



Informatica® Metadata Command Center
November 2025

Microsoft SharePoint Online Sources

© Copyright Informatica LLC 2023, 2025

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.

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation is subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License.

Informatica, Informatica Cloud, Informatica Intelligent Cloud Services, PowerCenter, PowerExchange, 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. Required third party notices are included with the product.

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 AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2025-11-19

Table of Contents

Preface.	4
Chapter 1: Introduction to Microsoft SharePoint Online catalog sources.	5
Extraction and view process.	6
About the Microsoft SharePoint Online catalog source.	6
Extracted metadata.	7
Chapter 2: Before you begin.	8
Verify permissions.	8
Permissions for metadata extraction.	8
Permissions to run data classification.	8
Permissions to run glossary association.	8
Verify authentication.	9
Access Control Service.	9
Microsoft Entra ID	12
Create a connection.	14
Access Control Service Authentication.	15
Microsoft Entra ID Authentication.	15
Chapter 3: Create catalog sources in Metadata Command Center.	17
Step 1. Register a catalog source.	18
Step 2. Configure capabilities.	19
Configure metadata extraction.	19
Filter guidelines and examples.	22
Configure data classification.	24
Configure glossary association.	24
Step 3. Associate stakeholders and asset groups.	25
Step 4. Run or schedule the job.	27
Step 5. Connect to referenced source systems.	28
Chapter 4: View results in Data Governance and Catalog.	30
View metadata extraction results.	31
View data lineage.	32
View lineage at the catalog source level.	32
View lineage at the data set level.	33
View lineage at the data element level.	33
View classified data.	33
View glossary associations.	34

Preface

Read *Microsoft SharePoint Online Sources* to learn how to register and configure Microsoft SharePoint Online sources in Metadata Command Center as catalog sources. After you configure a catalog source, you extract metadata and then view the results in Data Governance and Catalog.

CHAPTER 1

Introduction to Microsoft SharePoint Online catalog sources

You can use Metadata Command Center to extract metadata from a source system.

A source system is any system that contains data or metadata. For example, Microsoft SharePoint Online is a source system from which you can extract metadata through a Microsoft SharePoint Online catalog source with Metadata Command Center. A catalog source is an object that represents and contains metadata from the source system.

Before you extract metadata from a source system, you first create and register a catalog source that represents the source system. Then you configure capabilities for the catalog source. A capability is a task that Metadata Command Center can perform, such as metadata extraction, lineage discovery, data profiling, data classification, or glossary association.

When Metadata Command Center extracts metadata, Data Governance and Catalog displays the extracted metadata and its attributes as technical assets. You can then perform tasks such as analyzing the assets, viewing lineage, and creating links between those assets and their business context.

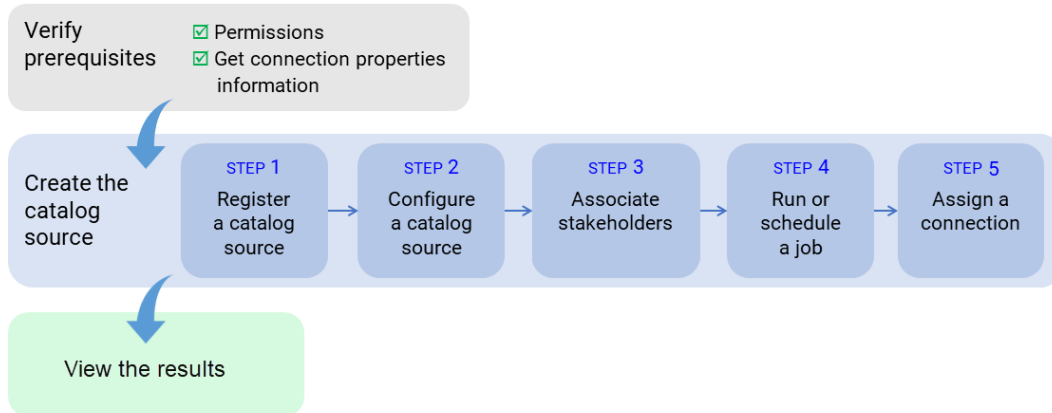
The following table describes the capabilities of the catalog source:

Capability	Description
Incremental metadata extraction	An incremental metadata extraction extracts only the changed and new objects since the last catalog source job run. Incremental metadata extraction doesn't remove deleted objects from the catalog and doesn't extract metadata of code-based objects if applicable.
Data Classification	Data classification is the process of identifying and organizing data into relevant categories based on the functional meaning of the data. Classifying data can help your organization manage risks, compliance, and data security.
Glossary Association	You can associate terms that are in the glossary with technical assets to provide user-friendly business names to technical assets. Glossary Association automatically associates glossary terms with technical assets or recommends glossary terms that you can manually associate with technical assets in Data Governance and Catalog.

Extraction and view process

To extract metadata from a source system, configure the catalog source and run the extraction job in Metadata Command Center. Then view the results in Data Governance and Catalog.

The following image shows the process to extract metadata from a source system:



After you verify prerequisites, perform the following tasks to extract metadata from Microsoft SharePoint Online:

1. Register a catalog source. Create a catalog source object, select Microsoft SharePoint Online, and specify values for connection properties.
2. Configure the catalog source. Specify the runtime environment and configure parameters for metadata extraction. Optionally, add filters to include or exclude source system assets from metadata extraction. You can also configure other capabilities such as data profiling and quality, data classification, or glossary association.
3. Optionally, associate stakeholders. Associate users with technical assets, giving the users permission to perform actions determined by their roles.
4. Run or schedule the catalog source job.
5. Optionally, if the catalog source job generates referenced asset objects, you can assign a connection to referenced source system assets.
You can view the lineage with object references without performing connection assignment. After connection assignment, you can view the objects.

After you run the catalog source job, you view the results in Data Governance and Catalog.

About the Microsoft SharePoint Online catalog source

You can use the Microsoft SharePoint Online catalog source to extract metadata from the Microsoft SharePoint Online source system.

Microsoft SharePoint Online is a web-based cloud service that integrates with Microsoft Office. It provides a platform to collaborate and allows users to store, share, and manage content.

Extracted metadata

You can use the Microsoft SharePoint Online catalog source to extract metadata from the Microsoft SharePoint Online source system.

Metadata Command Center extracts the following metadata from a Microsoft SharePoint Online source system:

- Site
- Subsite
- Document library
- Folder
- Hierarchical file
- Flat File
- XML File
- XSD File
- Attribute
- Element
- Hierarchical field
- Flat field

You can extract workbooks, worksheets, and columns from Microsoft Excel files.

The following table lists the structures associated with the file types that you can extract metadata from:

File Type	Partition structure
AVRO	Single partition, multiple partitions, schema merge
CSV	Single partition, multiple partitions, schema merge
JSON	Single partition, multiple partitions, schema merge
Parquet	Single partition, multiple partitions, schema merge
XML	Single partition, multiple partitions, schema merge

Excel file types

You can extract metadata from the following Microsoft Excel file types:

- Excel 97-2003 Workbook with XLS extension
- Excel Workbook with XLSX extension
- Excel Macro-Enabled Workbook with XLSM extension

CHAPTER 2

Before you begin

Before you can extract Microsoft SharePoint Online catalog source metadata, complete prerequisite tasks.

Ensure that the following prerequisites are met:

- Verify permissions
- Verify authentication
- Create a connection in Administrator

Verify permissions

To extract metadata and to configure other capabilities that a catalog source might include, you need account access and permissions on the source system. The permissions required might vary depending on the capability.

