WHERE MYCOLUMN = 10
```

Case-sensitive identifiers

By default, identifiers in IBM DB2, Microsoft SQL Server, and Oracle databases are not case sensitive. Database object names are stored in uppercase, but SQL queries can use any case to refer to them. For example, the following SQL statements access the table named *MYTABLE*:

```
SELECT * FROM mytable  
SELECT * FROM MyTable  
SELECT * FROM MYTABLE
```

To use case-sensitive identifiers, you must enclose the identifier within delimited characters in an SQL query. For example, the following SQL statement accesses a table named *MyTable*:

```
SELECT * FROM "MyTable"  
WHERE MYCOLUMN = 10
```

Identifier Properties

When you create most database connections, you must configure database identifier properties. The identifier properties that you configure depend on whether the database uses regular identifiers, uses keywords or special characters in identifiers, or uses case-sensitive identifiers.

Configure the following identifier properties in a database connection:

SQL Identifier Character

Type of character that the database uses to enclose delimited identifiers in SQL queries. The available characters depend on the database type.

Select (None) if the database uses regular identifiers. When the Data Integration Service generates SQL queries, the service does not place delimited characters around any identifiers.

Select a character if the database uses delimited identifiers. When the Data Integration Service generates SQL queries, the service encloses delimited identifiers within this character.

Support Mixed-case Identifiers

Enable if the database uses case-sensitive identifiers. When enabled, the Data Integration Service encloses all identifiers within the character selected for the **SQL Identifier Character** property.

In the Informatica client tools, you must refer to the identifiers with the correct case. For example, when you create the database connection, you must enter the database user name with the correct case.

When the **SQL Identifier Character** property is set to none, the **Support Mixed-case Identifiers** property is disabled.

Example: Database Uses Regular Identifiers

In this example, the database uses regular identifiers. No identifiers contain a reserved keyword or a special character. The database uses identifiers that are not case sensitive.

In the database connection, set the **SQL Identifier Character** property to (None). When **SQL Identifier Character** is set to none, the **Support Mixed-case Identifiers** property is disabled.

When the Data Integration Service generates SQL queries, the service does not place delimited characters around any identifiers.

Example: Database Uses Keywords or Special Characters in Identifiers

In this example, the database uses keywords or special characters in some identifiers. The database uses identifiers that are not case sensitive.

In the database connection, configure the identifier properties as follows:

1. Set the **SQL Identifier Character** property to the character that the database uses for delimited identifiers.

This example sets the property to `"` (quotes).

2. Clear the **Support Mixed-case Identifiers** property.

When the Data Integration Service generates SQL queries, the service places the selected character around identifiers that use a reserved keyword or that use a special character. For example, the Data Integration Service generates the following query:

```
SELECT * FROM "MYTABLE$@" /* identifier with special characters enclosed within
delimited
character */
WHERE MYCOLUMN = 10 /* regular identifier not enclosed within delimited character */
```

Example: Database Uses Case-Sensitive Identifiers

In this example, the database uses case-sensitive identifiers. The database might use keywords or special characters in some identifiers, or it might not.

In the database connection, configure the identifier properties as follows:

1. Set the **SQL Identifier Character** property to the character that the database uses for delimited identifiers.

This example sets the property to `"` (quotes).

2. Select the **Support Mixed-case Identifiers** property.

When the Data Integration Service generates SQL queries, the service places the selected character around all identifiers. For example, the Data Integration Service generates the following query:

```
SELECT * FROM "MyTable" /* case-sensitive identifier enclosed within delimited
character */
WHERE "MYCOLUMN" = 10 /* regular identifier enclosed within delimited character */
```

Searching for a Database Connection

You can search for a database connection. The Analyst tool highlights the first database connection in the list that has the search string. After you select a connection, you can test whether the connectivity is successful.

1. Click the **Find** icon.

The Find text field appears above the connection list.

2. Enter a search string.

The Analyst tool highlights the first connection name in the list that contains the search string.

Select a connection from the list and click the **Test** icon to test the connection.

Creating a Database Connection

You can create a database connection in the Analyst tool. Choose a simple connection to include basic database properties. Choose an advanced connection to include additional database specific properties.

1. Click **New** to open the **New Connection** dialog box.
2. Enter the following information:

| Option | Description |
|-------------|--|
| Name | Name of the connection. The name is not case sensitive and must be unique within the domain. The name cannot exceed 128 characters, contain spaces, or contain the following special characters: ~ ` ! \$ % ^ & * () - + = { [] \ : ; " ' < , > . ? / |
| ID | String that the Data Integration Service uses to identify the connection. The ID is not case sensitive. It must be 255 characters or less and must be unique in the domain. You cannot change this property after you create the connection. Default value is the connection name. |
| Description | Optional description for the connection. |

3. Select a database type.

Additional fields appear based on the database type you select.

4. Choose a simple connection or an advanced connection.

- To choose a simple connection, select **Simple Connection** and specify the connection properties.
- To choose an advanced connection, select **Advanced Connection** and specify additional database connection properties.

5. Click **OK**.

The Analyst tool tests the connection and displays the test status.

Editing a Database Connection

Edit a connection to make changes to the properties of the connection. You cannot change the ID of a connection.

1. Select a connection and click **Edit**.
The **Edit Connection** dialog box appears.
2. Make the required changes and click **OK**.
The Analyst tool validates the connection.
3. Click **OK** and then click **Close**.

Deleting a Database Connection

You can delete a database connection. You must have the write permission on the database connection to delete the connection.

1. Select the connection and click the **Delete** icon.
2. Click **Close**.

CHAPTER 4

Job Status Workspace

This chapter includes the following topics:

- [Job Status Workspace Overview, 42](#)
- [Accessing the Job Status Workspace, 42](#)
- [Job Properties, 43](#)
- [Monitoring Jobs, 44](#)

Job Status Workspace Overview

Use the **Job Status** workspace to monitor the status of ad hoc jobs such as profile, scorecard, and mapping specification jobs. Ad hoc jobs are jobs that users run from the Developer tool or the Analyst tool.

