Enterprise Data Catalog Tutorial
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 Agreement Supplemental License Terms, the BSD License (http://www.opensource.org/licenses/bsd-license.php), the new BSD License (http://www.opensource.org/licenses/bsd-license.php), 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.


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 merchantability, fitness for a particular purpose and non-infringement. INFORMatica 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, report them to us at infa_documentation@informatica.com.

Informatica products are warranted according to the terms and conditions of the agreements under which they are provided. INFORMatica 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 OR ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2018-10-04
# 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: Tutorial Objectives** ........................................ 9

  - Tutorial Objectives .............................................. 9

**Chapter 2: Enterprise Data Catalog Overview** ....................... 10

  - Overview .......................................................... 10
  - Objectives .......................................................... 10
  - Enterprise Data Catalog ........................................ 10
  - Key Features ........................................................ 11
  - Accessing Enterprise Data Catalog ............................ 12

**Chapter 3: Creating Users in Informatica Administrator** .......... 13

  - Overview .......................................................... 13
  - Objective ........................................................... 13
  - Creating a User Using Informatica Administrator ............ 13

**Chapter 4: Ingesting Metadata into the Catalog** .................... 15

  - Overview .......................................................... 15
  - Objectives .......................................................... 16
  - Creating a Resource ............................................. 16
  - Configuring metadata and profiling settings ................ 18
  - Running a Resource ............................................. 18
  - Monitoring a Task ............................................. 18
  - Browse Resources in the Catalog ............................... 19

**Chapter 5: Associating Business Context with Assets** .............. 20

  - Overview .......................................................... 20
  - Objectives .......................................................... 20
  - Associating Business Terms to Assets ....................... 21
  - Adding Custom Attributes to Catalog and Adding Values ... 22
# Table of Contents

Chapter 6: Discovering Assets Using Enterprise Data Catalog. 26

- Overview. 26
- Objectives. 26
- Search Features. 27
- Filter Search Results. 29
- Searching Using Business Terms. 31
- Searching Using Synonyms. 31

Chapter 7: Viewing Data Lineage Diagrams. 33

- Overview. 33
- Objectives. 33
- Using a Lineage Diagram to Perform Detailed Impact Analysis. 34
- Exporting a Lineage Diagram. 36

Chapter 8: Identifying Asset Relationships. 37

- Overview. 37
- Objective. 37
- Viewing Relationships. 37

Chapter 9: Identifying Value Frequency in Data Sources. 39

- Overview. 39
- Objectives. 39
- Enabling Value Frequency for Resources. 39
- Viewing Value Frequencies in Enterprise Data Catalog. 40

Chapter 10: Viewing Asset Transformation Logic. 42

- Overview. 42
- Objectives. 42
- Enabling Transformation Logic for Resources. 42
- Viewing Transformation Logic for Assets. 43

Chapter 11: Discovering Data Domains. 45

- Overview. 45
- Objectives. 45
- Creating Rule-based Data Domains. 46
- Running Data Domain Discovery for a Resource. 46
- Enabling Smart Data Domains. 47

Chapter 12: Searching Using Composite Data Domains. 50

- Overview. 50
- Objectives. 50
- Creating a Composite Data Domain. 51
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enabling Composite Data Domain Discovery for Resources</td>
<td>51</td>
</tr>
<tr>
<td>Searching for Assets Using Composite Data Domains</td>
<td>51</td>
</tr>
</tbody>
</table>
Preface

The Informatica Catalog Administrator Guide is written for administrators who manage and monitor Enterprise Data Catalog.

Informatica Resources

Informatica Network


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.
Tutorial Objectives

Welcome to the tutorial on Informatica Enterprise Data Catalog. The objective of this tutorial is to get you started with an overview on Enterprise Data Catalog. You will also learn about some of the key features of Enterprise Data Catalog by performing tasks in a configured environment.

This tutorial follows a concept-task approach to provide learning with a hands-on experience. This tutorial does not cover the advance features and all capabilities provided by Enterprise Data Catalog. See the following documents on the Informatica Knowledge Base portal for more information:

- Informatica Catalog Administrator Guide
- Informatica Enterprise Data Catalog User Guide

After completing this tutorial, you will be familiar with the important features and tasks in Enterprise Data Catalog.