Permissions for metadata extraction

To extract Microsoft SharePoint Online metadata, you need account access and permissions to the Microsoft SharePoint Online source system.

Verify that the administrator performs the following tasks:

- Can connect to the Microsoft SharePoint Online account to access the Microsoft SharePoint Online application.
- Configures the Microsoft SharePoint Online connector.
- Grants read permission for Site and its contents, such as Subsite and Document Library.

Permissions to run data classification

You can perform data classification with the permissions required to perform metadata extraction.

Permissions to run glossary association

You can perform glossary association with the permissions required to perform metadata extraction.

Verify authentication

You can configure Access Control Service and Microsoft Entra ID authentication types to access Microsoft SharePoint Online.

Before you configure the connection properties, ensure that you perform the prerequisite steps based on the type of authentication that you select.

Access Control Service

In Microsoft SharePoint Online, you can register applications in Access Control Service for app-only access, and the administrator can restrict site access through the SharePoint admin center.

Generate the Client ID and Client Secret

The client ID and client secret are required to generate a valid access token.

Perform the following steps to generate the client ID and client secret:

1. Log in to the Microsoft Sharepoint Online account.
2. Enter the following site or subsite URL:

Site: `https://<sitename.com>/_layouts/15/appregnew.aspx`

Subsite: `https://<sitename.com>/<subsitedomain>/_layouts/15/appregnew.aspx`

The **App Information** page appears.

3. Click **Generate** next to the **Client Id** field.

The value of the client ID is displayed in the **Client Id** field. The following image shows the **App Information** page where you can generate the values of the client ID and client secret:

Client Id:

Client Secret:

Title:

App Domain:
Example: "www.contoso.com"

Redirect URL:
Example: "https://www.contoso.com/default.aspx"

4. Click **Generate** next to the **Client Secret** field.

The value of the client secret is displayed in the **Client Secret** field.

5. Enter an appropriate title for the App in the **Title** field.
6. Enter an app domain name in the **App Domain** field.

For example, `www.google.com`

7. Enter a URL in the **Redirect URL** field.

For example, `https://localhost/`. You must enter the same redirect URL in the connection property.

8. Click **Create**.

The page redirects to the Microsoft Sharepoint Online page and the following message appears:

The app identifier has been successfully created.

The values of the client Id, client secret, title, and redirect URL are displayed.

Generate the Bearer Realm

A bearer realm is a unique ID provided for each user. You must generate the bearer realm to obtain the authorization code.

Perform the following steps to generate the bearer realm:

1. Open the Google PostMan application.
2. Enter the following site or subsite URL in the Google PostMan application:

Site: `https://<sitename.com>/_layouts/15/appregnew.aspx`

Subsite: `https://<sitename.com>/<subsiteidomain>/_layouts/15/appregnew.aspx`

The following image shows the **BearerToken** page where you can generate the value of the bearer realm:



3. Select the **GET** method.
4. On the **Headers** tab, enter **Authorization** in the **Key** field and **Bearer** in the **Value** field.
5. Click **Send**.
6. Select the **Headers** tab in the **Response** header.

The bearer realm value appears in the **WWW-Authenticate** section. For example:

`Bearer realm="77baf95d-f3e0-42b-aa08-9b798b8c177b"`

Generate the Authorization Code

You must generate the authorization code to gain access to the current site and to generate a valid refresh token.

Perform the following steps to generate the authorization code:

1. Enter the following site or subsite URL in the Google chrome browser:

Site: `https://<site.sharepoint.com>/_layouts/15/OAuthAuthorize.aspx?`

`client_id=<client_GUID>&scope=<app_permissions_list>&response_type=code&redirect_uri=<redirect_uri>`

For example, `https://icloudconnectivitydev.sharepoint.com/_layouts/15/oauthauthorize.aspx?`

`client_id=ecea5b1b-80e4-4f3e-a269-48b85c1797a8&`

`scope=AllSites.Manage&response_type=code&redirect_uri=https%3A%2F%2Flocalhost%2F`

Subsite: `https://<site.sharepoint.com>/<subsiteidomain>/_layouts/15/OAuthAuthorize.aspx?`

`client_id=<client_GUID>&scope=<app_permissions_list>&response_type=code&redirect_uri=<redirect_uri>`

For example, `//informaticaone.sharepoint.com/sites/TEST/_layouts/15/oauthauthorize.aspx?`

`client_id=ecea5b1b-80e4-4f3e-a269-48b85c1797a8&`

`scope=AllSites.Manage&response_type=code&redirect_uri=https%3A%2F%2Flocalhost%2F`

Microsoft Entra ID

You can use Microsoft Entra ID to securely access and manage Microsoft SharePoint Online data.

Register the Azure application with Azure Active Directory

To establish a connection with Microsoft SharePoint Online, you need to provide the Microsoft SharePoint Online client ID and client secret in the connection properties.

You can get the client ID and client secret by registering your application in Azure Active Directory (AAD) through the Microsoft Identity platform.

1. Log in to portal.azure.com using your Microsoft SharePoint Online credentials.
2. Go to **App Registrations** in the Azure Services section.
3. Click **New Registration**.
4. Specify a display name for your application and supported account type, enter the redirect URI and then click **Register**.

Ensure that you select either the Single tenant or Multitenant account type. You can't use the personal Microsoft account type.

A client ID is generated. Ensure that you copy the client ID and keep it handy to use when you generate an authorization code and configure a Microsoft SharePoint connection.

5. Click **Add a Certificate or Secret**.
6. Click **New client secret**, and then add the description and the expiry time.

A client secret value is generated. Ensure that you copy the client secret from the Value column and keep it handy to use when you generate an authorization code and configure a Microsoft SharePoint connection.

Certificates (0) Client secrets (1) Federated credentials (0)

A secret string that the application uses to prove its identity when requesting a token. Also can be referred to as application password.

+ New client secret

Description	Expires	Value ⓘ	Secret ID
kk_secret	9/29/2025	z4~*****	z4~*****

7. Now, click **API permissions** in the left pane.
8. Click **Add a permission**.
9. Click **SharePoint**, and then click **Delegated permission** on the Request API permissions page.
10. Select the permissions that the client application must have on behalf of the signed-in user.

The following list outlines the permissions and the levels of access each permission provides:

- **AllSites.FullControl**. Full Control access.
- **AllSites.Manage**. Read and write access.
- **AllSites.Read**. Read access.
- **AllSites.Write**. Write access.

Consider selecting the AllSites.Manage permission to ensure appropriate access to Microsoft SharePoint Online.

11. Click **Add Permissions**.

Generate the authorization code

To generate an authorization code, select the GET method with the necessary query parameters, and retrieve the code from the redirect URL after you authenticate to the SharePoint Online application.

1. Open the PostMan application.
2. In Postman, enter one of the following URLs based on your account type:
 - For a single tenant account, enter the following URL: `https://login.microsoftonline.com/<Single_Tenant_Id_value>/oauth2/v2.0/authorize`
 - For a multi-tenant account, enter the following URL: `https://login.microsoftonline.com/organizations/oauth2/v2.0/authorize`

Replace `<Single_Tenant_Id_value>` with the tenant ID found in the overview section of your registered application if you are working with a single tenant account. For multi-tenant accounts, use the organizations endpoint.