You can monitor the status of ad hoc jobs such as data preview for assets and drill-down operations on profiles. For example, you might need to view the status of a data preview job for a mapping specification if the Analyst tool could not perform the data preview. You can filter by job type to narrow the results to data preview jobs.

By default, you can monitor jobs that you run. If you have the appropriate privilege, you can also view jobs that other users run.

When you select a job, you can view logs for the job, view context for the job, or cancel the job. You can also view properties and messages for the job in the job panel.

Note: You might not be able to view job status if the Analyst tool uses the HTTPS security protocol and the Administrator tool uses the HTTP security protocol. Contact an administrator to configure the HTTPS security protocols for both tools.

Accessing the Job Status Workspace

Access the Job Status workspace to view and monitor jobs.

- From the **Manage** menu, select **Job Status**.
The **Job Status** workspace appears.

Job Properties

You can view the properties for each job, such as the job state, the user who started the job, and the duration of the job.

You can view the following job properties:

Name

Name of the job.

Type

Type of job. You can filter by a particular job type to view the status of a job. Select **Custom** to filter by multiple job types. You can choose the following options:

- Preview
- Mapping
- Reference Table
- Enterprise Discovery Profile
- Profile
- Scorecard

State

State of the job. You can filter by a particular job state to view the progress of a job. Select **Custom** to filter by multiple job states. You can choose to view the following states:

- Running. The Analyst Service is running the job.
- Completed. The Analyst Service successfully completes the job.
- Failed. The Analyst Service encounters a fatal error while processing the job.
- Aborted. The Analyst Service aborted the job.
- Canceled. You chose to cancel a running job.
- Queued. The Analyst Service has queued the job for processing.
- Unknown. The Analyst Service cannot determine the state of a job.

Job ID

Unique identifier for the job.

Started By

Name of the user who started the job.

Start Time

Start time of the job. You can filter by a particular start time. Select **Custom** to filter by date and time range. You can choose to view one of the following options for start times:

- Last 30 minutes
- Last 4 hours
- Last 1 day
- Last 1 week

Elapsed Time

The amount of time that the job ran. Select **Custom** to filter by date and time range.

End Time

The time that the job ended. You can filter by a particular end time. Select **Custom** to filter by date and time range. You can choose the following options for end time:

- Last 30 minutes
- Last 4 hours
- Last 1 day
- Last 1 week

User Security Domain

Security domain for the user name. Security domain can be Native, LDAP, or the Kerberos domain.

Monitoring Jobs

You can monitor the status of jobs related to a data preview or a profile drill down.

You can perform the following tasks when you monitor jobs:

Search for a job.

Search for a job by a job status property or through a search filter. After you apply a search filter, you can clear the filter.

To search by a job status property, enter a job status property in the search field.

To search by applying filters, click the filter menu on a job status property. Optionally, enter a custom filter for the **Start Time** and **Elapsed Time** properties.

To clear search filters, click the **Reset Filters** icon.

View the context of a job.

View a job in the context of other jobs that started around the same time as the selected job.

To view the context of a job, from the Actions menu, select **View Context**. The Analyst tool displays a list of jobs that started around the same time as the selected job.

Refresh the list of jobs.

To refresh the list of jobs, from the Actions menu, select **Refresh**.

Request notifications for new jobs.

To request notifications for new jobs, from the Actions menu, select **New Job Notifications**.

Cancel a job.

You can cancel a running job. You might want to cancel a job that hangs or that is taking an excessive amount of time to complete.

To cancel a job, from the Actions menu, click **Cancel Selected Job**.

View job log events.

You can view log events for a selected job. The event severity values are Info, Error, Warning, Trace, Debug, Fatal. Default is Info.

To view log events for a job, from the Actions menu, click **View Logs for Selected Object**. The Analyst tool creates a text file that contains the logs. You can open or download the text file to view the logs.

CHAPTER 5

Projects Workspace

This chapter includes the following topics:

- [Projects Workspace Overview, 45](#)
- [Accessing the Projects Workspace, 45](#)
- [Manage Projects and Folders, 46](#)
- [Project Security, 47](#)

Projects Workspace Overview

Use the **Projects** workspace to manage projects and folders and assign permissions on projects and folders. The projects and folders appear in the **Projects** panel.

A project is the top-level container that you use to store folders and repository content. You can also store assets in the Analyst tool in projects. Use projects to organize and manage folders and assets.

Use folders to organize project contents. Create folders to group assets based on business needs. You can create a folder in a project or in another folder. When you create a project or folder, the Analyst tool stores the project or folder in the Model repository.

For example, you need to assess data quality on multiple systems structured by region in a country. You create projects named East and West to correspond with data for East and West regions. You create folders named Customers and Accounts in the East and West projects to organize the data in these projects. You can import assets such as table objects and flat file objects in the Customers and Accounts folders.

Accessing the Projects Workspace

Access the **Projects** workspace to manage projects and folders.

- From the **Manage** menu, select **Projects**.
The **Projects** workspace appears.

Manage Projects and Folders

You can perform tasks to manage projects and folders in the **Projects** workspace.

You can perform the following tasks on a project or folder:

Create a project or folder.

Create a project to store data objects and assets in the Analyst tool. You can create folders in projects.

From the Actions menu in the **Projects** panel, click **New > Project** or click **New > Folder** and enter a project or folder name or optional description.

Duplicate a project or folder.

Duplicate a project or folder within a project to use the same contents to perform different tasks. For example, duplicate the Customers project that contains customer address tables to use the same tables for a Customer_Accounts project.

Select the project or folder that you want to duplicate. You cannot duplicate a project in another project with the same name. You cannot duplicate a folder within a project to another folder in a different project. Duplicating a project does not duplicate the user permissions on the project. The owner of the project gets all permissions by default on the duplicate project.

From the Actions menu in the **Projects** panel, click **Duplicate** and enter a project or folder name or optional description.

Rename a project or folder.