The following are the objectives of this tutorial:

- Create users to access Enterprise Data Catalog.
- Discover data assets in the catalog using multiple search methods and filters.
- Create resources to extract metadata from data sources.
- Associate business context with assets and enrich assets with custom attributes.
- Understand and configure data domains.
- Understand data lineage, change impact, and relationships.
- Create composite data domains to search for entities across schemas in a data source.
- View transformation logic for assets.
- Identify the frequency of values in data sources.
Enterprise Data Catalog Overview

This chapter includes the following topics:

- Overview, 10
- Objectives, 10
- Enterprise Data Catalog, 10
- Key Features, 11
- Accessing Enterprise Data Catalog, 12

Overview

In this chapter, you will learn about Enterprise Data Catalog and understand its key features.

Objectives

- Understand the basic concepts of Enterprise Data Catalog and the key features.
- Access Enterprise Data Catalog.

Enterprise Data Catalog

Enterprise Data Catalog helps you analyze and understand large volumes of metadata in the enterprise. You can extract physical and business metadata for many objects, organize the metadata based on business concepts, and view data lineage and relationships for each object.

The catalog serves as a centralized repository that stores all metadata extracted from different external sources. The sources include databases, data warehouses, business glossaries, data integration resources, business intelligence reports, and more. For ease of search, the catalog maintains an indexed inventory of all the assets in an enterprise. Assets represent the data objects such as tables, columns, reports, views, and schemas. Metadata and statistical information in the catalog include profile results, information about data domains, and information about data relationships.
You can use Enterprise Data Catalog to perform the following tasks:

- Find available assets.
- Explore assets to verify the quality of data, such as profiling information.
- View lineage for assets.
- View relationships between assets.
- Enrich assets by tagging them with additional attributes.

Key Features

This deep dive tutorial focuses on the following key features of Enterprise Data Catalog:

- Data discovery. Semantic search, dynamic filtering, data lineage, and relationships for assets across the enterprise.
- Data Classification. Automatic and manual annotations to classify data aiding in governance and discovery.
- Resource administration using Catalog Administrator that includes the following tasks:
  - Resource management. Create, edit, and remove resources.
  - Schedule management. Create, edit, and remove schedules.
  - Attribute management. View system-defined attributes for metadata object types. Create custom attributes and assign to metadata object types, such as tables, views, and columns.
  - Connection management. View automatically assigned connections and schemas. Assign schemas and connections to resources. Un-assign user-assigned connections.
  - Profile configuration management. Create and edit reusable profile-definition settings.
  - Resource monitoring. Monitor resources and tasks.
  - Data domain management. Create and edit data domains and data domain groups. Assign logical data domains to data domain groups.
  - Composite data domain management. Create and edit composite data domains.
  - Monitor status of tasks and progress.
  - Transformation logic for assets. View the transformation logic for data structures such as tables and columns.
  - Value frequency for assets.

Accessing Enterprise Data Catalog

After deploying Enterprise Data Catalog, you can access Enterprise Data Catalog and Catalog Administrator using the following URLs:

- **URL to access Enterprise Data Catalog:** http://<host>:<port>/ldmcatalog
  
  **Note:** The <host> is the gateway node host name. The <port> is the Informatica Catalog Service port number. The default port number is 6705 for an HTTP connection.

- **URL to access Catalog Administrator:** http://<host>:<port>/ldmadmin
  
  **Note:** The <host> is the gateway node host name. The <port> is the Informatica Administrator port number. The default port number is 6705.
Creating Users in Informatica Administrator

This chapter includes the following topics:

- **Overview, 13**
- **Objective, 13**
- **Creating a User Using Informatica Administrator, 13**

## Overview

In this chapter, you will learn how to create users in Informatica Administrator.

## Objective

Create a user account in Informatica Administrator. You can access the Informatica Administrator using a URL in the following format: 