3. Select the **GET** method.
4. On the **Params** tab, enter the name and value.

To authenticate and verify access permissions, enter the following query parameters:

```
client_id=<client_id> &response_type=code &redirect_uri=<redirect_URI>
&scope=<sharepoint_url>/<delegated permission> offline_access
&client_secret=<client_secret_value>
```

The scope query contains delegated permissions for your Azure application. If you selected **AllSites.Manage** as the delegated permission when you registered the Azure application with Azure Active Directory, specify the permission in the scope query parameter as shown in the following example:

```
client_id=<client_id> &response_type=code &redirect_uri=<redirect_URI>
&scope=<sharepoint_url>/AllSites.Manage offline_access
&client_secret=<client_secret_value>
```

5. Copy the URL and paste it in the browser.
6. Enter the SharePoint Online log in credentials.
7. Verify and click **Accept** on the consent screen.

The redirect URL page includes the authorization code as a query string in the following format:

```
https://<redirect_url>/?code=<authcode>
```

Ensure that you copy the authorization code and keep it handy to use when you generate a refresh token.

Generate the refresh token

Generate the refresh token in the PostMan application.

1. In the PostMan application, enter one of the following URLs based on your account type:
 - For a single tenant account, enter the following URL: `https://login.microsoftonline.com/<Single_Tenant_Id_value>/oauth2/v2.0/token`
 - For a multi-tenant account, enter the following URL: `https://login.microsoftonline.com/organizations/oauth2/v2.0/token`

Replace `<Single_Tenant_Id_value>` with the tenant ID found in the overview section of your registered application if you are working with a single tenant account. For multi-tenant accounts, use the organizations endpoint.

2. Select the **POST** method.

3. On the **Header** tab, enter **Content-Type** in the **Key Name** field and **application/x-www-form-urlencoded** in the **Value** field.

4. On the **Body** tab, enter the XML request in the following format:

```
grant_type=authorization_code &client_id=<client_id>&client_secret=<client_secret_value>
&code=<auth_code> &redirect_uri=<redirect_url>
```

You need to enter the client ID and client secret that you generated when you registered the Azure application with Azure Active Directory.

5. Click **Send**.

The refresh token is generated on the **Response** tab.

Ensure that you copy the refresh token and keep it handy to use when you configure a Microsoft SharePoint Online connection.

Create a connection

Before you configure the Microsoft SharePoint Online catalog source, create a connection object in Administrator.

1. In Administrator, select **Connections**.
2. Click **New Connection**.
3. Enter the following connection details:

Property	Description
Connection Name	Name of the connection. Each connection name must be unique within the organization. Connection names can contain alphanumeric characters, spaces, and the following special characters: _ . + -, Maximum length is 255 characters.
Description	Description of the connection. Maximum length is 4000 characters.
Use Secret Vault	Stores sensitive credentials for this connection in the secrets manager that is configured for your organization. This property appears only if secrets manager is set up for your organization. When you enable the secret vault in the connection, you can select which credentials that the Secure Agent retrieves from the secrets manager. If you don't enable this option, the credentials are stored in the repository or on a local Secure Agent, depending on how your organization is configured. Note: If you're using this connection to apply data access policies through pushdown or proxy services, you cannot use the Secret Vault configuration option. For information about how to configure and use a secrets manager, see "Secrets manager configuration" in the Administrator help.
Runtime Environment	The name of the runtime environment where you want to run tasks. For more information about how to configure and use the runtime environments, see <i>Runtime Environments</i> in the Administrator help.

4. Select the authentication type to connect to Microsoft SharePoint Online and enter the required properties. You can use the following authentication types:
 - Access Control Service
 - Microsoft Entra ID
5. Click **Test Connection**.

Access Control Service Authentication

You can use the Access Control Service authentication to access the SharePoint API.

The following table describes the basic connection properties for Access Control Service authentication:

Property	Description
Account types	The tenant that you want to use to access the application. Select None .
Single tenant id	Required only when you select the Single tenant account type. The unique ID of the organization to manage and control access to resources, applications, devices, and services.
Client_Id	Client ID of Microsoft SharePoint Online required to generate a valid access token.
Client_Secret	Client secret of Microsoft SharePoint Online required to generate a valid access token.
Refresh_Token	Refresh token of Microsoft SharePoint Online.
Redirect_URL	URL where you want to redirect from the Microsoft SharePoint Online account.
URL	URL to the Microsoft SharePoint Online account.
Attachment_File_Path	Directory on the Secure Agent machine where you want to download or attach files to Microsoft SharePoint Online.

The following table describes the advanced connection properties for Access Control Service authentication:

Property	Description
Subsite_URL	URL of the Microsoft SharePoint Online account within the Microsoft SharePoint site. Enter the subsite URL if you have organized data and set up subsite accounts in the Microsoft SharePoint Online application. For more information about sites and subsites in Microsoft SharePoint Online account, see Create sites and subsites . If you do not enter a subsite URL, the Microsoft SharePoint Online Connector reads files from the URL that you specify in the URL property.

Microsoft Entra ID Authentication

You can use the Microsoft Entra ID authentication to access Microsoft SharePoint resources securely.

The following table describes the basic connection properties for Microsoft Entra ID authentication:

Property	Description
Account types	The tenant that you want to use to access the application. Select from the following options: <ul style="list-style-type: none">- Single tenant. Select if your target audience is inside your organization.- Multi tenant. Select if your target audience includes businesses or educational customers and requires multi-tenancy support.- Default is None.
Single tenant id	Required only when you select the Single tenant account type. The unique ID of the organization to manage and control access to resources, applications, devices, and services.
Client_Id	The client identifier issued during the application registration process. You can get the client ID by registering your application in Azure Active Directory (AAD) through the Microsoft Identity platform.
Client_Secret	The client secret issued during the application registration process. You can get the client secret by registering your application in Azure Active Directory (AAD) through the Microsoft Identity platform.
Refresh_Token	Refresh token of Microsoft SharePoint Online.
Redirect_URL	Does not apply to Microsoft Entra ID authentication.
URL	URL to the Microsoft SharePoint Online account.
Attachment_File_Path	Directory on the Secure Agent machine where you want to download or attach files to Microsoft SharePoint Online.

The following table describes the advanced connection properties for Microsoft Entra ID authentication:

Property	Description
Subsite_URL	URL of the Microsoft SharePoint Online account within the Microsoft SharePoint site. Enter the subsite URL if you have organized data and set up subsite accounts in the Microsoft SharePoint Online application. For more information about sites and subsites in Microsoft SharePoint Online account, see Create sites and subsites . If you do not enter a subsite URL, the Microsoft SharePoint Online Connector reads files from the URL that you specify in the URL property.

CHAPTER 3

Create catalog sources in Metadata Command Center

Use Metadata Command Center to configure a catalog source for Microsoft SharePoint Online and run the catalog source job.

When you configure a catalog source, you define the source system that you want to extract metadata from. Configure filters to include or exclude source system metadata before you run the job. Optionally, configure other capabilities, such as lineage discovery, data profiling and quality, data classification, relationship discovery, and glossary association.