Rename a project or folder after you create it to change the project or folder name to fit the particular business use or naming convention. Select the project or folder that you want to rename.

From the Actions menu in the **Projects** panel, click **Edit** and enter another project or folder name.

Edit a project or folder description.

Edit the description of a project or folder after you create it. Select the project or folder that you want to edit.

From the Actions menu in the **Projects** panel, click **Edit** and enter a project or folder description.

Delete a project or folder.

Delete a project or folder when you no longer need it. Select the project or folder that you want to delete. Before you delete a project or folder, verify that the contents are not used in other projects or folders.

From the Actions menu in the **Projects** panel, click **Delete**.

Refresh a project or folder.

Refresh the contents of a project or folder to view the latest contents and project permissions.

From the Actions menu in the **Projects** panel, click **Refresh**.

Move a folder.

Move folders within another folder in a project to organize project content into a hierarchy of folders.

You cannot move a folder into one of its own child folders in a project. Select the folder that you want to move.

From the Actions menu in the **Projects** panel, click **Move**.

View or assign permissions on a project.

View or assign project permissions to users or groups. Select the project that you want to assign or view permissions on.

View permissions on a project in the **Direct Permissions** panel.

Assign permissions on a project from the **Edit Permissions** dialog box.

Project Security

Manage permissions on projects in the Analyst tool to control access to projects. You can add users to a project and assign permissions for users on a project.

Even if a user has the privilege to perform certain actions, the user may also require permission to perform the action on a particular asset.

When you create a project, you are the owner of the project by default. The owner has all permissions, which you cannot change. The owner can assign permissions to users.

You can assign the following permissions to a user or group:

Read

The user or group can open, preview, export, validate, and deploy all assets in the project. The user or group can also view project details.

Write

The user or group has read permission on all assets in the project. Additionally, the user or group can edit all assets in the project, edit project details, and delete all assets in the project.

Grant

The user or group has read permission on all assets in the project. Additionally, the user or group can assign permissions to other users or groups.

Project Permissions

Assign project permissions to users or groups. Project permissions determine whether a user or group can view assets, edit assets, or assign permissions to others. Permissions can be direct, inherited, or effective permissions.

Direct permissions are permissions that are assigned directly to a user or group. When users and groups have permission on an object, they can perform administrative tasks on that object if they also have the appropriate privilege. You can edit direct permissions.

Inherited permissions are permissions that users inherit. When users have permission on a project, they inherit permission on all folders and data objects in the project. When groups have permission on a project, all subgroups and users that belong to the group inherit permission on the project. For example, a project has a folder named Customers that contains multiple folders. If you assign a group permission on the project, all subgroups and users that belong to the group inherit permission on the Customers folder and on all folders in the folder.

Effective permissions are a superset of all permissions for a user or group. These include direct permissions and inherited permissions.

Users assigned the Administrator role for a Model Repository Service inherit all permissions on all projects in the Model Repository Service. Users assigned to a group inherit the group permissions.

Assigning Direct Permissions on a Project

You can add users to a project and assign direct permissions on a project to restrict, provide access, or manage the assets within the project.

1. Select a project on which you want to assign direct permissions.
2. Click the **Edit Permissions** icon.
The **Edit Permissions** dialog box appears.
3. Select users, groups, or both from the **Users and groups** panel.
4. Optionally, click the **Add Users and Groups** icon to add users and groups to the project.
The **Add Groups and Users** dialog box appears.
5. Select the users and groups to which you want to assign permissions.
6. Click **Next**.
7. Select the users and groups permissions.
8. Click **Save**.
9. Optionally, choose to filter the list of users and groups by name, security domain, or type of user or group.
 - To filter by name, enter a name or string above the **Name** field.
 - To filter by security domain, click the filter menu above the **Security Domain** field.
 - To filter by type, click the filter menu above the **Type** field and select user or group.
10. Select or clear the **Read**, **Write**, and **Grant** permissions in the **Permissions** panel.
11. Click **OK**.

Viewing Permissions on a Project

When you view permissions on a project, you can view the origin of effective permissions. Permission details display direct permissions assigned to the user or group, direct permissions assigned to parent groups, and permissions inherited from parent objects.

1. Select a project for which you want to view permissions.
2. Click the **Effective Permissions** icon.
The **Effective Permissions** dialog box appears.
3. View the effective permissions for users and groups. The permissions you see include both direct and inherited permissions.
4. Optionally, choose to filter the list of users and groups by name, security domain, or type of user or group.
 - To filter by name, enter a name or string above the **Name** field.
 - To filter by security domain, click the filter menu above the **Security Domain** field.
 - To filter by type, click the filter menu above the **Type** field and select user or group.
5. Click **Close**.

CHAPTER 6

Model Repository

This chapter includes the following topics:

- [Model Repository Overview, 49](#)
- [Informatica Analyst Assets, 49](#)
- [Repository Asset Locks, 50](#)
- [Team-based Development with Versioned Objects, 51](#)

Model Repository Overview

The Model repository is a relational database that stores the metadata for projects and folders.

Each time you open the Analyst tool, you connect to the Model repository to access projects and folders.

When you edit an asset, the Model repository locks the asset for your exclusive editing. An administrator can also integrate the Model repository with a third-party version control system. With version control system integration, you can check assets out and in.

Informatica Analyst Assets

You can manage assets within some workspaces. An asset is a type of object that you use to support business operations within the enterprise.

For example, a profile is an asset that an analyst can create to discover the content, quality, and structure of a data source.

You can create the following types of assets:

Glossary assets

Create Glossary assets in the **Glossary** workspace. You can create the following types of Glossary assets:

- **Business term.** A word or phrase that uses business language to define relevant concepts for business users in an organization.
- **Business initiative.** A business decision that results in bulk changes to a collection of Glossary assets.
- **Category.** A descriptive classification of business terms and policies.

- Glossary. A collection of categories, business terms, and policies.
- Policy. The business purpose, process, or protocol that governs business practices that are related to business terms.