http://<host>:<port>/administrator/

## Creating a User Using Informatica Administrator

1. Log in to Informatica Administrator using the http://<host>:<port>/administrator/ URL.
   
   **Note:** `<Host>` and `<port>` in the URL represent the host name and port number of the master gateway node.

2. In the Administrator tool, click the **Security** tab.

3. On the **Security Actions** menu, click **Create User**.
4. Enter the following details for the user:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Login Name</td>
<td>Login name for the user account. The login name for a user account must be unique within the security domain to which it belongs. The name is not case sensitive and cannot exceed 128 characters. It cannot include a tab, newline character, or the following special characters: + * \ &lt; &gt; ; / % ? &amp; The name can include an ASCII space character except for the first and last character. All other space characters are not allowed. <strong>Note:</strong> Data Analyzer uses the user account name and security domain in the format UserName@SecurityDomain to determine the length of the user login name. The combination of the user name, @ symbol, and security domain cannot exceed 128 characters.</td>
</tr>
<tr>
<td>Password</td>
<td>Password for the user account. The password can be from 1 through 80 characters long.</td>
</tr>
<tr>
<td>Confirm Password</td>
<td>Enter the password again to confirm. You must retype the password. Do not copy and paste the password.</td>
</tr>
<tr>
<td>Full Name</td>
<td>Full name for the user account. The full name cannot include the following special characters: &lt; &gt; &quot; <strong>Note:</strong> In Data Analyzer, the full name property is equivalent to three separate properties named first name, middle name, and last name.</td>
</tr>
<tr>
<td>Description</td>
<td>Description of the user account. The description cannot exceed 765 characters or include the following special characters: &lt; &gt; &quot;</td>
</tr>
<tr>
<td>Email</td>
<td>Email address for the user. The email address cannot include the following special characters: &lt; &gt; &quot; Enter the email address in the format UserName@Domain.</td>
</tr>
<tr>
<td>Phone</td>
<td>Telephone number for the user. The telephone number cannot include the following special characters: &lt; &gt; &quot;</td>
</tr>
</tbody>
</table>

5. Click **OK** to save the user account.
CHAPTER 4

Ingesting Metadata into the Catalog

This chapter includes the following topics:

- Overview, 15
- Objectives, 16
- Creating a Resource, 16
- Configuring metadata and profiling settings, 18
- Running a Resource, 18
- Monitoring a Task, 18
- Browse Resources in the Catalog, 19

Overview

In this chapter, you will learn about how to ingest metadata into Enterprise Data Catalog by creating a resource and configuring various metadata extraction properties for the resource.

A resource is a catalog object that represents an external data source or metadata repository from where scanners extract metadata. A resource represents an instance of a specific resource type. The basic metadata operations, such as extraction and storage of metadata, are performed at the resource level.

A resource can be of resource types, such as relational databases, Business Glossary classification, and business intelligence sources. A resource might have an associated schedule. Each resource can extract both source metadata and profile metadata from the external data sources.

A scanner is a pluggable component of Enterprise Data Catalog that extracts specific metadata, such as source metadata or profile metadata, from external data sources and stores the metadata in the catalog. A scanner typically maps to a single resource type. However, there can be more than one scanner for a resource type. Examples are profiling scanner and lineage analyzer. A scanner performs a scan job on the metadata sources to fetch metadata into the catalog. When you have scanners for newer resource types ready, you can plug in those scanners to Enterprise Data Catalog without having to upgrade Enterprise Data Catalog.
Objectives

- Creating a resource
- Configuring metadata and profiling settings
- Running a resource
- Monitoring a task
- Browsing resources in the catalog

Creating a Resource

1. Log in to Catalog Administrator using the following URL: http://<host>:6705/ldadmin


3. Enter ResourceData in the Name box.

4. Optional. Enter a description for the resource in the Description box.
5. Select **Amazon S3** as the resource type from the **Select Resource Type** dialog box and click **Select**.

![Select Resource Type dialog box with Amazon S3 selected]

6. Enter the connection properties for the resource

**Connection Properties**

- **Amazon Web Services Bucket URL**: BucketName.s3.amazonaws.com *(Amazon Web Services URL to access a bucket)*
- **AWS Access Key ID**: *(Amazon Web Services access key ID to sign requests that you send to AWS)*
- **AWS Secret Access Key**: *(Amazon Web Services secret access key to sign requests that you send to AWS)*
- **Amazon Web Services Bucket Name**: *(Amazon Web Services bucket name that Live Data Map needs to scan)*
- **Source Directory**: *(The source directory on S3 from where metadata must be extracted)*

![Connection Properties dialog box with details filled in]

7. Click **Test Connection** to verify that the connection properties are correct.

8. Click **Next** to configure metadata and profiling settings for the resource.
Configuring metadata and profiling settings

Enter the details as shown in the following images to specify the metadata and profiling settings for the resource:

1. Click **Next** to go to the **Custom Attributes** page. You can use the page to bulk assign custom attributes to all objects imported from the catalog. In this session, we will skip assigning any custom attributes.

2. Click **Next** to go to the **Schedule** page. On this page, you can specify schedules for loading source metadata and profiling metadata. You can create a global schedule if required. Enterprise Data Catalog uses the schedules to run the metadata extraction and data profiling jobs at the specified frequency. For this session, we will skip schedule creation.