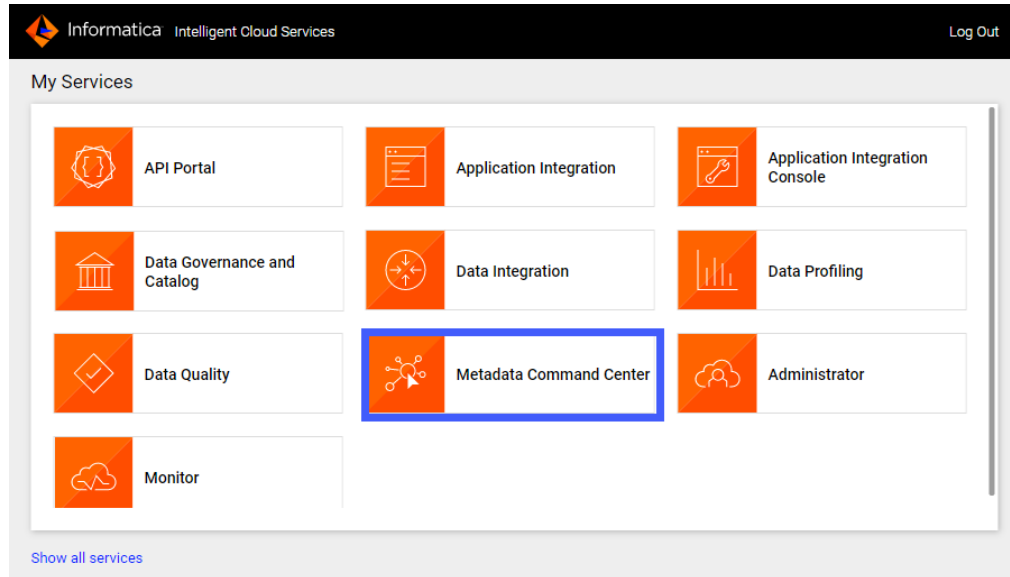
To provide stakeholders access to technical assets, you can assign access through stakeholder roles. You can also associate technical assets extracted from the catalog source to asset groups. If your catalog source references other source systems, you can create a connection assignment to the endpoint catalog source to view complete lineage.

Step 1. Register a catalog source

When you register a catalog source, provide general information and connection values.

1. Click **Metadata Command Center**.

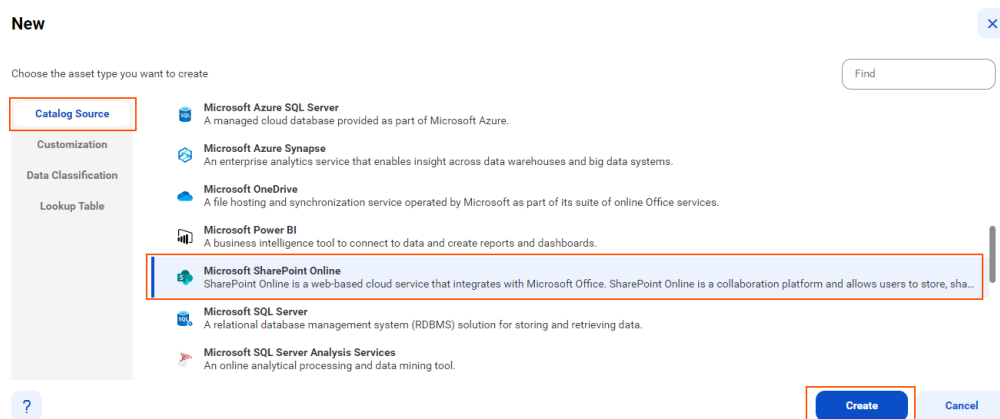
The following image shows the Metadata Command Center box on the **My Services** page:



The Metadata Command Center home page appears.

2. Click **New**.
3. Select **Catalog Source** from the list of asset types.
4. Select **Microsoft SharePoint Online** from the list of catalog source types.
5. Click **Create**.

The following image shows where you choose the source system:



The **New Catalog Source** page opens.

6. In the **General Information** section, enter a name and an optional description for the catalog source.

Note: You can rename a catalog source after you create it, but to apply the change to all associated objects you must rerun the metadata extraction job.

After you save the catalog source, you can update the description in Metadata Command Center and Data Governance and Catalog. The update appears only in the service in which you update it.

7. In the **Connection Information** area, select the connection that you created in Administrator.

Note: To create or edit a catalog source, you need permissions on the connection to the source system. Select a connection that you have access to, or ask the administrator to grant the necessary permissions to the connection that you want to use.

8. Click **Connection Properties** to expand and view the connection properties for the selected connection.
9. Click **Test Connection** to test your connection to the source system.
10. Click **Next**.

The **Configuration** page appears.

Step 2. Configure capabilities

When you configure the Microsoft SharePoint Online catalog source, you define the settings for the metadata extraction capability and other optional capabilities.

The metadata extraction capability extracts source metadata from external source systems. You can also configure other capabilities that the catalog source includes.

You can save the catalog source configuration at any point after you enter the connection information. After you save the catalog source, you can choose to run the catalog source job. To run the job once, click **Run**. To run metadata extraction and other capabilities on a recurring schedule, configure schedules on the **Schedule** tab.

Configure metadata extraction

When you configure the Microsoft SharePoint Online catalog source, you choose a runtime environment, define filters, and enter configuration parameters for metadata extraction.

1. In the **Connection and Runtime** area, choose a serverless runtime environment or the Secure Agent group where you want to run catalog source jobs.

Note: Serverless runtime environment options are available if the catalog source works with a serverless runtime environment.

2. Choose to retain, delete, or deprecate objects that are deleted from the source system in the catalog with the **Metadata Change Option**.
 - **Retain.** Retains objects that are deleted from the source system in the catalog. If you update or add a filter, the catalog retains objects extracted from the previous job and extracts additional objects that match the current filter. Objects deleted from the source system are not deleted from the catalog. Enrichments added on deleted objects and relationships are retained.
 - **Delete.** Deletes metadata from the catalog based on objects deleted from the source system and changes you make to the filter. Enrichments added on deleted objects and relationships are also permanently lost. Objects renamed in the source system are removed and recreated in the catalog.

- **Deprecate.** The lifecycle of objects imported into the catalog moves to Obsolete based on objects deleted from the source system and changes you make to the filter. This does not impact enrichments added on deprecated objects and relationships. Objects renamed in the source system are removed and recreated in the catalog. When you run the catalog source job again for other capabilities such as data classification, relationship discovery, or glossary association, the job doesn't consider obsolete objects. Obsolete objects remain in the catalog until they are purged when you run a **Purge Obsolete Objects** job on the **Explore** page.

Note: You can also change the configured metadata change option when you run a catalog source.

3. In the **Filters** area, define one or more filter conditions to apply for metadata extraction:
 - a. Select **Yes** to view filter options.
 - b. From the Include/Exclude list, choose to include or exclude metadata based on the filter parameters.
 - c. From the Object type list, select an object type, depending on the object that you want to extract metadata from.
 - d. Enter the path to the object as the filter value.

The following image shows the filter condition options:

Filters

Specify metadata filters: ☐ No ☒ Yes

[Show supported wildcards and examples](#)

Include or exclude metadata ▼ Select the object type ▼ Enter a value to specify the object location + 🗑️

File
Folder
Path

- e. To define an additional filter with an OR condition, click the **Add** icon.
4. In the **Configuration Parameters** area, enter the catalog source configuration options and enable or disable partition detection in files.