Discovery assets

Create Discovery assets in the **Discovery** workspace. You can create the following types of Discovery assets:

- Profile. A single data object profile or enterprise discovery profile that discovers column characteristics and data domains in a single data source or multiple data sources.
- Flat file data object. A representation of data based on a flat file.
- Table data object. A representation of data based on a relational table.

Design assets

Create Design assets in the **Design** workspace. You can create the following types of Design assets:

- Mapping specification. A template that describes the movement and transformation of data from a source to a target.
- Reference table. A table that contains the standard version and alternative versions of a set of data values.
- Rule specification. An object that represents the logic in a business rule.

Scorecards assets

Open scorecard assets in the **Scorecards** workspace. A scorecard is a graphical representation of the quality measurements in a profile.

Repository Asset Locks

The Model repository locks assets to prevent users from overwriting work. The Model repository can lock any asset that the Analyst tool displays in the Library workspace, except for projects and folders.

When you begin to edit an asset in the Analyst tool, the Model repository locks the asset so that other users cannot save changes to it. When you save the asset, you retain the lock. When you close the asset, the Model repository unlocks the asset.

If you open an asset that another user has locked, the Analyst tool notifies you that the asset is locked by another user. The object might be locked in the Analyst tool or the Developer tool. You can choose to review the asset in read-only mode, or you can save the asset with another name.

The Model repository retains asset locks if the Analyst tool shuts down. When you connect to the Model repository again, you can continue to edit the assets that you have locked. To edit an asset that is locked by another user, contact that user or the administrator.

The **Properties** view for each locked asset displays the time and date it was locked and the user ID of the lock owner.

Rules and Guidelines for Asset Lock Management

Consider the following rules and guidelines when you manage asset locks:

- The Model repository does not lock the asset when you open it. The Model repository locks the asset only after you begin to edit it. For example, the Model repository locks a mapping specification when you insert a cursor in an editable field or rename the asset.
- You can use more than one client tool to develop an asset. For example, you can edit an asset on one machine and then open the asset on another machine and continue to edit it. When you return to the first machine, you must close the asset and reopen it to regain the lock. The same principle applies when a user with administrative privileges unlocks an asset that you had open.
- An administrator can revoke your write permission on an asset you have locked, or reassign the lock to another user. In this case, you cannot edit or save the asset. You can save the asset with another name.

Team-based Development with Versioned Objects

Team-based development is the integration of the Model repository with a third-party version control system. The version control system saves multiple versions of assets and assigns each version a version number. You can check assets out and in and undo the checkout of assets.

The Model repository protects assets from being overwritten by other members of the development team. If you open an asset that another user has checked out, you receive a notification that identifies the user who has it checked out. You can open a checked out asset in read-only mode, or you can save it with a different name.

Use the **My Checked Out Assets** view to manage assets that you have checked out. For example, you might want to undo a checkout to delete changes to an asset.

When the connection to the version control system is active, the Model repository has the latest version of each asset.

The Model repository maintains the state of checked-out assets if it loses the connection to the version control system. While the connection with the version control system is down, you can continue to open, edit, save, and close assets. The Model repository tracks and maintains asset states.

When the connection is restored, you can resume version control system-related actions, such as checking in or undoing the checkout of assets. If you opened and edited an asset during the time when the connection was down, the Model repository checks out the asset to you.

Versioned Asset Management

When the Model repository is integrated with a version control system, you can manage versions of assets. For example, you can check out and check in assets, undo checkouts, and view assets that you have checked out.

You can perform the following actions:

Check out an asset.

When you check out an asset, it maintains a checked-out state until you check it in or undo the checkout. You can view assets that you have checked out in the **My Checked Out Assets** view. To check out an asset, right-click the asset in the Object Library and select Check Out.

Undo the checkout of an asset.

When you undo a checkout, you check in the asset without changes and without incrementing the version number or version history. All changes that you made to the asset after you checked it out are lost. To undo a checkout, you can use the **My Checked Out Assets** view.

Check in an asset.

When you check in an asset, the version control system updates the version history and increments the version number. You can add check-in comments up to a 4 KB limit. To check in an asset, use the **My Checked Out Assets** view or the object right-click menu.

Delete an asset.

A versioned asset must be checked out before you can delete it. If it is not checked out when you perform the delete action, the Model repository checks the asset out to you and marks it for delete. To complete the delete action, you must check in the asset.

When you delete a versioned asset, the version control system deletes all versions.

To delete an asset, you can use the **My Checked Out Assets** view.

My Checked Out Assets View

The **My Checked Out Assets** view lists all the assets that you have checked out.

You can perform the following actions in the **My Checked Out Assets** view:

- Undo the checkout of an asset.
- Check in an asset.
- Delete an asset.

To access the view, right-click an asset in the **My Checked Out Assets** view, and select an action.

Deleting an Asset

When you delete an asset that is under version control, you mark the asset for deletion and then check it in.

1. Right-click the asset in the **Library Navigator** view or the **My Checked Out Assets** view and choose **Delete**.
2. Select the asset in the **My Checked Out Assets** view and choose **Check In**.

The asset is deleted from the Model repository.

CHAPTER 7

Data Objects

This chapter includes the following topics:

- [Data Objects Overview, 53](#)
- [Flat File Data Objects, 54](#)
- [Table Data Objects, 59](#)
- [Synchronize Data Objects, 60](#)
- [Viewing Data Objects, 62](#)
- [Editing Data Objects, 62](#)

Data Objects Overview

A data object represents the source from which you want to extract metadata. You can import flat files and tables as data objects to analyze the structure of the data.

Flat file data objects and table data objects are Discovery assets that you can use as the starting point for a collaborative project in your organization. You can add data objects by importing them into the Analyst tool. You can create a profile for the source data that the table data objects and flat file data objects represent. When you run the profile, the Analyst tool connects to the database table or flat file. You can then use the table data objects and flat file data objects to perform tasks such as data analysis or data integration tasks.

When you import a data object you must access the source to extract the metadata. Access relational sources through a connection object available in the Analyst tool. Access flat file sources through the network path.