Running a Resource

▶ **Click Save and Run.**

Monitoring a Task

1. After scanners start extracting metadata from resources, you can monitor the resource statistics in the **Monitoring** tab. You can view a consolidated summary of resources with different task status updates. The monitoring statistics include the task type, resource name, schedule, task status, and the start and end time for the task.
2. Click **Refresh** to see the status of the tasks. You can also click the **Progress** tab to see the status of the tasks that are currently running.

3. After the metadata extraction task is complete, you can view the assets extracted by the Azure Microsoft SQL Server resource in the catalog.

**Browse Resources in the Catalog**

- You can browse the resources in the library, by clicking on **Open > Resources** or you can also search for the resource name.
Chapter 5

Associating Business Context with Assets

This chapter includes the following topics:

- Overview, 20
- Objectives, 20
- Associating Business Terms to Assets, 21
- Adding Custom Attributes to Catalog and Adding Values, 22

Overview

A Business Glossary comprises of online glossaries of business terms and policies that define important concepts within an organization. Analysts create and publish terms that include information such as descriptions, relationships to other terms, and associated categories. These terms can be associated to the related assets in Enterprise Data Catalog for classifying the assets. You can associate business terms from Business Glossary or Axon.

In this chapter, you will learn how to use business glossaries in the Enterprise Data Catalog that are created in Informatica Analyst. You will also learn how to associate business terms with assets, how to create custom attributes in the catalog and add values to the custom attribute. Finally, you will use the discovery features in Enterprise Data Catalog to search for assets using business terms and custom attributes.

Objectives

- Associating business terms to assets
- Adding Custom Attributes to Catalog and adding values
Associating Business Terms to Assets

1. Log in to Enterprise Data Catalog.
2. Search for a report and open the report.
3. Click **Associate a Business Term**.

4. Select a business term from the **Glossary** drop-down list, and choose **Certified**.

5. Click **OK**.
   You can now search for assets using business terms. For example, search for the business term **certified** to get the associated assets.
Adding Custom Attributes to Catalog and Adding Values

You can create custom attributes based on the search filters that you need to use in Enterprise Data Catalog where you search for metadata. Custom attributes help you quickly find specific metadata.

1. Log in to the Catalog Administrator using the following URL: `http://<host>:6705/ldmadmin` and select Manage > Attributes.

2. Select New from Attributes.

3. Enter a name and description for the custom attribute. For example, enter the name as `New_Attribute`.
   a. Choose String as the Data Type.
   b. Select Allow Filtering to use the custom attribute as a filter type.
   c. Search for Table from Object Types and select Table to enable the custom attribute to be assigned to relational tables.
4. Log in to Enterprise Data Catalog and click **Application Configuration**.

5. In the **Application Configuration** dialog box, go to **Custom Attributes** section and click **Add**.
6. Select the custom attribute you created in step 3 of this procedure and click **OK**.

7. Search for a relational table asset.
8. Open the relational table asset.
9. Click **Edit Custom Attribute**.
10. Select the new attribute, enter the text **Assignment** for the attribute, and click **OK**.

You can now search for the string **Assignment** and view the relational table asset in the search results.
CHAPTER 6

Discovering Assets Using Enterprise Data Catalog

This chapter includes the following topics:

- **Overview, 26**
- **Objectives, 26**
- **Search Features, 27**
- **Filter Search Results, 29**
- **Searching Using Business Terms, 31**
- **Searching Using Synonyms, 31**

Overview

In this chapter, you will learn how to discover assets using search and dynamic filtering capabilities in Enterprise Data Catalog. You will also learn to explore search by business terms and synonyms.

You can search for the required assets in the catalog. You can search based on the name of the asset or perform a generic search using the wildcard character asterisk (*) or the wildcard character question mark (?). Enterprise Data Catalog also displays probable matches when you type the name of a required asset.

From the search results displayed, you can sort the results based on the asset name, the relevance, or the system attributes or the custom attributes. You can use the search filters displayed to filter the search results and view additional details for the displayed assets. After searching and finding the required asset, you can annotate and enrich the required assets with custom attributes.

You can sort the displayed assets based on the name of the listed assets, the relevance, or the system attributes or the custom attributes selected.

Objectives

- Learn about search features such as suggestions, wildcards, and asset type inference
- Filter the search results
- Search using business terms
• Search using synonyms

Search Features

1. Launch the Enterprise Data Catalog login page using the following URL: http://<host>:6705/lmcatalog
2. Log in to Enterprise Data Catalog.
1. Type a term. For example, **customer**. Enterprise Data Catalog displays suggestions based on the assets in the catalog that match the search string.

2. You can also search for an asset using a wildcard.
3. Enterprise Data Catalog displays all assets that match the search string along with the asset type.

Filter Search Results

To filter the search results by resource name, for example, CRM, perform the following steps:

1. Click **Show All from Resource Name**.

2. Select a resource from the list of all resource names and click **OK**.
You can filter based on the asset type. For example, you can Select **Table** from **Asset Type** to filter the assets based on the table asset type.
You can also use one of the following search options to view the required details:

- Search using keywords such as tables, files, or reports. For example, if you search using the string **tables with customer**. Enterprise Data Catalog returns the following results:
  - Tables with the word customer in the name.
  - Tables with descriptions that contain the word customer.
  - Tables that have the associated business term customer.
  - Tables that contain columns with the word customer in the column name.
  - Tables that contain columns for which the customer data domain is inferred or assigned.

- Similarly, you can search using the keywords files or reports. For example, searching using **Sales Files** or **Sales Reports** lists the files or reports that include the term Sales.

### Searching Using Business Terms

You can search for assets associated with business terms, for example **Sale price**.

![Enterprise Data Catalog displaying search results for Sale price](image)

Enterprise Data Catalog displays the business term along with all associated assets.

### Searching Using Synonyms

You can use synonyms defined in Business Glossary to search for the required data asset. For example, search for **Cell** and press **ENTER**.
<table>
<thead>
<tr>
<th>Resource Name</th>
<th>Asset Type</th>
<th>Resource Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile</td>
<td>Mobile</td>
<td>Business Plan</td>
</tr>
<tr>
<td>Mobile</td>
<td>Mobile</td>
<td>Company Profile</td>
</tr>
<tr>
<td>Mobile</td>
<td>Mobile</td>
<td>Sales Data</td>
</tr>
</tbody>
</table>

Search Results

- Mobile
  - Asset Type: Mobile
  - Resource Type: Business Plan
  - Last Updated: Jan 10, 2021, 11:12 AM

- Mobile
  - Asset Type: Mobile
  - Resource Type: Company Profile
  - Last Updated: Jan 5, 2021, 10:43 PM

- Mobile
  - Asset Type: Mobile
  - Resource Type: Sales Data
  - Last Updated: Jan 4, 2021, 9:35 PM
CHAPTER 7

Viewing Data Lineage Diagrams

This chapter includes the following topics:

- Overview, 33
- Objectives, 33
- Using a Lineage Diagram to Perform Detailed Impact Analysis, 34
- Exporting a Lineage Diagram, 36

Overview

In this chapter, you will learn how to use the lineage feature in Enterprise Data Catalog to visualize data provenance and how to export lineage to a file. You will also learn how to use the detailed impact analysis reports in Enterprise Data Catalog to understand impact of changes in assets or ETL flows.

Lineage describes the flow of data from the origins to an asset. Lineage shows you where the data for an asset comes from and which assets affect the asset that you are studying. When you view an asset in a lineage and impact diagram, the lineage includes the asset that you are viewing and all of the upstream assets in the data flow.

Impact describes the flow of data from an asset to the destinations. Impact shows you where the data is used and which assets might be affected if you change the asset that you are studying. When you view an asset in a lineage and impact diagram, the impact includes the asset that you are viewing and all of the downstream assets in the data flow.

Objectives

- Use a lineage diagram to perform detailed impact analysis
- Export lineage diagrams
Using a Lineage Diagram to Perform Detailed Impact Analysis

The data flow for an asset includes two components:

- **Lineage.** Shows the flow of data from the origin to an asset.
- **Impact.** Shows the flow of data from an asset to the endpoints.

Perform the following steps to view the lineage and impact for an asset:

1. Type the name of a report in the Enterprise Data Catalog **Search** box and click the **Search** icon.
2. Click the report to view metadata associated with it.
3. Click the **Lineage and Impact** tab. The **Data Lineage** report provides a summarized view of lineage with endpoints.
4. Use the lineage sliders to add the required lineage levels in the lineage diagram.

Alternatively, you can also expand a specific lineage path by clicking the (+) icon on the link:

Click **Close (X)** to collapse individually expanded paths that appear on the side pane.
You can use the **Zoom In** icon to view further details in a specific portion of the diagram.

5. Click Show Business Terms to display business terms in the lineage diagram.

6. Click the plus icon next to the asset to view the metric and expand all columns in the lineage diagram that directly affect the metric.
7. Click the metric to view the details of the metric.

Exporting a Lineage Diagram

Click Export to export the lineage diagram.
Overview

In this chapter, you will learn how to use the Relationships view to identify the relationship between the selected asset and other assets in Enterprise Data Catalog. The view displays the asset relationships in a graphical form. The related assets that you view for a selected asset depend on the asset type. For example, if the selected asset is a column, the Relationships view shows all the data domains, similar columns, business term, and users that are related to the column. If the selected asset is a data domain, the Relationships view shows all the columns, business term, users, and data domain groups that are related to the data domain.

By default, the Relationships view displays the relationships for the selected asset. You can hover the cursor over any related asset in the Relationships view for additional details such as the asset type, resource type, and the time and date the asset was last updated.

Objective

Use a Relationship view to identify the relationship between the selected asset and other assets in the catalog.

Viewing Relationships

1. Open a report in Enterprise Data Catalog.
2. Click the Relationships tab.
3. The Relationships view displays the relationship that the report has with other assets in the catalog.

4. To expand the asset type circle, click the number icon at the top of the circle, and then click Show All to display all the assets of the same asset type which are included in the selected asset. When you click Show All, a dialog box appears, which provides you details about all the assets of the same asset type.

5. You can adjust the Relationships view to view specific assets or asset relationships in detail. You can zoom in, zoom out, reset to the original view size, and search for related assets.
Overview

Value frequency identifies the frequency of values in the columns. You can identify the frequency of values after you enable data similarity for a resource. Based on your business requirement, you can use the value frequency to analyze data in a resource. You can infer value frequency from relational tables and flat files in resources.

Objectives

- Enabling value frequency for resources
- Viewing value frequencies in Enterprise Data Catalog

Enabling Value Frequency for Resources

Configure the column similarity properties to prepare data to identify similar columns based on the source data and to compute the frequency of values in the columns.
The following table describes the property that you can configure in the Similarity Profile and Value Frequency Settings section of the Metadata Load Settings tab for a resource:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profiling Run Option</td>
<td>Select Data preparation for Similarity and Value frequency option.</td>
</tr>
</tbody>
</table>

**Viewing Value Frequencies in Enterprise Data Catalog**

You can view the value frequency for view column, table column, CSV field, XML file field, and JSON file data assets in the Asset Details view. In the value frequency section, you can view the distribution of values across the data asset, the maximum and minimum values in the data asset, the distribution of distinct values in the data asset, and a list of values and their frequency in a tabular format.

The value distribution chart for a data asset appears by default after you run the similarity profile and value frequency on the data asset in the Catalog Administrator. You can view the value frequency section after the administrator assigns the Data Privileges: View Data privilege in the Administrator tool and you have data read permission for the data asset. If the data asset has sensitive data, then you can view the sensitive data in the data asset after the administrator assigns the Data Privileges: View Sensitive Data privilege. For more information about privileges and permissions, see the Informatica Administrator Reference for Catalog guide.

The following image displays the value frequency section in the Asset Details view:

1. Value distribution chart. Displays the number of NULL values, distinct values, and non-distinct values in the first 1000 rows in the data asset.

2. Maximum and minimum value. Displays the maximum value and minimum value in the data asset.
3. **Slider.** Displays the values in the slider along with the frequency and percentage of values in a tabular format below the graph.

The value distribution for the first 1000 rows in the data asset appears as a horizontal bar chart. The horizontal bar chart displays the distribution of NULL values, distinct values, and non-distinct values in different colors. When you move the pointer over the chart, the number of rows and the percentage of values pertaining to the data point appears in a data label. A data label is similar to a tool tip and it displays the data that pertains to the data point in the chart over which you hover the pointer.

After you run the value frequency for a data asset, you can view the following information about the data asset in the value frequency section:

- **Max.** Displays the maximum value in the data asset. After you click Max, the maximum value along with the frequency and percentage appears in a tabular format below the value frequency graph.
- **Min.** Displays the minimum value in the data asset. After you click Min, the minimum value and the next 19 values in the data asset along with the frequency and percentage appear in a tabular format below the value frequency graph.
- **Value frequency graph.** The graph displays the distribution of distinct values in the data asset as vertical bars in the column chart. The X-axis in the column chart represents values, and the Y-axis represents the frequency of values. The values are arranged from maximum to minimum number of occurrences in the chart. The chart displays the first 1000 distinct values in the data asset and has a slider. The values within the slider appear in a tabular format below the value frequency graph along with the frequency and percentage of the value.
- **Values in tabular format.** The values along with the frequency and percentage appear in a tabular format. The values within the slider appear in the table or after you click the Max or Min value. At a time, you can view only 20 values in the table.
Overview

You can display the transformation logic for assets in the Lineage and Impact view in Enterprise Data Catalog. A transformation is a repository object that generates, modifies, or passes data from source to target connections.

You can choose to display the transformation logic for assets that contain transformations in Lineage and Impact view.

Objectives

- Enabling transformation logic for resources
- Viewing transformation logic for assets

Enabling Transformation Logic for Resources

You can enable transformation logic for resources by selecting the Detailed Lineage property in the Metadata and Profiling Settings tab for a resource, in Catalog Administrator.
Viewing Transformation Logic for Assets

You can display the transformation logic for assets in the Lineage and Impact view in Enterprise Data Catalog. A transformation is a repository object that generates, modifies, or passes data from source to target connections.

You can choose to display the transformation logic for assets that contain transformations in Lineage and Impact view.

When you click the transformation logic button in the lineage and impact view, the view displays transformation icons on the assets that contain transformations. Assets that display transformation icons in Lineage and Impact view are target assets in the transformation logic view. Lineage and Impact view displays transformation logic for resource types such as PowerCenter, Informatica Cloud Services, and Cloudera Navigator. The transformation icon for PowerCenter and Informatica Cloud Services resource types indicate that a mapping exists between source assets and target assets. A mapping is a set of inputs and outputs that represent the data flow between sources and targets. A mapping contains components such as source objects, target objects, and transformations. The transformation icon for Cloudera Navigator resource type indicates that a data flow relation type exists between source assets and target assets.

Transformation logic view displays all the mappings or data flow relation types between source assets and target assets for the resource type that you select. The transformation logic view, by default, displays the source assets, transformations, and target assets. The view also displays the properties panel to view details about assets in a mapping or data flow relation type.

If an asset contains multiple transformation logic between the source assets and target assets, the lineage and Impact view displays a drop-down menu when you hover over the transformation icon. The drop-down menu lists all the resource types that contain transformation logic. You can select a resource from the drop-down menu for which you want to view the transformation logic.

The transformation view displays transformation logic for data structures such as tables and columns. The view displays various types of transformations such as filter, joiner, lookup, expression, sorter, union, and aggregate.

The following image shows a sample transformation logic for a relational table in the Lineage and Impact view:
1. Source asset
2. Transformation logic
3. Target asset
4. Menu
5. Fit to window
6. Zoom controls
7. Show business terms
This chapter includes the following topics:

- **Overview, 45**
- **Objectives, 45**
- **Creating Rule-based Data Domains, 46**
- **Running Data Domain Discovery for a Resource, 46**
- **Enabling Smart Data Domains, 47**

### Overview

In this chapter, you will learn how to manage data domains using Catalog Administrator. Managing data domains involves creating new data domains. You will also learn how to use column data similarity and smart data domain inference to tag data domains.

A data domain is a predefined or user-defined Model Repository object based on the semantics of column data or a column name. Examples include Social Security number, phone number, and credit card number. You use rules to define data and column name patterns that match source data and metadata and create data domains in Catalog Administrator. A data domain helps you find important data that remains undiscovered in a data source.

Data similarity refers to identification of similar columns based on source data available in an enterprise. As a data analyst or a data architect, you can scan your enterprise data to find similar data and then attach data domains to similar data patterns. This process helps you to search and discover assets of interest in the catalog faster.

If you enable data domain discovery on resources, Enterprise Data Catalog uses defined data domains to infer matching column data or column name patterns from metadata extracted by resources. After identifying matching column data or column names, Enterprise Data Catalog marks the inferred data domains in the catalog for approval or rejection. You can approve or reject inferred data domains (smart data domains) in Enterprise Data Catalog.

### Objectives

- Creating rule-based data domains
- Running data domain discovery for a resource
Creating Rule-based Data Domains

1. Log in to Catalog Administrator.
2. Click New > Data Domain.
3. Enter CameoCategory in the Domain Name box.
4. Enter Identifies Cameo Category Codes in the Description box.
5. Select Yes for the Specify Rules for Data Domain option.
6. Specify a data rule for the data domain.
7. Click Save.

Running Data Domain Discovery for a Resource

1. Log in to Catalog Administrator.
2. Open a resource from the Library.
3. Click Edit.
4. Click the Metadata Load Settings tab.
5. Click Select and add the required data domains. For example, CameoCategory from the Data Domains dialog box.
6. Click **Save and Run**.

7. Log in to Enterprise Data Catalog after the resource run is complete.

8. Type **CameoCategory** in the **Search** box and click **Search**. The list of inferred data domains and associated columns appear in the search results.

9. Open the associated columns in Enterprise Data Catalog and curate the domains by clicking the tick icon on the inferred domain.

---

**Enabling Smart Data Domains**

1. Search for a column in a table in Enterprise Data Catalog.

2. Open the column and confirm that the column is from the metadata source from which you extracted metadata.

3. Click **Edit** in the **Data Domain** section.
4. In the Add Domain field, enter CustomerTier as the name of the data domain and click Add.

5. Optional. Enter a description for the data domain.
6. Log in to Catalog Administrator.
7. Click Open > Resources.
8. Open the required resource and run the resource.

9. Monitor the resource run process to completion.

After the resource run is complete, Enterprise Data Catalog propagates the CustomerTier domain to all similar columns. Searching for CustomerTier displays all the similar columns in Enterprise Data Catalog.
Chapter 12

Searching Using Composite Data Domains

This chapter includes the following topics:

- **Overview, 50**
- **Objectives, 50**
- **Creating a Composite Data Domain, 51**
- **Enabling Composite Data Domain Discovery for Resources, 51**
- **Searching for Assets Using Composite Data Domains, 51**

Overview

In this chapter, you will learn about composite data domains and how to create and manage composite data domains.

A composite data domain is a collection of data domains or other composite data domains linked using rules. A composite data domain helps you search for the required details of an entity across multiple schemas defined for the database. An entity refers to a particular subject for which you need all the associated information. For example, for an entity such as customer, you might want to search for all the associated details such as the customer name, age, location, and contact number. The details might be spread across multiple tables and columns in the database. Composite data domains help you define a search query that includes the entity details that you want to find from multiple schemas defined for the database.

After you create composite data domains and enable composite data domain discovery for resources, you can search for the entity using composite data domains in Enterprise Data Catalog. Enterprise Data Catalog finds the assets associated with the entity in the catalog.

Objectives

- Creating a composite data domain
- Enabling composite data domain discovery for resources
- Searching for assets using composite data domains
Creating a Composite Data Domain

As a catalog administrator, make sure that you identify the existing data domains, create additional data domains or composite data domains if required for the new composite data domain, and understand the combination of rules required to create the composite data domain. Make sure that data domains or other composite data domains exist before you create a new composite data domain.

2. In the Name box, specify the name for the composite data domain as cameo.
3. In the Description box, specify the description as Cameo category search.
4. Specify the data domains that you want to include with the composite data domain in the Specify Match Criteria section. You can click Select Data Domain to select the required data domains from the Data Domains dialog box. Select the CameoCategory data domain that you created.
5. Click Save.

Enabling Composite Data Domain Discovery for Resources

As a one-time prerequisite step, verify that you run profiling before you run composite data domain discovery. Alternatively, Enterprise Data Catalog discovers composite data domains if inferred data domains exist in your environment. Enterprise Data Catalog discovers data domains that are not enabled for discovery, during the composite data domain discovery.

The following options in the Composite Domain Discovery section in the Metadata and Profiling Settings tab for a resource help you to enable and configure composite data domain discovery for the resource:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable Composite Domain Discovery</td>
<td>Select the option to enable composite data domain discovery for the resource.</td>
</tr>
<tr>
<td>Select Composite Data Domain</td>
<td>Select the composite data domains that the resource must use for inference. Select the following option to specify the composite data domain: Specific Composite Data Domains. Selects specific composite data domains. When you select this property, the Composite Data Domains text box appears. Click Select... to select the CameoCategory data domain from the Select Composite Data Domains dialog box.</td>
</tr>
</tbody>
</table>

Note: If you enable composite data domain discovery and data domain discovery for the resource, Enterprise Data Catalog discovers all data domains included in the composite data domains for the resource.

Searching for Assets Using Composite Data Domains

The Asset Details view in Enterprise Data Catalog displays the Composite Data Domains section if tabular assets in resources have associated composite data domains.
To search for assets using the cameo composite data domain, perform the following steps:

1. Click Add in the Composite Data Domains section of the Filter By panel in Enterprise Data Catalog.
2. Type cameo and press Enter. Enterprise Data Catalog refreshes the search results based on the definitions specified in the composite data domain, cameo.

The Asset Details view for a composite data domain displays information, such as the asset type, resource that contains the composite data domain, and date that the composite data domain was last updated in the catalog. The view also displays specific details, such as the overview, custom attributes, and resources where the composite data domain is found.