The following table describes the properties that you can enter:

Property	Description
Extract Group Elements from Hierarchical Files	<p>Select one of the following options to extract group or leaf elements from hierarchical files:</p> <ul style="list-style-type: none"> - Yes. Extracts group elements from hierarchical files with the complete hierarchy of hierarchical fields. You can view the hierarchy of hierarchical files in the Hierarchy tab of assets in Data Governance and Catalog. - No. Extracts only leaf elements from hierarchical files without the complete hierarchy of hierarchical fields. <p>You can extract group elements from hierarchical files for the following file types:</p> <ul style="list-style-type: none"> - AVRO. Extracts and groups hierarchical files and hierarchical fields. - Parquet. Extracts and groups hierarchical files and hierarchical fields. - JSON. Extracts and groups hierarchical files and hierarchical fields. - XML. Extracts and groups elements and attributes. For XML file types, a maximum depth of 1000 elements is permitted within a single element in the hierarchy. - XSD. Extracts and groups elements and attributes. <p>Attention: If you modify the Extract Group Elements from Hierarchical Files field and run the catalog source again, the asset page doesn't display the hierarchical elements in the correct hierarchy groups. If you modify the property value, purge the catalog source before you run it again.</p>
Enable Extension-Based File Type Detection	<p>Select one of the following options to detect file types by file extensions or by parsing the file contents:</p> <ul style="list-style-type: none"> - Yes. Detects file types by file extensions. - No. Parses the file contents to detect file types. <p>Note: You can detect file types by file extensions for the following file types:</p> <ul style="list-style-type: none"> - CSV - TSV - TXT - XML
Use First Row as Header of Delimited Files	<p>Select one of the following options to use the first row as the header or detect headers automatically for delimited files:</p> <ul style="list-style-type: none"> - Yes. <ul style="list-style-type: none"> Detects column headers based on the following rules: <ul style="list-style-type: none"> - Duplicate headers get suffixed with '#' followed by a number, for example, ABC#1, ABC#2. The detection is not case-sensitive. - Empty column header values appear as UnknownColumn<position>, for example UnknownColumn2. - The header row in the file is detected even if it has a different number of columns than the data rows. - No. Detects headers automatically for delimited files.
Headers of Delimited Files	<p>Specify values to determine headers of delimited files. Separate multiple values by commas. If any value from the list is found in the first row of the delimited file, then the first row is used as the header.</p> <p>Note: This parameter appears only if you choose No for the Use First Row as Header of Delimited Files parameter.</p>
Treat Files Without Extension As	<p>Select one of the following options to identify files without an extension:</p> <ul style="list-style-type: none"> - Parquet - AVRO - JSON

Property	Description
Enter File Delimiter	<p>Specify the file delimiter if the file from which you extract metadata uses a delimiter other than the following list of delimiters:</p> <ul style="list-style-type: none"> - Comma (,) - Horizontal tab (\t) - Semicolon (;) - Colon (:) - Pipe symbol () <p>Enclose the delimiter in single quotes, such as '\$'. Use a comma to separate multiple delimiters.</p> <p>Note: Adding a custom delimiter overrides the default list of delimiters. If you specify a delimiter, characters from the default list are not considered as delimiters.</p>
Files to be excluded during partition discovery	<p>Specify the regular expression of the files that you want to exclude during partition discovery.</p> <p>Enclose each regular expression in double quotes such as, ".json","Customer.csv","Parquet.*". Use a comma to separate multiple regular expressions.</p>
Partition Detection	<p>Enable or disable detection of partitions in files.</p> <p>Enable partition detection to identify horizontally partitioned files and publish them in a directory and files organized in hierarchical Hive-style directory structures as a single partitioned file.</p>

5. Click **Next**.

The **Associations** page appears.

Filter guidelines and examples

You can add metadata extraction filters when you configure the catalog source. To create a filter, you can use choose from file names, folder names, or paths.

Consider the following rules and guidelines when you enter filter values:

- Filters are case-sensitive.
- Use an asterisk to represent multiple characters in a folder name, file name, and a single folder level in a folder hierarchy. For example, A* matches A, Ab, ABC.
- For file filters, specify only the file name.
- If a file name contains an asterisk, the filter considers it as a wildcard and not a special character. To ignore an asterisk as a wildcard, enclose it in double quotes (") in the filter.
- Use a forward slash as a separator in path hierarchies. You can add a path in folder and path filters.
- Use an asterisk as a path placeholder in folder and path filters. For example, folder1/*/folder3. Here, the filter includes all folders under folder1.
- Use two asterisks to indicate zero or more levels of folders in folder and path filters. The pattern with two asterisks is recursive. The processing time is longer as the data volume increases.

Important: It is recommended that you either use only a path filter or use a combination of a folder and a file filter.

Examples

You can include or exclude metadata from folders, files, or paths.

Folder filters

Folder filters apply to folders included in the source system.

For example:

- To include or exclude metadata from 'Folder2' located inside 'Folder1', select **Folder** as the object type and enter `Folder1/Folder2` in the value field.
- To include or exclude metadata from 'Folder2' located in any folder under 'Folder1', select **Folder** as the object type and enter `Folder1/*/Folder2` in the value field.
- To include or exclude metadata from 'Folder2' located two levels under 'Folder1', select **Folder** as the object type and enter `Folder1/*/*/Folder2` in the value field.
- To include or exclude metadata from 'Folder2' located at any level under 'Folder1', select **Folder** as the object type and enter `Folder1/**/Folder2` in the value field. This is a recursive search, and therefore the processing time can be longer.

File filters

File filters apply to the files included in folders that you filter. The file filter is recursive. If you don't provide any folder filters, the file filters apply to the entire folder hierarchy.

For example:

- To include or exclude metadata from all files with the name 'File1.csv' located in the source directory, select **File** as the object type and enter `File1.csv` in the value field. Metadata Command Center recursively searches for files that match the filter criteria in all folders in the source directory.
- To include or exclude metadata from all files with names that start with 'File' and end with 'ame.csv', select **File** as the object type and enter `File*ame.csv` in the value field.
- To include or exclude metadata from all files with names that end with 'File.csv', select **File** as the object type and enter `*File.csv` in the value field.
- To include or exclude metadata from all files with the name 'File' and files that start with the name 'File' followed by one or more characters, select **File** as the object type and enter `File*` in the value field.
- To include or exclude metadata from all files with names that contain the word 'File', select **File** as the object type and enter `*File*` in the value field.
- To include or exclude metadata from all files with the name 'Fi*le.csv', select **File** as the object type and enter `Fi"*le.csv` in the value field.

Path filters

Path filters apply to the files and folders in the path that you filter. The path filter is non-recursive. If you provide only the file or folder names, the path filters apply to the first level files or directories.

For example:

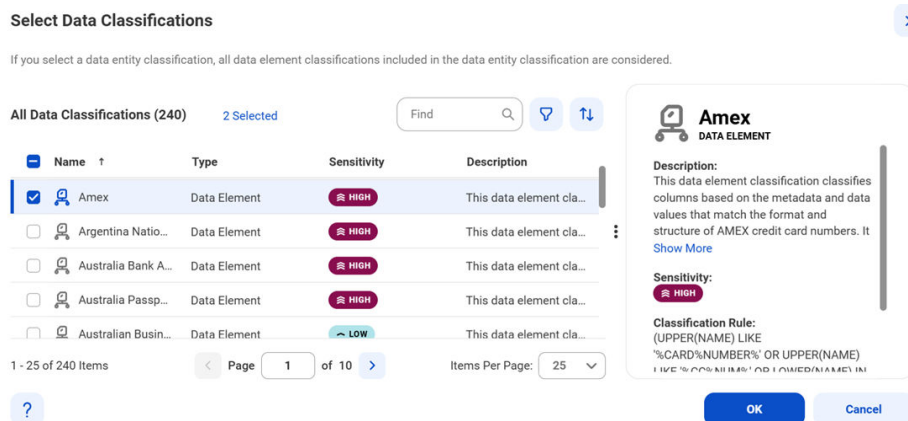
- To include or exclude metadata from files and folders with names that start with 'Item1' in the first level directory, select **Path** as the object type and enter `Item1*` in the value field.
- To include or exclude metadata from the 'File1' file in the 'Folder1' folder, select **Path** as the object type and enter `Folder1/File1` in the value field.
- To include or exclude metadata from files or folders with names that contain the word 'Subfolder' in the 'Folder1' folder, select **Path** as the object type and enter `Folder1/*Subfolder*` in the value field.
- To include or exclude metadata from files or folders with the name 'File1' in any subfolder of the 'Folder1' folder, select **Path** as the object type and enter `Folder1/*/File1` in the value field.