Create flat file data objects and table data objects in the **Discovery** workspace. Use the workspace hover menu or use the **New Assets** panel to create data objects. You can also create data objects from the **New** menu in the Analyst tool header. After you add the data objects to the project or folder, you can view data objects in the **Projects** panel in the **Library** workspace.

Flat File Data Objects

A flat file data object contains the metadata for a flat file. Use a flat file data object as the starting point in a collaborative project. When you add a flat file data object, the Analyst tool connects to the network path location or the location where you upload the source flat file to extract metadata.

To add a flat file data object, you must select the flat file, configure the file options, and configure the column data types. After you add the flat file data object, you can preview its properties and column data.

You can add a flat file data object as fixed-width or delimited. When you add a flat file data object as fixed-width, you can format data by fixed-width column breaks. When you add a flat file data object as delimited, you can format data by delimiters such as commas for column breaks.

You can also synchronize the changes to the flat file data object to get the updated metadata if the source flat file changes.

Import Flat File Data Objects

You can add flat file data objects in the Analyst tool by importing the flat files into projects or folders. When you import a flat file data object, you can choose to upload a flat file from your local machine or you can choose a network path. Choose a network path to import a flat file data object if the flat file is greater than 10 MB.

When you upload a flat file from your local machine, the Analyst tool uploads a copy of the flat file to a flat file cache directory in the Informatica services installation directory that the Analyst tool can access. Contact an administrator to configure the flat file cache that the Analyst tool uses for the network path. When you choose a network path, you can specify the location to the flat file on your local machine.

You can synchronize the changes to the flat file data object if you modify the source flat file.

When you import a flat file data object, the Analyst tool infers the Numeric or String data types for flat file fields based on the first 10,000 rows.

Flat File Options

When you import a flat file data object, you can configure the flat file options for each column in **Add Flat File** wizard. The options that you configure determine how the wizard reads the data from the source flat file.

You can configure the following flat file options in the **Add Flat File** wizard:

Code Page

Code page of the data in the flat file object. Select a code page that matches the code page of the data in the flat file object.

Delimiters

Character used to separate columns of data. Use the Other field to enter a different delimiter. Printable characters must be different from the escape character and the quote character if selected. You can input the following non-printing multibyte characters: \1, \01, or \001.

Text Qualifier

Quote character that defines the boundaries of text strings. Choose No Quote, Single Quote, or Double Quotes. If you select a quote character, the wizard ignores delimiters within pairs of quotes.

Column Names

Option to import column names from the first line. Select this option if column names appear in the first row. The wizard uses data in the first row in the preview for column names.

If the first row contains numeric characters, the wizard uses COLUMNx as the default column name. If the first row contains special characters, the wizard converts the special characters to underscore and uses the valid characters in the column name. The wizard skips the following special characters in a column name: " . + - = ~ ` ! % ^ & * () [] { } ' \ " ; : ? , < > \ \ | \t \r \n. Default is not enabled.

Values

Option to start value import from a line. Indicates the row number in the preview at which the wizard starts reading when it imports the file.

Flat File Data Types

Configure data types for the data in each column in the **Add Flat File** wizard. The data types you configure determine how the wizard imports the data from the source flat file.

Configure the following data types:

- Bigint. You can specify the format in the **Numeric Format** window. You can use the default or specify another numeric format and choose to make this the default numeric format.
- Datetime. You can specify the format in the **Datetime Format** window. You can use the default or specify another datetime format and choose to make this the default datetime format.
- Double. You can specify the format in the **Numeric Format** window. You can use the default or specify another numeric format and choose to make this the default numeric format.
- Int. You can specify the format in the **Numeric Format** window. You can use the default or specify another numeric format and choose to make this the default numeric format.
- Nstring. You can specify a value for precision. You cannot specify a format.
- Number. You can specify values for precision and scale. You can specify the format in the **Numeric Format** window. You can use the default or specify another numeric format and choose to make this the default numeric format.
- String. You can specify a value for precision. You cannot specify a format.

Datetime Data Types

When you configure the datetime data type, you can specify the format in the **Datetime Format** window. You can use the default or specify another datetime format and choose to make this the default datetime format.

You can specify the following datetime format strings as part of the date:

AM, a.m., PM, p.m.

Meridian indicator. Use any of these format strings to specify AM and PM hours. AM and PM return the same values as a.m. and p.m.

DAY

Name of day, including up to nine characters. The DAY format string is not case sensitive.

DD

Day of month.

DDD

Day of year including leap years.

DY

Abbreviated three-character name for a day. The DY format string is not case sensitive.

HH, HH12

Hour of the day.

HH24

Hour of the day from 0-23, where 0 is 12AM.

J

Modified Julian Day.

MI

Minutes from 0-59.

MM

Month

MONTH

Name of month, including up to nine characters. Case does not matter.

MON

Abbreviated three-character name for a month. Case does not matter.

MS

Milliseconds from 0-999.

NS

Nanoseconds from 0-999999999.

RR

Four-digit year. Use when source strings include two-digit years.

SS

Seconds from 0-59.

SSSSS

Seconds since midnight.

US

Microseconds from 0-999999.

Y

The current year with the last digit of the year replaced with the string value.

YY

The current year with the last two digits of the year replaced with the string value.

YYY

The current year with the last three digits of the year replaced with the string value.

YYYY

Four digits of a year. Do not use this format string if you are passing two-digit years. Use the RR or YY format string instead.

Adding a Delimited Flat File

When you import a flat file data object to a project or folder, you can set delimiters to format the data. You can change column attributes to match the data preview.

1. On the **New** header, click **Flat File Data Object**.
The **Add Flat File wizard** appears.
2. Choose to browse for a location or enter a network path to import the flat file.
 - To browse for a location, select **Browse and Upload** and click **Choose File** to select the flat file from a directory that your machine can access.
 - To enter a network path, select **Enter a Network Path** and configure the path and file name of the file.