- To include or exclude metadata from all files and subfolders in the 'Folder1' folder, select **Path** as the object type and enter `Folder1/*` in the value field.
- To include or exclude metadata from files or folders with the name 'File1' located at any level in the 'Folder1' folder, select **Path** as the object type and enter `Folder1/**/File1` in the value field. This is a recursive search, and therefore the processing time can be longer.

Configure data classification

Enable the data classification capability to identify and organize data into relevant categories based on the functional meaning of the data.

1. Click the **Data Classification** tab.
2. Select **Enable Data Classification**.
3. Choose one or both of the following options:
 - **Generated Data Classifications.** CLAIRE automatically generates data classifications for the data elements.
 - **Data Classification Rules.** Choose from predefined or custom data classifications.
 1. Click **Add Data Classification**. The following image shows the **Select Data Classifications** dialog box:



2. Select the data classifications that you want to use.
3. Click **OK**.

Configure glossary association

Enable the glossary association capability to associate glossary terms with technical assets, or to get recommendations for glossary terms that you can manually associate with technical assets in Data Governance and Catalog.

Metadata Command Center considers all published business terms in the glossary while making recommendations to associate your technical assets.

1. Click the **Glossary Association** tab.
2. Select **Enable Glossary Association**.
3. Select **Enable auto-acceptance** to automatically accept glossary association recommendations.
4. Specify the **Confidence Score Threshold for Auto-Acceptance** to set a threshold limit based on which the glossary association capability automatically accepts the recommended glossary terms.

Note: Specify a percentage from 80 to 100. If the score is higher than the specified limit, the glossary association capability automatically assigns a matching glossary term to the data element.

5. Select **Enable Below-threshold Recommendations** to receive glossary association recommendations below the auto-acceptance threshold. If you enable auto-acceptance, you can enable below-threshold recommendations to receive glossary recommendations below the auto-acceptance threshold.
6. Specify the **Confidence Score Threshold for Recommendations** to set a threshold based on which the glossary association capability makes recommendations
If you enable auto-acceptance, specify a percentage from 80 to the selected auto-acceptance threshold. You can accept or reject the recommended glossary terms that fall within this range in Data Governance and Catalog.
If you disable auto-acceptance, specify a percentage from 80 to 100 inclusive.
7. Choose to automatically assign business names and descriptions to technical assets. You can then choose to retain existing assignments and only assign business names and descriptions to assets that don't have assignments, or allow overwrite of existing assignments.
By default, existing assignments are retained.
8. Optional. Choose to ignore specific parts of data elements when making recommendations. Select **Yes** and enter prefix and suffix keyword values as needed.
Click **Select** to enter a keyword. You can enter multiple unique prefix and suffix keywords. Keyword values are case insensitive.
9. Optional. Choose specific top-level business glossary assets to associate with technical assets. Selecting a top-level asset selects its child assets as well. Select **Top-level Glossary Assets** and specify the assets on the **Select Assets** page.
10. Optional. Choose to use abbreviations and synonym definitions from lookup tables for accurate glossary association. Select **Yes** to enable, and then click **Select** to upload a lookup table.
11. Click **Next**.
The **Associations** page appears.

Step 3. Associate stakeholders and asset groups

Associate users or user groups within a stakeholder role as stakeholders for technical assets in Data Governance and Catalog. Also, you can choose to assign technical assets extracted from the catalog source to asset groups. You can then use access policies to control permissions on assets that are assigned to asset groups.

Verify that the administrator assigned users and user groups to the stakeholder role that you want to associate with technical assets.

1. To associate users or user groups as stakeholders with technical assets extracted from the catalog source, perform the following steps:
 - a. On the **Associations** page, click **Stakeholders**.
 - b. Select **Assign Stakeholders**.
 - c. Select a stakeholder role.

- d. Click **Select** to add users and user groups from the stakeholder role as stakeholders for the technical assets.

The **Add Users & User Groups** dialog box displays a list of users and user groups assigned to the selected stakeholder role.

- e. Select one or more users or user groups to assign as stakeholders for the technical assets, and click **OK**.

Only the selected users and user groups belonging to the specified stakeholder role are granted the permissions to technical assets.

- f. To assign users or user groups from another stakeholder role, click **Add** and then repeat the steps.

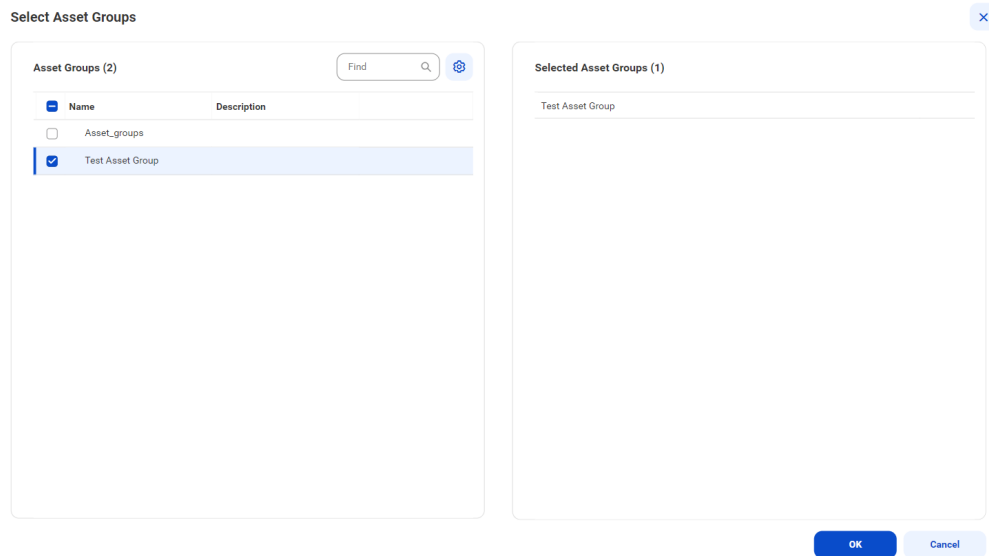
2. To assign asset groups to technical assets extracted from the catalog source, perform the following steps:

- a. On the **Associations** page, click **Asset Groups**.
- b. Select **Assign Asset Groups**.
- c. Click **Select**.

The **Select Asset Groups** dialog box displays the list of asset groups.

If you enabled an access policy that includes an asset group, you can only view assets that belong to that asset group.

3. Select the asset groups to which you want to assign technical assets extracted from the catalog source, and click **OK**.



4. Choose to save and run the job or to schedule a recurring job.
 - To save and run the job, click **Save** and then **Run**.
 - To schedule a recurring job, click **Next** to open the **Schedule** page.

Step 4. Run or schedule the job

Choose to run a catalog source job manually, or configure it to run on schedule.

Note: You can't run multiple jobs simultaneously.

You can choose to perform a full or an incremental metadata extraction. A full metadata extraction extracts all objects from the source to the catalog. An incremental metadata extraction extracts only the changed and new objects since the last successful catalog source job run. Incremental metadata extraction doesn't remove deleted objects from the catalog and doesn't extract metadata of code-based objects if applicable.

When you run an incremental metadata extraction job with a filter to include metadata from objects, the job extracts only the objects that have the latest timestamp since the last successful job.

Note: The incremental extraction option appears if it is available for the catalog source.

Run the job manually

Click **Save** to save the catalog source and click **Run**. On the **Run Catalog Source Job** window, click **Run** to run the job.

You can override the capabilities that you selected while configuring your catalog source on the **Configuration** page. The first time you run the catalog source job, the metadata extraction capability is mandatory. From the second run onwards, you can choose to override the configured metadata change option. You can retain, delete, or deprecate objects that are deleted from the source in the catalog. For subsequent runs of the catalog source job, the metadata extraction capability is optional.

Note: You can choose incremental metadata extraction for subsequent runs only after one full metadata extraction job completes successfully. Incremental metadata extraction jobs run with the **Retain** metadata change option even if you set the option to **Delete** or **Deprecate** in the catalog source.

Note: To run a catalog source job, you need permissions on the connection to the source system. To run a catalog source job for catalog sources that reference other source systems, you need permissions on the connections for all the reference source systems.

Run the job on a schedule

You can choose to run metadata extraction and other capabilities on a recurring schedule. You can't choose incremental metadata extraction and full metadata extraction in the same schedule. To create a schedule for incremental metadata extraction, you must have completed at least one full metadata extraction job successfully. If not, first create a schedule for a full metadata extraction.

If an incremental metadata extraction is scheduled to run when the last run details aren't available, the job first performs a full metadata extraction, followed by incremental metadata extraction on subsequent runs.

For example, this can happen in the following scenarios:

- You create schedules for both incremental metadata extraction and full metadata extraction, but schedule the incremental extraction to run before the first full metadata extraction job.
- You create schedules for both incremental metadata extraction and full metadata extraction, but delete the full metadata extraction schedule before its first run.

1. On the **Schedule** tab, select **Run on Schedule**.
The **Schedule** configuration page opens.
2. Click the checkbox corresponding to each capability that you want to include in the schedule.
3. Enter the start date, time zone, and the interval at which you want to run the job.
4. You can manage additional schedules using the following options:
 - To create a new schedule, click the **Add** button.
 - To delete a schedule, click the **Delete** button.
 - To enable or disable a schedule, click the **Enable Schedule** toggle button.

Note: You can create a maximum of one schedule per capability that you enable. If you purged a catalog source or did not run the metadata extraction job, the catalog source job runs metadata extraction before running other scheduled capabilities.

Note: To create a schedule, you need permissions on the connection to the source system. If you lose permissions on the connection after you create a schedule, the scheduled jobs continue to run.

5. Click **Save** to save the schedule.

Monitor job status

After the job runs, you can monitor the status of the job on the **Overview** page of the job.

For more information about job monitoring, see *Administration*.

Step 5. Connect to referenced source systems

When you run the catalog source job, if the catalog source references another source system, a reference catalog source and connection get created that point to the reference source system. To view the complete lineage for your catalog source, you can perform connection assignment from the reference catalog source connection to the objects in the reference source system. A reference source system might be a database,

such as Oracle. You must first create and run an endpoint catalog source that connects to the reference source system.

Before you assign a connection, ensure that you have created and run an endpoint catalog source for each reference source system.

Note: If the source schema contains case-sensitive tables or if the reference objects contain multiple objects with the same name in different cases, perform case-sensitive connection assignment to get correct lineage.

If you enabled the lineage discovery capability for your catalog source, you can either curate the CLAIRE recommended endpoint objects on the **Related Catalog Sources** tab or assign connections manually.

For more information about related catalog sources and lineage discovery, see *Lineage discovery* in the *Administration* help.

Preview Notice: Effective in the April 2023 release, creation of connection assignments for a Microsoft SharePoint Online catalog source is available for preview. Preview functionality is supported for evaluation purposes but is unwarranted and is not supported in production environments or any environment that you plan to push to production. Informatica intends to include the preview functionality in an upcoming release for production use, but might choose not to in accordance with changing market or technical circumstances. For more information, contact Informatica Global Customer Support.

1. On the **Configure** page, select the **Lineage** tab, and then select the **Lineage Discovery** tab. On the **Catalog Sources** panel, select the required catalog source and click the **Assign Connections** tab.
The **Assign Connections** tab displays a list of assigned and unassigned connections along with details for each connection. Use filters to view the connections based on the connection names. Click the **Add Filter** menu to add filters.
2. Select the connection to the reference source system and click **Assign**.
The connection name appears prefixed to the reference catalog source name on the **Hierarchy** tab of your catalog source in Data Governance and Catalog.
The **Assign Connection** dialog box appears with a list of recommended objects from the endpoint catalog sources. Click **All** to view all endpoint catalog source objects.
3. In the **Assign Connection** dialog box, select one or more catalog sources to assign to the selected connection and click **Assign**.

When you click **Assign**, Metadata Command Center creates links between matching objects in the connected catalog sources, and it calculates the percentage of matched and unmatched objects. The higher the percentage of matched objects, the more accurate the lineage that you view in Data Governance and Catalog.

CHAPTER 4

View results in Data Governance and Catalog

After Metadata Command Center runs a job, you can view the results in Data Governance and Catalog where the catalog source and its elements are called technical assets. You can view a catalog source as a hierarchy. Expand each technical asset to see its components.

When referenced source systems are connected to a catalog source, you can expand the hierarchy to see details about the technical asset's component elements.

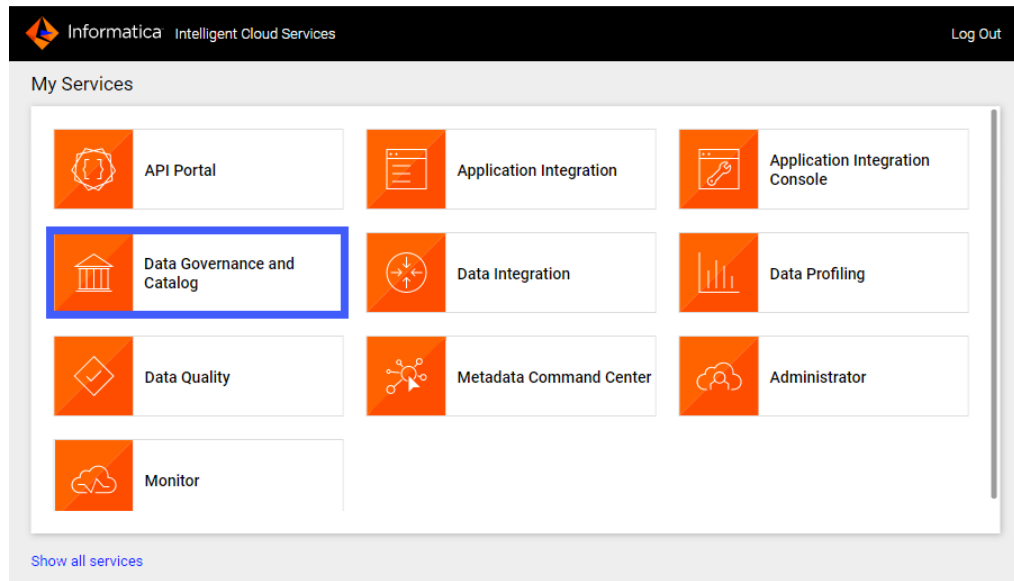
You can view the data lineage of an asset contained within a catalog source to see individual elements such as data sources, calculations, and filters. When you view data lineage, you can see the individual upstream elements that contribute data or expressions to each component of a data flow or catalog source.