3. Click **Next**.
4. Accept the default **Delimited** option.
5. Click **Next**.
6. Configure the flat file options, and preview the flat file data.
Note: Select a code page that matches the code page of the data in the file.
7. Optionally, click the **Refresh** icon in the **Preview** panel to update preview changes to the flat file data.
8. Click **Next**.
9. Optionally, change the **Column Attribute**.
10. Click **Next**.
11. Configure the name, optional description, and the location in the **Folders** panel where you want to add the flat file.
The **Flat Files** panel displays the flat files that exist in a project or folder.
12. Click **Finish**.
The Analyst tool displays the data preview for the flat file on the **Data Preview** tab. View the properties for the flat file on the **Properties** tab.

Adding a Fixed-width Flat File

When you import a fixed-width flat file to a project or folder, you can set column breaks to format the data.

1. On the **New** header, click **Flat File Data Object**.
The **Add Flat File wizard** appears.
2. Choose to browse for a location or enter a network path to import the flat file.
 - To browse for a location, select **Browse and Upload** and click **Choose File** to select the flat file from a directory that your machine can access.
 - To enter a network path, select **Enter a Network Path** and configure the path and file name of the file.
3. Click **Next**.
4. Select **Fixed-width**.
5. Click **Next**.
6. Configure the flat file options, and preview the flat file data.
Note: Select a code page that matches the code page of the data in the file.
7. Optionally, click the **Refresh** icon in the **Preview** panel to update preview changes to the flat file data.

8. Choose to set, remove, move, or edit column breaks.
 - To set a column break, click within the **Preview** panel.
 - To remove a column break, double-click the column break.
 - To move column breaks, drag them.
 - To edit column breaks, click the **Edit Breaks** icon and use the **Edit Breaks** dialog box to modify column breaks.
9. Click **Next**.
10. Optionally, change the **Column Attribute**.
11. Click **Next**.
12. Configure the name, optional description, and the location in the **Folders** panel where you want to add the flat file.

The **Flat Files** panel displays the flat files that exist in a project or folder.
13. Click **Finish**.

The Analyst tool displays the data preview for the flat file on the **Data Preview** tab. View the properties for the flat file on the **Properties** tab.

Rules and Guidelines for Flat Files

Consider the following rules and guidelines while working with flat files:

Upload small files to an Informatica services installation directory.

Upload files up to 10 MB to an Informatica services installation directory on the machine where the Analyst tool runs. The Analyst tool accesses this location to extract flat file metadata that does not change frequently. When you use files of sizes up to 10 MB, the Analyst tool accesses a copy of the file in the Informatica services installation directory. If you modify the original file, you need to upload the file again.

Upload large files to a network path location.

Enable the Analyst tool to connect to a network path location for files greater than 10 MB. The Analyst tool accesses this location to extract flat file metadata that changes frequently. The network path location should be a shared directory or file system that the Analyst tool can access. When you use files greater than 10 MB, the Analyst tool can connect to the flat file in the network path. If you modify the original flat file, you must refresh the flat file in the Analyst tool. Refreshing metadata for a large flat file can take time.

Blank data rows are not imported.

The Analyst tool does not import the blank rows above the first data row, blank middle rows, and blank rows after the last data row when importing a flat file.

Refresh the data preview.

After a preview, you can change the row number at which the **Add Flat File** wizard starts reading when it imports the file. This row number corresponds with the preview. If you choose to import column names from the first line, refresh the preview to update the row numbers for the preview data.

Table Data Objects

A table data object contains the metadata for a relational database source in the Analyst tool. Use table data objects to analyze source data. When you add a table data object, the Analyst tool uses a database connection to connect to the source database to extract metadata.

You can add table data objects in the Analyst tool by importing the tables into projects or folders. Before you import a table data object, you select or create a database connection and select the database table that you want to add. You can add multiple tables from a connection as data objects. You can also search for a table or table schema when you import a table data object.

Use the **New Table** wizard to add a table data object to the project or folder. Use the Connections workspace to create a database connection to connect to the source table when you import it as a table data object.

Adding a Table

Use the **New Table** wizard to add a table data object to a project. Add the table data object that you want to analyze source data for. To add a table data object, select a connection, select the schema and tables, and add the table data object.

1. On the **New** header, click **Table Data Object**.

The **New Table** wizard appears.

2. Select a connection.

3. Click **Next**.

4. Optionally, unselect **Show Default Schema Only** to show all schemas associated with the selected connection.

5. Select the table you want to add.

6. Optionally, choose to search for a table by table name or schema name, or by table name and schema name.

- To search for a table by table name, enter a table name in the **Tables** search box and click the **Find** icon to search by table name. Click the **Clear** icon to display all tables by name.
- To search for a table by schema name, enter a table schema name in the **Schema** search box and click the **Find** icon to search by table schema name. Click the **Clear** icon to display all schemas by name.
- To search for a table by table name and schema name, enter a table name in the **Tables** search box and a schema name in the **Schema** search box and click the **Find** icon to display all tables by name in schemas by name. Click the **Clear** icon to display all schemas by name.

7. Optionally, on the **Properties** tab, view the properties and column metadata for the table.

8. Optionally, click the **Data Preview** tab to view the columns and data for the table.

9. Click **Next**.

10. Select a project or folder in the **Folders** panel where you want to add the table.

The **Tables** panel displays that tables that exist in the project or folder.

11. Click **Finish**.

Rules and Guidelines for Tables

Consider the following rules and guidelines when you work with tables:

- The Analyst tool displays the first 100 rows by default when you preview the data for a table. The Analyst tool might not display all the data columns in a wide table.
- The Analyst tool can import wide tables with more than 30 columns to profile data. When you import a wide table, the Analyst tool does not display all the columns in the data preview. The Analyst tool displays the first 30 columns in the data preview. However, you can include all the columns in the wide tables and flat files for profiling.
- You can import tables and columns with lowercase and mixed-case characters.
- You can import tables that have special characters in the table or column name. When you import a table that has special characters in the table or column name, the Analyst tool converts the special character to an underscore character in the table or column name. You can use the following special characters in table or column names:

" \$. + - = ~ ` ! % ^ & * () [] { } ' \ " ; : / ? , < > \ \ | \t \r \n

- You can import tables and columns with Microsoft SQL92 or Microsoft SQL99 reserved words such as "concat" into the Analyst tool.
- You can use an ODBC connection to import Microsoft SQL Server, MySQL, Teradata, and Sybase tables in the Analyst tool. The ODBC connection requires a user name and password.
- When you use a Microsoft SQL Server connection to access tables in a Microsoft SQL Server database, the Analyst tool does not display the synonyms for the tables.
- When you preview relational table data from an Oracle, IBM DB2, IBM DB2 for zOS, IBM DB2/iOS, Microsoft SQL Server, and ODBC database, the Analyst tool cannot display the preview if the table, view, schema, synonym, and column names contain mixed case or lower case characters. To preview data in tables that reside in case sensitive databases, set the **Support Mixed Case Identifiers** attribute to true in the connections for Oracle, IBM DB2, IBM DB2 for zOS, IBM DB2/iOS, Microsoft SQL Server, and ODBC databases in the Developer tool or Administrator tool.
- You can view comments for the source database table after you import the table into the Analyst tool. To view source table comments, use an additional parameter in the JDBC connection URL used to access metadata from the database. In the **Metadata Access String** option in the database connection properties, use `CatalogOptions=1` or `CatalogOptions=3`. For example, use the following JDBC connection URL for an Oracle database connection:

```
Oracle: jdbc:informatica:oracle:// <host_name>:<port>;SID=<database
name>;CatalogOptions=1
```

Synchronize Data Objects

Synchronize the changes to a flat file or table data object with its external data source. If the external flat file or table data source changes, you can synchronize the changes to the flat file or table data object in the Analyst tool.

You can synchronize the changes to a data object with its external data source from the **Discovery** workspace. You can also open a data object in the **Library** workspace and right-click the data object to synchronize it with its external data source.

Synchronizing a Flat File Data Object

You can synchronize the changes to an external flat file data source with its data object in the Analyst tool. Use the **Synchronize Flat File** wizard to synchronize the data objects.

1. Open the **Library** workspace.
2. In the **Projects** section, select a flat file data object from a project.
The Analyst tool displays the data preview for the flat file in the **Data Preview** tab.
3. Click the **Properties** tab.
4. From the Actions menu, click **Synchronize**.
The **Synchronize Flat File** wizard appears.
5. Choose to browse for a location or enter a network path to import the flat file.
 - To browse for a location, click **Choose File** to select the flat file from a directory that your machine can access.
 - To enter a network path, select **Enter a Network Path** and configure the file path and file name.
6. Click **Next**.
7. Choose to import a delimited or fixed-width flat file.
 - To import a delimited flat file, accept the **Delimited** option.
 - To import a fixed-width flat file, select the **Fixed-width** option.
8. Click **Next**.
9. Configure the flat file options for the delimited or fixed-width flat file.
10. Click **Next**.
11. Optionally, change the column attributes.
12. Click **Next**.
13. Accept the default name or enter another name for the flat file.
14. Optionally, enter a description.
15. Click **Finish**.
A synchronization message prompts you to confirm the action.
16. Click **Yes** to synchronize the flat file.
A message that states synchronization is complete appears. To view details of the metadata changes, click **Show Details**.
17. Click **OK**.

Synchronizing a Relational Data Object

You can synchronize the changes to an external relational data source with its table data object. External data source changes include adding, changing, and removing source columns and rule columns.

1. Open the **Library** workspace.
2. In the **Projects** section, select a table data object from a project.
The Analyst tool displays the data preview for the table on the **Data Preview** tab.
3. Click the **Properties** tab.
4. From the Actions menu, click **Synchronize**.

- A message prompts you to confirm the action.
5. To complete the synchronization process, click **Yes**.
A synchronization status message appears.
 6. A message that states synchronization is complete appears.
To view details of the metadata changes, click **Show Details**.
 7. Click **OK**.

Viewing Data Objects

You can view properties for each data object in a project or folder. You can open the data object to preview data in a tab. You can preview the contents of data objects and object types to view the structure of data and analyze data quality results.

1. Open the **Library** workspace, and look up the data object from the **Projects** or **Assets** section.
The Analyst tool displays the data objects in the asset list.
2. Select a data object.
The Analyst tool displays the data preview for the flat file data object or table data object on the **Data Preview** panel.
3. Click the **Properties** tab.
The Analyst tool displays the properties for the flat file data object or table data object.

Editing Data Objects

You can edit the name and description properties of tables and flat files while viewing the tables and flat files.

1. Open the **Library** workspace, and look up the data object from the **Projects** or **Assets** section.
The Analyst tool displays the data objects in the asset list.
2. Select a data object.
The Analyst tool displays the data preview for the flat file or table on the **Data Preview** panel.
3. Click the **Properties** tab to view the table or flat file properties in the **Properties** panel.
4. From the Actions menu, click **Edit** to edit the data object.
The **Edit** dialog box appears.
5. Enter a name and an optional description.
Optionally, for table data objects, enter an owner name.
6. Click **OK**.

CHAPTER 8

Search

This chapter includes the following topics:

- [Search Overview, 63](#)
- [Search Results, 63](#)

Search Overview

You can search for assets such as data objects, mapping specifications, and profiles. Use the search box in the Analyst tool header to perform the search. You can limit search results to a workspace or all workspaces that you have the privilege to access.

The Search Service must be enabled to perform a search in the Analyst tool.

When you perform a search from the Analyst tool header, a search panel appears at the bottom of the workspace that you are in. The name of the search panel appears as **Search <workspace name>** if you searched by a workspace or **Search All** if you searched by all workspaces. You can close the search panel.

You can enter another search query in the **Search** box of the search panel. The Analyst tool displays the number of results found and lists the search results.

You can apply search filters from the **Filter** panel to narrow the search results. You can also refine the search query to use keyword matches, wildcards, and operators.

Search Results

When you perform a search, the Analyst tool displays the number of search results and lists the search results in the search panel. The search panel appears at the bottom of the workspace that you are in.

The search results include assets, related assets, business terms, and policies. The results can also include column profile results and domain discovery results from a profiling warehouse.

You can apply filters to narrow the search results. Apply filters to search results from the **Filter** panel within the search panel. You can configure filter properties when you apply filters to the search results. You can hide the **Filter** panel or open it again.

You can sort or group assets in the search results by asset property. You can select an asset from the search results and open the asset in its workspace.

Tip: If the search does not return results, then you might not have permission to view projects in the workspace. Ask your administrator to verify that you have Read privilege for the projects in the workspace.

Search Query

Use keyword matches, wildcards, or operators to refine a search query.

You can use the following characters in a search query:

Keywords

Use an exact keyword match in the search. Enclose a search query in quotation marks (" ") to search for an exact keyword match. The Analyst tool returns assets with the name that matches the keyword exactly.

Wildcards

Use the * and ? wildcard characters in the search. Use wildcard characters to define one or more characters in a search. Use wildcards as a suffix or infix in a search.

. Represents characters. For example when you search for `customer`, the Analyst tool can return `customer`, `customer_name`, and `CustomerID`. Use * with at least one character. You cannot use * at the beginning of a search query.

?. Represents a single character. For example when you search for `Customer?`, the Analyst tool can return `Customer1`, `Customer2`, and `CustomerA`.

Operators

Use the + or Space operators in the search.

+. Includes the search term. For example, to include both sales and data use the following query: `+sales +data`.

Space. Includes either one of the search terms. For example, `sales data`.

Search Properties

You can apply filters to search for assets in the search results. You can hide the **Filter** panel if you do not need to add filters. Business glossary users can specify additional asset state filters.

You can use the following filter properties:

Search

Enter a search string in the **Find** box of the **Filter** panel.

Location

Location of the asset in the business glossary or the repository.

Time (last updated)

Time that the asset was last updated. You can select the following times:

- From the beginning
- Past 1 hour
- Past 24 hours
- Past week
- Past month
- Past year

Created by

Name of the user who created at least one asset in the asset list. Select **All** to select all users.

APPENDIX A

Configure the Web Browser

This appendix includes the following topic:

- [Configure the Web Browser , 65](#)

Configure the Web Browser

You can use Microsoft Internet Explorer or Google Chrome to launch the Analyst tool in the Informatica platform.

To use the Analyst tool, configure the following options in the browser:

Scripting and ActiveX

Enable the following controls on Microsoft Internet Explorer:

- Active scripting
- Allow programmatic clipboard access
- Run ActiveX controls and plug-ins
- Script ActiveX controls marked safe for scripting

To configure the controls, click **Tools > Internet options > Security > Custom level**.

Trusted sites

Configure the browser to allow access to the Analyst tool. In Microsoft Internet Explorer, add the Analyst tool URL to the list of trusted sites. In Google Chrome, add the Analyst tool host name to the whitelist of trusted sites.

INDEX

A

- accessing the Library workspace
 - Job Status [16](#)
- assets
 - Informatica Analyst [13](#), [49](#)
 - search [63](#)

C

- checking assets out and in [51](#)
- connections
 - database connection [40](#)
 - database identifier properties [37](#)
- Connections workspace
 - creating [40](#)
 - deleting [41](#)
 - editing [41](#)
 - searching [40](#)

D

- data objects
 - assets [53](#)
 - editing [62](#)
 - flat file data object [54](#)
 - table data object [59](#)
 - viewing [62](#)
- database connections
 - identifier properties [37](#)
- delimited identifiers
 - database connections [38](#)

F

- flat file data object
 - data types [55](#)
 - datetime data types [55](#)
 - delimited [57](#)
 - fixed-width [57](#)
 - flat file options [54](#)
 - importing [54](#)
 - synchronizing [61](#)

H

- HDFS connections
 - properties [36](#)
- header
 - Informatica Analyst [10](#)
- Hive connections
 - properties [30](#)

I

- IBM DB2 connections
 - properties [19](#)
- identifiers
 - delimited [38](#)
 - regular [37](#)
- Informatica Analyst
 - assets [13](#), [49](#)
 - header [10](#)
 - interface [10](#)
 - workspaces [11](#)
- Informatica Analyst interface
 - log in [14](#)
- interface
 - Informatica Analyst [10](#)

J

- JDBC connections
 - properties [21](#)
- job status
 - monitoring [44](#)
 - properties [43](#)
- Job Status
 - accessing the Library workspace [16](#)
- Job Status workspace
 - accessing [42](#)

L

- Library
 - library tasks [16](#)
 - workspaces [15](#)

M

- Model repository
 - checking assets out and in [51](#)
 - description [49](#)
 - non-versioned [51](#)
 - team-based development [51](#)
 - versioned [51](#)
- MS SQL Server connections
 - properties [23](#)

O

- ODBC connections
 - properties [27](#)
- Oracle connections
 - properties [28](#)

P

project permissions
 direct permissions [48](#)
 effective permissions [48](#)
Projects workspace
 accessing [45](#)
 manage projects [46](#)
 permissions [47](#)
 project permissions [47](#)

R

regular identifiers
 database connections [37](#)

S

search
 filter properties [64](#)
 search results [63](#)
 search syntax [64](#)

T

table data object
 adding [59](#)

table data object (*continued*)
 synchronizing [61](#)
tags
 assigning [17](#)
 creating [17](#)
 removing [17](#)
team-based development [51](#)

V

versioning [51](#)

W

workspaces
 Connections workspace [18](#)
 Data Domains workspace [11](#)
 Design [11](#)
 Discovery workspace [11](#)
 Exceptions workspace [11](#)
 Glossary Security workspace [11](#)
 Glossary workspace [11](#)
 Informatica Analyst [11](#)
 Job Status workspace [42](#)
 Library [15](#)
 Projects workspace [45](#)
 Scorecards workspace [11](#)
 Start workspace [11](#)