View metadata extraction results

After a job runs in Metadata Command Center, view the results in Data Governance and Catalog. You can view details about source system contents as hierarchical displays.

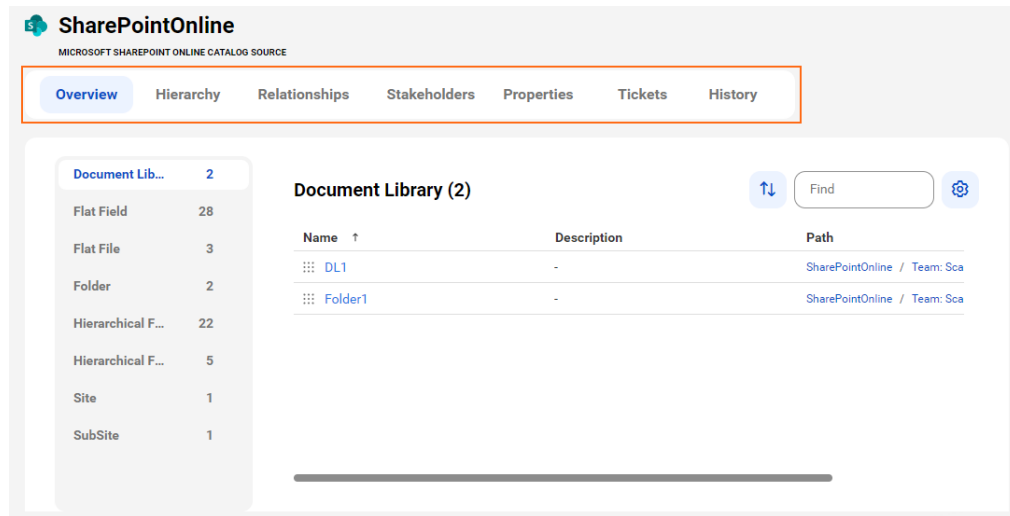
1. Log in to Informatica Intelligent Cloud Services and select Data Governance and Catalog from the **My Services** page.

The following image shows the **My Services** page:



2. On the Data Governance and Catalog home page, click the number in the **Technical Assets** panel.
The **Technical Assets** page opens.
3. Select **Catalog Source** in the **Filter** list.
The list of catalog sources opens.
4. Search for the catalog source from which you extracted metadata, and click the name.
The **Overview** tab of the asset opens.

The following image shows a sample asset page:



5. View the asset from different perspectives by clicking on the tabs.

For more information about working with assets, see "Working with Assets" in *Cloud Data Governance and Catalog* online help.

View data lineage

Data lineage is a visual representation of the flow of data across the systems in your organization. Lineage depicts how the data flows from the system of its origin to the system of its destination.

Data lineage views are available for technical assets in the catalog source. You can view lineage at the catalog source, data set, or data element level.

The lineage at the catalog source level shows how data flows from one catalog source to another. The lineage at the data set and the data element levels show how other technical assets such as files or tables contribute to the selected asset.

If linking catalog sources is available for your catalog source, you can use Metadata Command Center to generate data lineage based on rules or by generating automated lineage with CLAIRE. You can choose source and target catalog sources and objects to link and generate lineage.

To determine whether linking catalog sources is available for your catalog source, navigate to the **Configuration** tab of the **Link Catalog Sources** page. The catalog source must appear in the list of source and target catalog sources.

For information about linking catalog sources, see *Link catalog sources* in the Administration help.

View lineage at the catalog source level

The catalog source level shows how data flows from one catalog source to another with the lineage aggregating data from the data set and data element levels.

To view data lineage at the catalog source level, open a technical asset, click the **Lineage** tab, and then verify that the level is set to **Catalog Source Level**.

Preview Notice: Effective in the April 2023 release, data lineage views of technical assets in the Microsoft SharePoint Online catalog source lineage are available for preview. Preview functionality is supported for evaluation purposes but is unwarranted and is not supported in production environments or any environment that you plan to push to production. Informatica intends to include the preview functionality in an upcoming release for production use, but might choose not to in accordance with changing market or technical circumstances. For more information, contact Informatica Global Customer Support.

View lineage at the data set level

The data set level displays individual sets of data in the data flow.

To view lineage at the data set level, open a technical asset, click the **Lineage** tab, and then verify that the level is set to **Data Set Level**.

View lineage at the data element level

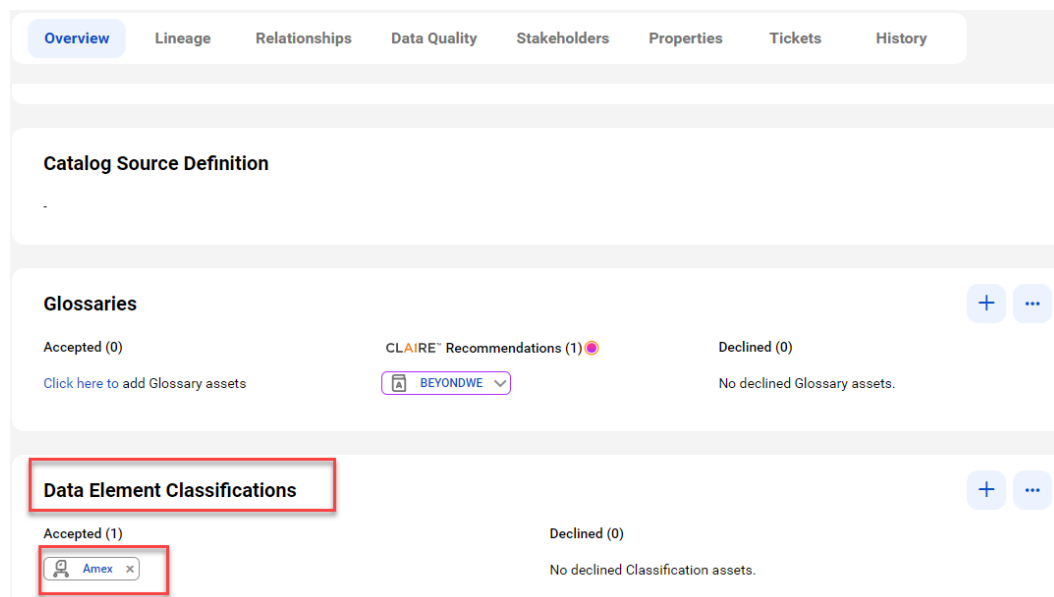
The data element level displays details of the data set level. At the data element level, you can see the input sources for expressions or commands and calculations or transformations on the data.

To view data lineage at the data element level, open a technical asset, click the **Lineage** tab, and then verify that the level is set to **Data Element Level**.

View classified data

When you add data classification rules to a catalog source in Metadata Command Center, the system identifies the columns and tables that match the rules and displays one or more matched data classifications on the column or table asset pages in Data Governance and Catalog.

The following image shows a column asset page with the inferred data element classifications that match the column data and metadata:



For more information about data classification assets, see *Asset Details* in the Data Governance and Catalog help.

View glossary associations

When you enable the glossary association capability for a catalog source in Metadata Command Center, you can view the accepted glossary assets in Data Governance and Catalog.

The **Overview** tab for a technical asset in the catalog source displays glossary assets in the Accepted and CLAIRE Recommendations sections.

The **Glossaries** panel shows the automatically accepted and CLAIRE® recommended terms.

The following image shows a sample asset page:

