



Informatica® Metadata Command Center
November 2025

Microsoft Azure Data Lake Storage Gen2 Sources

© Copyright Informatica LLC 2024, 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: 2025-11-20

Table of Contents

Preface.	5
Chapter 1: Introduction to Microsoft Azure Data Lake Storage Gen2 catalog sources.	6
Extraction and view process.	7
About the Microsoft Azure Data Lake Storage Gen2 catalog source.	8
Extracted metadata.	8
Data profiling for Microsoft Azure Data Lake Storage Gen2 objects.	9
Compatible connectors.	10
Chapter 2: Before you begin.	11
Verify permissions.	11
Permissions to extract metadata.	11
Permissions to run data profiles.	11
Permissions to perform data classification.	12
Permissions to perform relationship discovery.	12
Permissions to perform glossary association.	12
Create a connection	12
Service Principal Authentication.	13
Shared Key Authentication.	14
Managed Identity Authentication.	14
Import a relationship inference model.	15
Chapter 3: Create catalog sources in Metadata Command Center.	16
Step 1. Register a catalog source.	16
Step 2. Configure capabilities.	18
Configure metadata extraction.	18
Filter guidelines and examples.	22
Configure data profiling and quality.	23
Configure data classification.	27
Configure relationship discovery.	28
Configure glossary associations.	29
Step 3. Associate stakeholders and asset groups.	29
Step 4. Run or schedule the job.	31
Chapter 4: View results in Data Governance and Catalog.	33
View metadata extraction results.	33
View data lineage.	35
View lineage at the catalog source level.	35
View lineage at the data set level.	36

View lineage at the data element level.	36
View data profiling results	36
View data observability results	37
View classified data.	38
View glossary associations.	38

Preface

Read *Microsoft Azure Data Lake Storage Gen2 Sources* to learn how to register and configure Microsoft Azure Data Lake Storage Gen2 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 Azure Data Lake Storage Gen2 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 Azure Data Lake Storage Gen2 is a source system from which you can extract metadata through a Microsoft Azure Data Lake Storage Gen2 catalog source. 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:

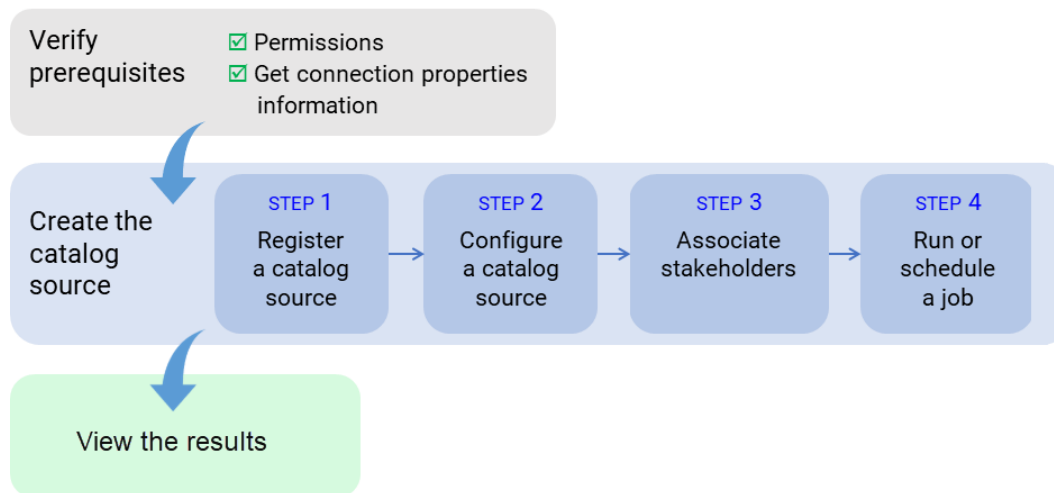
Capability	Description
Incremental metadata extraction	An incremental metadata extraction extracts only the changed and new objects since the last catalog source job run. Incremental metadata extraction doesn't remove deleted objects from the catalog and doesn't extract metadata of code-based objects if applicable.
Serverless Runtime Environment	A serverless runtime environment is an advanced serverless deployment solution that doesn't require downloading, installing, configuring, or maintaining a Secure Agent or Secure Agent group. You can use a serverless runtime environment in the same way that you use a Secure Agent when you configure a catalog source.
Data Profiling and Quality	<ul style="list-style-type: none">- Data Profiling. Assesses source metadata and analyzes the collected statistics to discover content and structure, such as value distribution, patterns, and data types.- Data Quality. Measures the reliability of the data and enables data usage.- Data Observability. Identifies anomalies in the characteristics of the data.
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.

Capability	Description
Relationship Discovery	The relationship discovery capability identifies pairs of similar columns and relationships between tables within a catalog source.
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 Azure Data Lake Storage Gen2:

1. Register a catalog source. Create a catalog source object, select the source system, 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.

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

About the Microsoft Azure Data Lake Storage Gen2 catalog source

You can use the Microsoft Azure Data Lake Storage Gen2 catalog source to extract metadata from a Microsoft Azure Data Lake Storage Gen2 source system.

Microsoft Azure Data Lake is a scalable data storage and analytics service that is hosted on Azure. Configure an Azure Data Lake Store (ADLS) catalog source to extract metadata from the data lake store.

Extracted metadata

You can use the Microsoft Azure Data Lake Storage Gen2 catalog source to extract metadata from a Microsoft Azure Data Lake Storage Gen2 source system.

Metadata Command Center extracts the following metadata from an Azure Data Lake Storage Gen2 source system:

- File
- Folder
- Container
- Iceberg Table
- Iceberg Field

You can extract metadata from the following file types:

- AVRO
- Compressed files
 - TAR
 - ZIP
- Delimited files
 - CSV
 - TSV
- JSON
- Microsoft Excel files
 - Excel 97-2003 Workbook with XLS extension
 - Excel Workbook with XLSX extension
 - Excel Macro-Enabled Workbook with XLSM extension
- Parquet
- TXT
- XML
- XSD

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

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

You can extract metadata from XML and XSD file formats as XML file objects. Metadata Command Center extracts only elements and attributes from XML and XSD files. If the size of the XML file exceeds 100 KB, Metadata Command Center extracts metadata from the initial 100 KB of the file. However, for XSD file types, Metadata Command Center extracts complete metadata.

Data profiling for Microsoft Azure Data Lake Storage Gen2 objects

Configure data profiling to run profiles on the metadata extracted from a Microsoft Azure Data Lake Storage Gen2 source system. You can run data profiles on the following Microsoft Azure Data Lake Storage Gen2 objects:

- Avro
- CSV
- JSON
- Parquet

You can view the profiling statistics in Data Governance and Catalog. The data profiling task runs profiles on the following data types for AVRO, CSV, JSON, and Parquet file formats:

File format	Supported data type
AVRO	- INT - STRING
CSV	- STRING

File format	Supported data type
JSON	<ul style="list-style-type: none"> - ARRAY - BOOLEAN - NUMBER (double) - NUMBER (float) - NUMBER (int) - NUMBER (long) - OBJECT - STRING
Parquet	<ul style="list-style-type: none"> - BOOLEAN - INT32 - INT64 - FLOAT - DOUBLE - DATE - DECIMAL - STRING

Compatible connectors

Before you configure a Microsoft Azure Data Lake Storage Gen2 catalog source, you must connect to the Microsoft Azure Data Lake Storage Gen2 source system.

Use the Microsoft Azure Data Lake Storage Gen2 connector to connect to the Microsoft Azure Data Lake Storage Gen2 source system.

For information about configuring a connection, see *Connections* in the Administrator service.

CHAPTER 2

Before you begin

Before you create a catalog source, ensure that you have the information required to connect to the source system.

Perform the following tasks:

- Assign the required permissions.
- Configure a connection to the Microsoft Azure Data Lake Storage Gen2 source system in Administrator.
- Optionally, if you want to identify pairs of similar columns and relationships between tables within a catalog source, import a relationship inference model.

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 to extract metadata

Ensure that you have the required permissions to enable metadata extraction and to access the Windows or the Linux file system.

- Create a user account for the Informatica user to access the Microsoft Azure Data Lake Storage Gen2 source system.
- Grant read permission to the user account.

Permissions to run data profiles

Ensure that you have the required permissions to run profiles.

Grant the following permissions:

- Read permission. Required to read the data and metadata.
- Write permission. Required to write the profiling results to the staging location.
- Delete permission. Required to delete the profiling results from the staging location.
- Execute permission. Required to perform the test connection of query executions.

Permissions to perform data classification

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

Permissions to perform relationship discovery

You can perform relationship discovery with the permissions required to perform metadata extraction.

Permissions to perform glossary association

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

Create a connection

Before you configure the Microsoft Azure Data Lake Storage Gen2 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. Select a Secure Agent, Hosted Agent, or serverless runtime environment. Do not use a Hosted Agent if you use the connection in mappings in advanced mode. For more information about how to configure and use the runtime environments, see <i>Runtime Environments</i> in the Administrator help.

Property	Description
Account Name	Microsoft Azure Data Lake Storage Gen2 account name or the service name.
File System Name	The name of the file system in the Microsoft Azure Data Lake Storage Gen2 account.
Directory Path	<p>The path of a directory without the file system name.</p> <p>You can select from the following directory structures:</p> <ul style="list-style-type: none"> - / for root directory - /dir1 - dir1/dir2 <p>Default is /.</p>

- Select the authentication type to connect to Microsoft Azure Data Lake Storage Gen2 and enter the required properties. You can use the following authentication types:
 - Service Principal Authentication
 - Shared Key Authentication
 - Managed Identity Authentication
- Click **Test Connection**.

Service Principal Authentication

This authentication method uses the client ID, client secret, and tenant ID to connect to Microsoft Azure Data Lake Storage Gen2.

The following table describes the connection properties for the Service Principal Authentication type:

Property	Description
Client ID	<p>The client ID of your application.</p> <p>Specify the client ID for your application registered in the Azure Active Directory.</p>
Client Secret	<p>The client secret key generated for the client ID.</p> <p>Specify the client secret key to complete the OAuth authentication in the Azure Active Directory.</p>
Tenant ID	The directory ID of the Azure Active Directory.
Endpoint Suffix	<p>The type of Microsoft Azure endpoints.</p> <p>Select one of the following endpoints:</p> <ul style="list-style-type: none"> - core.windows.net. Connects to Azure endpoints. - core.usgovcloudapi.net. Connects to US government Microsoft Azure Data Lake storage Gen2 endpoints. - core.chinacloudapi.cn. Connects to Microsoft Azure Data Lake storage Gen2 endpoints in the China region. <p>Default is core.windows.net.</p> <p>Note: You cannot configure the Azure Government endpoints for mappings in advanced mode.</p>

Shared Key Authentication

This authentication method uses the account key to connect to Microsoft Azure Data Lake Storage Gen2.

The following table describes the connection properties for the Shared Key Authentication type:

Property	Description
Account Key	The account key for the Microsoft Azure Data Lake Storage Gen2 account.
Endpoint Suffix	<p>The type of Microsoft Azure endpoints.</p> <p>Select one of the following endpoints:</p> <ul style="list-style-type: none">- core.windows.net. Connects to Azure endpoints.- core.usgovcloudapi.net. Connects to US government Microsoft Azure Data Lake storage Gen2 endpoints.- core.chinacloudapi.cn. Connects to Microsoft Azure Data Lake storage Gen2 endpoints in the China region. <p>Default is core.windows.net.</p> <p>Note: You cannot configure the Azure Government endpoints for mappings in advanced mode.</p>

Managed Identity Authentication

This authentication method uses the identities that are assigned to applications in Azure to access Azure resources in Microsoft Azure Data Lake Storage Gen2.

When you create a Microsoft Azure Data Lake Storage Gen2, select the Azure virtual machine on which you have installed the Secure Agent. If you enable system assigned identity, assign the required role or permissions to the Azure virtual machine to run the mappings and tasks. If you enable user assigned identity, assign the required role or permissions to the user assigned identity.

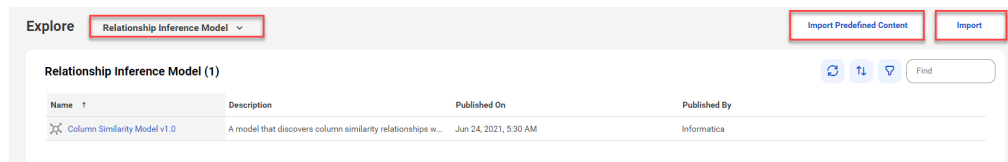
The following table describes the connection properties for the Managed Identity Authentication type:

Property	Description
Client ID	<p>The client ID of your application.</p> <p>To use managed identity authentication, specify the client ID for the user-assigned managed identity.</p> <p>Leave the field blank in the following scenarios:</p> <ul style="list-style-type: none">- If the permission is provided by system-assigned managed identity.- If there is no system-assigned identity but only a single user-assigned managed identity.
Endpoint Suffix	<p>The type of Microsoft Azure endpoints.</p> <p>Select one of the following endpoints:</p> <ul style="list-style-type: none">- core.windows.net. Connects to Azure endpoints.- core.usgovcloudapi.net. Connects to US government Microsoft Azure Data Lake storage Gen2 endpoints.- core.chinacloudapi.cn. Connects to Microsoft Azure Data Lake storage Gen2 endpoints in the China region. <p>Default is core.windows.net.</p> <p>Note: You cannot configure the Azure Government endpoints for mappings in advanced mode.</p>

Import a relationship inference model

Import a relationship inference model if you want to configure the relationship discovery capability. You can either import a predefined relationship inference model, or import a model file from your local machine.

1. In Metadata Command Center, click **Explore** on the navigation panel.
2. Expand the menu and select **Relationship Inference Model**. The following image shows the **Explore** page with the **Relationship Inference Model** menu:



3. Select one of the following options:
 - **Import Predefined Content.** Imports a predefined relationship inference model called Column Similarity Model v1.0.
 - **Import.** Imports the predefined relationship inference model from your local machine. Select this if you previously imported predefined content into your local machine and the inference model is stored on the machine.
To import a file, click **Choose File** in the **Import Relationship Inference Model** window and navigate to the model file on your local machine. You can also drag and drop the file.

The imported models appear in the list of relationship inference models on the **Relationship Discovery** tab.

CHAPTER 3

Create catalog sources in Metadata Command Center

Use Metadata Command Center to configure a catalog source for Microsoft Azure Data Lake Storage Gen2 and run the catalog source job.

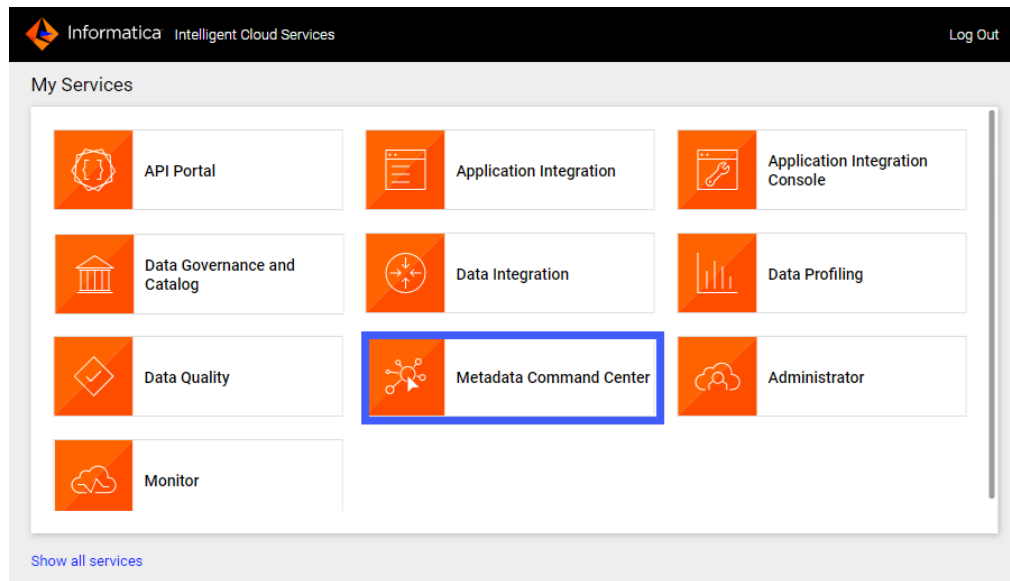
When you configure a catalog source, you define the source system that you want to extract metadata from. Configure filters to include or exclude source system metadata before you run the job. Optionally, configure other capabilities, such as lineage discovery, data profiling and quality, data classification,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

Step 1. Register a catalog source

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

1. Log in to Informatica Intelligent Cloud Services.
The **My Services** page appears.
2. Click **Metadata Command Center**.

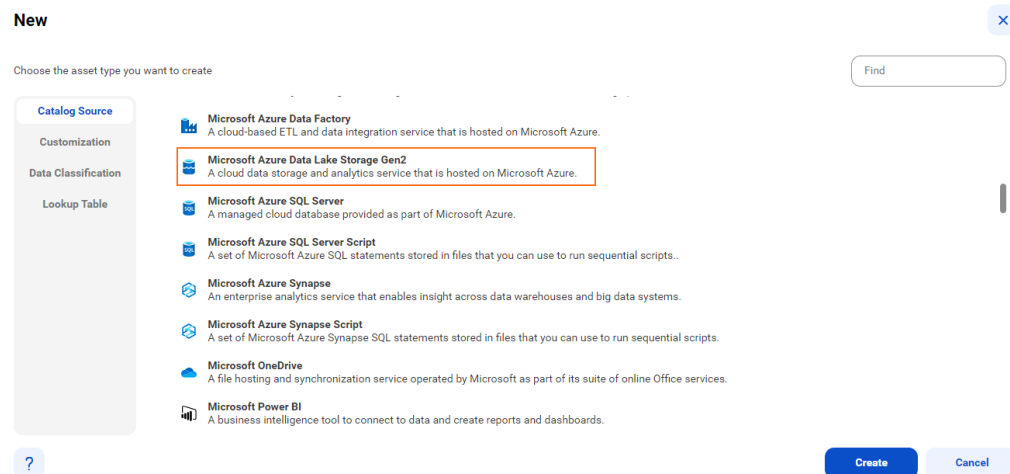
The following image shows the Metadata Command Center box on the **My Services** page:



The Metadata Command Center home page appears.

3. Click **New**.
4. Select **Catalog Source** from the list of asset types.
5. Select Microsoft Azure Data Lake Storage Gen2 from the list of catalog source types.

The following image shows the Microsoft Azure Data Lake Storage Gen2 catalog source:



6. Click **Create**.
- The **New Catalog Source** page opens.
7. 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.
 8. 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.

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

The **Configuration** page appears.

Step 2. Configure capabilities

When you configure the Microsoft Azure Data Lake Storage Gen2 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 Azure Data Lake Storage Gen2 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.

To define filters, you can either select an object type and enter the path to the object as the filter value, or select an object from a list of objects available in the source system.

- 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. Perform one of the following steps:
 - From the Object type list, select an object type, depending on the object that you want to extract metadata from. Enter the path to the object as the filter value.
 - In the filter value field, click the Search button and select an object from a list of objects available in the source system.
The Object type field updates based on the selected object.

If you select an object type and then click the Search button, the list of objects includes all object types, but you can only select objects that match the selected object type.

You can edit the filter value after you select an object from the list.

Note: You can only search for object types that work with the search functionality. If you don't see the Search button for the selected object, enter the object path as the filter value.

Note: If the object metadata is available in Data Governance and Catalog, a check mark appears next to the object.

Note: To select an object, you need to have permissions on the connection to the source system.

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

- d. To define an additional filter with an OR condition, click the **Add** icon.
4. Optionally, in the **Configuration Parameters** area, enter properties to override default context values and job parameters.

The following table describes the properties that you can enter for Catalog Source Configuration Options:

Parameter	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 use to identify files without an extension:</p> <ul style="list-style-type: none"> - Parquet - Avro - JSON

Parameter	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. Use a comma to separate multiple regular expressions.</p> <p>For example, enter ". *json","Customer.csv","Parquet.*".</p>

The following table describes the properties that you can enter for Partitioned File Configuration:

Parameter	Description
Custom Partition Detection Configuration File	<p>The configuration file in JSON format to detect custom partitions in the source system. The configuration file defines the pattern of the non-Hive style custom partitions.</p> <p>Note: The system automatically detects partitions if the date format of the partition key is in any of the following formats:</p> <ul style="list-style-type: none"> - dd-MM-yy - dd-MM-yyyy - dd-MMM-yyyy <p>For example, the system interprets dd-MM-yyyy as the partition format and extracts <code>Customer</code> as a hierarchical file instead of a folder in the following partition pattern:</p> <pre>Customer __01-01-2022 __ __ __dataFile1.parquet __ __ __dataFile2.parquet __ __ __... __02-02-2022 __...</pre> <p>The system doesn't automatically detect partitions with other patterns, such as MM-dd-yyyy. If the pattern is different, you need to define it in JSON format. For example,</p> <pre>{"CustomPartitionPatterns": ["MM-dd-yyyy"]}</pre> <p>To detect the epoch time format, define it in JSON format as:</p> <pre>{"CustomPartitionPatterns": ["@"]}</pre>
Partitioned Pruning Configuration Options	<p>The configuration file in JSON format for partition pruning. The configuration file contains additional information to identify partitions and determine the relevant schema in the source system.</p>

Parameter	Description
Enable Partition Detection	Specify if you want to enable partition detection.
Enable Partition Pruning	Specify if you want to enable partition pruning. Default is Yes. Note: Applicable if you enable partition detection.

5. Configure additional capabilities for the catalog source by clicking on the tabs.

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 profiling and quality

Enable the data profiling capability to evaluate the quality of metadata extracted from the Microsoft Azure Data Lake Storage Gen2 source system.

1. Click the **Data Profiling and Quality** tab.
2. Expand **Data Profiling** and select **Enable Data Profiling**.

Note: Ensure that you have permissions on all the staging connections that you use in your data profiling configuration. You can't run the job if you don't have permissions on the connections that you use. Select

connections that you have access to, or ask the administrator to grant the necessary permissions on the connections that you want to use.

3. In the **Connection and Runtime** area, choose the Secure Agent group where you want to run catalog source jobs.
4. Optionally, in the **Filters** area, specify additional filters in addition to metadata filters:
 - 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 **File** or **Folder**.
 - d. Enter the path to the object as the filter value.

Filters can contain the following wildcards:

- Question mark. Represents a single character.
- Asterisk. Represents multiple characters.

Examples:

- File: You extracted metadata from a folder that contains multiple files and now you want to run a profile on a specific file. Select **File** from the Object type option and select **Named** from the Filter Criteria option, and then enter the file name in the input field.
- Folder: You extracted metadata from a folder that contains multiple folders and now you want to run a profile on a specific folder. Select **Folder** from the Object type option and select **Named** from the Filter Criteria option, and then enter the folder name in the input field.

To include or exclude multiple objects, click the **Add** icon to add filters with the AND condition.

5. In the **Parameters** area, configure the parameters.

The following table describes the parameters that you can enter:

Parameter	Description
Modes of Run	Determine the type of data that you want the data profiling task to collect. Choose one of the following options: <ul style="list-style-type: none">- Keep signatures only. Collects only aggregate information such as data types, average, standard deviation, and patterns.- Keep signatures and values. Collects both signatures and data values.
Profiling Scope	Determine whether you want to run data profiling only on the changes made to the source system or on the entire source system. Choose one of the following options: <ul style="list-style-type: none">- Incremental. Includes only source metadata that is changed or updated since the last profile run.- Full. Includes the entire metadata that is extracted based on the filters applied for extraction.
Sampling Type	You can run the data profiling task on all rows for a Microsoft Azure Data Lake Storage Gen2 source system.

Parameter	Description
Elastic Runtime Environment	Select an elastic runtime environment for complex file types, including Avro and Parquet. Note: To run a profile on an Avro or Parquet files, use a secure agent configured with an advanced cluster. The same agent can also process delimited files. If the agent has no advanced cluster capabilities, the data profiling task runs profiles only on delimited files. For more information about advanced clusters, see Advanced Clusters help.
Staging Connection	The staging connection where data profiling results are stored temporarily during the profile run. Applicable for data profiling on complex file types. When you create a staging connection in Administrator, use the same client secret that you used to create a source connection object.
Maximum Precision of String Fields	The maximum precision value for profiles on string data type.
Text Qualifier	The character that defines string boundaries. If you select a quote character, profiling ignores delimiters within the quotes. Select a qualifier from the list. Note: Default is Double Quote.
Code Page for Delimited Files	Select a code page that the Secure Agent can use to read and write data. Use this option to ensure that profile results for assets with non-English characters don't include junk characters. Default value is UTF-8. Choose one of the following options: <ul style="list-style-type: none"> - MS Windows Latin 1. Select for ISO 8859-1 Western European characters. - UTF-8. Select for Unicode and non-Unicode characters. - Shift-JIS. Select for double-byte characters. - ISO 8859-15 Latin 9 (Western European). - ISO 8859-2 Eastern European. - ISO 8859-3 Southeast European. - ISO 8859-5 Cyrillic. - ISO 8859-9 Latin 5 (Turkish). - IBM EBCDIC International Latin-1.
Escape character for delimited files	You can specify an escape character if you need to override the default escape character. An escape character ignores a delimiter character in an unquoted string if the delimiter is part of the string value. If you specify an escape character, the data profiling task overrides the default escape character that the Metadata Extraction job detects and considers the specified escape character. It then reads the delimiter character as a part of the string value. If you don't specify an escape character, the data profiling task considers the default escape character that the Metadata Extraction job detects and reads the delimiter character as a part of the string value.
Read Multiple Line JSON Files	By default, the data profiling job reads each JSON schema as a single line. Select Yes to read input that spans across multiple lines.

6. Expand **Data Quality** and select **Enable Data Quality**.

Note: You can click **Use Data Profiling Parameters** to use the same parameters as in the **Data Profiling** section.

Note: Ensure that you have permissions on all the staging and flat file connections that you use in your data quality configuration. You can't run the job if you don't have permissions on the connections that you use. Select connections that you have access to, or ask the administrator to grant the necessary permissions on the connections that you want to use.

7. In the **Connection and Runtime** area, choose the Secure Agent group where you want to run catalog source jobs.
8. In the **Parameters** area, configure the parameters.

The following table describes the properties that you can enter:

Parameter	Description
Data Quality Rule Automation	<p>Enable the option to automatically create or update rule occurrences for data elements in the catalog source.</p> <p>Choose one of the following options:</p> <ul style="list-style-type: none"> - Apply on Data Elements linked with Business Dataset. Creates rule occurrences for all data elements that are linked with business data sets in the catalog source. - Apply on all Data Elements. Creates rule occurrences for all data elements in the catalog source.
Cache Result	<p>Specify how you want to preview rule occurrence results.</p> <p>Select Agent Cache if you want to generate a cache file in the runtime environment and to preview the cached results faster in subsequent data preview runs. The results are cached for seven days by default after the first run in the runtime environment.</p> <p>Select No Cache if you don't want to cache the preview results and view the live results.</p>
Run Rule Occurrence Frequency	Specify whether you want to run data quality rules based on the frequency defined for the rule occurrence in Data Governance and Catalog.
Sampling Type	You can run the data quality task on all rows for a Microsoft Azure Data Lake Storage Gen2 source system.
Elastic Runtime Environment	Select an elastic runtime environment for complex file types, including Avro and Parquet.
Staging Connection	<p>The staging connection where data quality results are stored temporarily during the execution.</p> <p>Applicable for data quality on complex file types.</p>
Maximum Precision of String Fields	The maximum precision value for profiles on string data type.
Text Qualifier	<p>The character that defines string boundaries. If you select a quote character, profiling ignores delimiters within the quotes. Select a qualifier from the list.</p> <p>Note: Default is Double Quote.</p>

Parameter	Description
Code page for delimited files	<p>Select a code page that the Secure Agent can use to read and write data. Use this option to ensure that rule results for assets with non-English characters don't include junk characters. Default value is UTF-8.</p> <p>Choose one of the following options:</p> <ul style="list-style-type: none"> - MS Windows Latin 1. Select for ISO 8859-1 Western European characters. - UTF-8. Select for Unicode and non-Unicode characters. - Shift-JIS. Select for double-byte characters. - ISO 8859-15 Latin 9 (Western European). - ISO 8859-2 Eastern European. - ISO 8859-3 Southeast European. - ISO 8859-5 Cyrillic. - ISO 8859-9 Latin 5 (Turkish). - IBM EBCDIC International Latin-1.
Escape character for delimited files	<p>You can specify an escape character if you need to override the default escape character. An escape character ignores a delimiter character in an unquoted string if the delimiter is part of the string value.</p> <p>If you specify an escape character, the data quality task overrides the default escape character that the Metadata Extraction job detects and considers the specified escape character. It then reads the delimiter character as a part of the string value. If you don't specify an escape character, the task considers the default escape character that the Metadata Extraction job detects and reads the delimiter character as a part of the string value.</p>
Read Multiple Line JSON Files	By default, the data quality job reads each JSON schema as a single line. Select Yes to read input that spans across multiple lines.

- To enable the data observability capability, expand **Data Observability** and select **Enable Data Observability**.

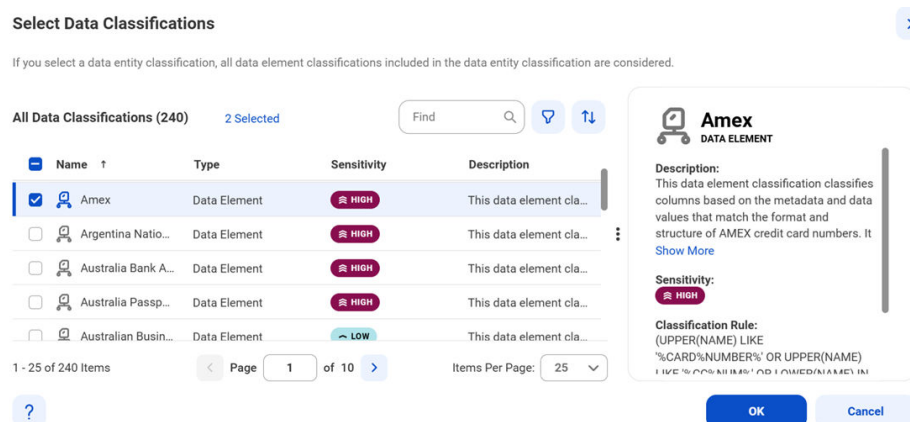
Configure data classification

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

- Click the **Data Classification** tab.
- Select **Enable Data Classification**.
- 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 relationship discovery

Enable the relationship discovery capability to identify pairs of similar columns and relationships between tables within a catalog source.

Before you configure relationship discovery, perform the following tasks:

- Import a relationship inference model. For more information about importing a relationship inference model, see [“Import a relationship inference model” on page 15](#).
- Enable data profiling on the **Data Profiling and Quality** tab, and select **Keep Signatures and Values** as the run mode in the **Parameters** section. These configurations enable you to retain values of the columns in the profiling results and discover relationships.

1. Click the **Relationship Discovery** tab.
2. Select **Enable Relationship Discovery**.
3. In the **Column Similarity** area, select the **Relationship Inference Model**.

Note: The relationship inference models that you imported appear in the **Relationship Inference Model** field.

4. In the **Joinable Tables Relationship** area, specify the **Containment Score Threshold** to identify joinable table relationships within the catalog source. This score is an indicator of the data overlap between any two given columns which determines whether the tables are joinable.

Note: A higher score means that the objects have more overlapping data and a lower score means lesser overlapping data between the two objects. A containment score threshold lower than 0.4 might result in a large number of false positives.

After you run the catalog source job, you can view the inferred relationships on the **Relationships** tab of the extracted assets in Data Governance and Catalog.

Configure glossary associations

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.

<input type="checkbox"/> Full Name	Email	User Name	Status
<input type="checkbox"/> gov owner_09			Active

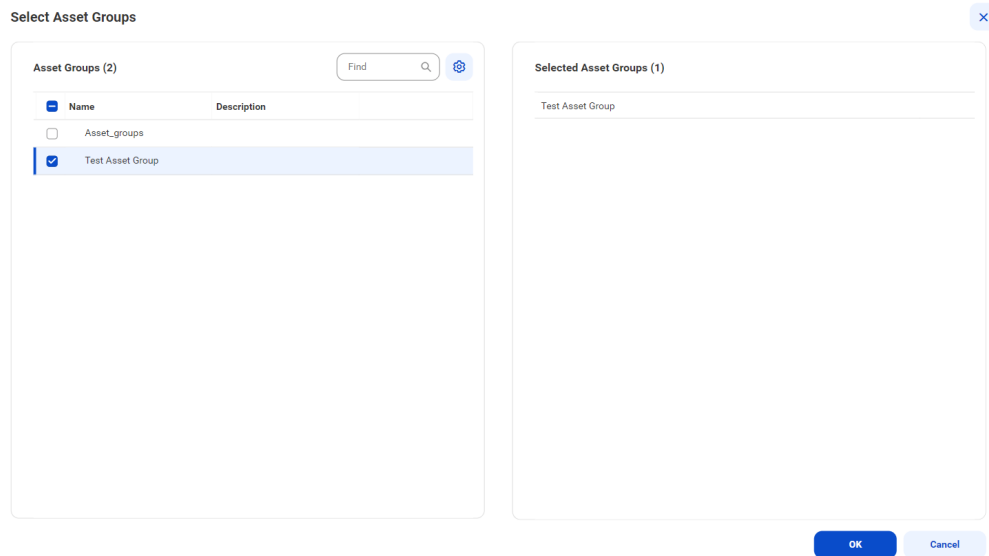
- e. Select one or more users or user groups to assign as stakeholders for the technical assets, and click **OK**.

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*.

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

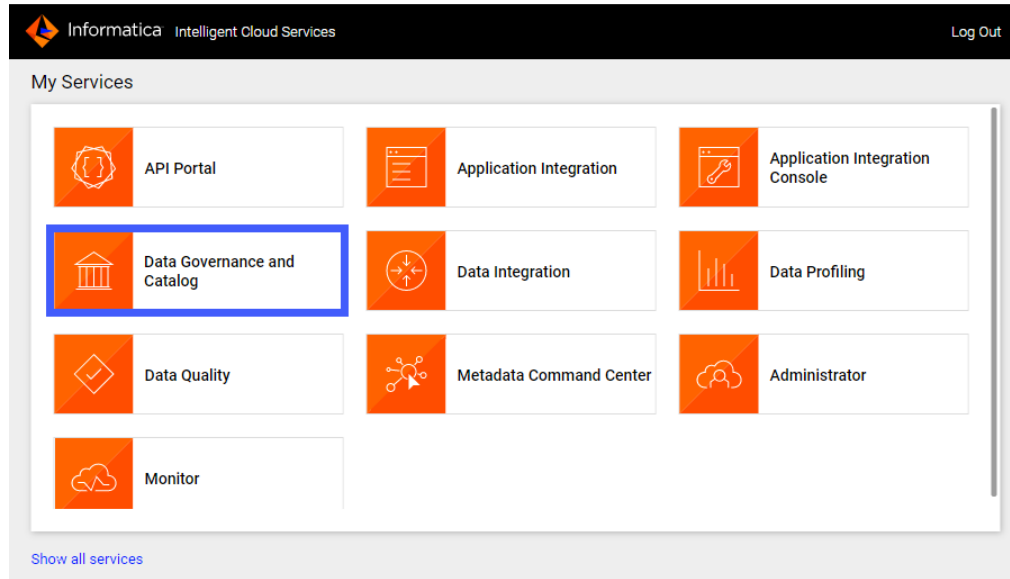
After a job runs in Metadata Command Center, view the results in Data Governance and Catalog. You can view details about source system contents in a hierarchical structure and trace data lineage.

1. Log in to Informatica Intelligent Cloud Services.

The **My Services** page appears.

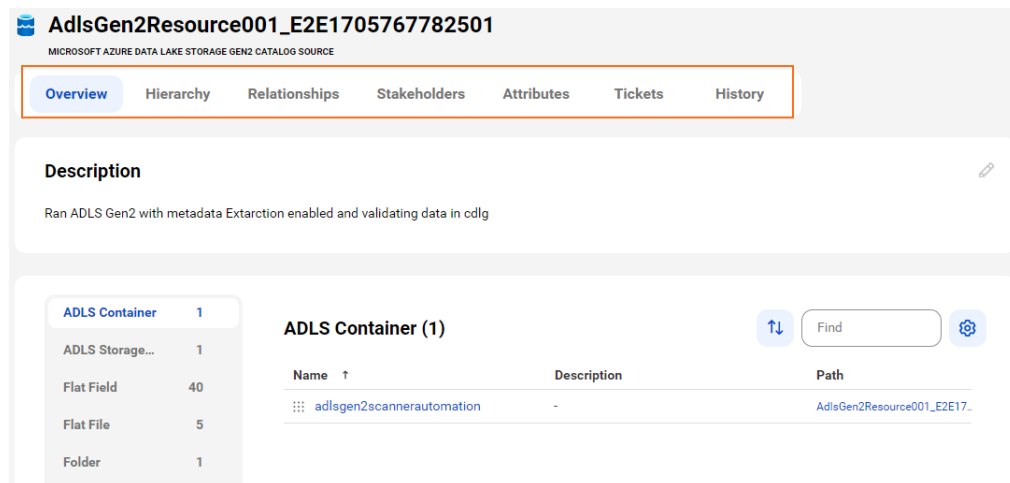
2. Click Data Governance and Catalog.

The following image shows the Data Governance and Catalog box on the **My Services** page:



3. On the Data Governance and Catalog home page, click the number in the **Technical Assets** panel. The **Technical Assets** page opens.
4. Select **Catalog Source** in the **Filter** list. The list of catalog sources opens.
5. 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 Microsoft Azure Data Lake Storage Gen2 catalog source asset page:



Note: If a new record is added to a file in the Microsoft Azure Data Lake Storage Gen2 source system, the timestamp of the file gets updated. If you then run an incremental metadata extraction job, the file gets extracted because objects are extracted based on the updated timestamp.

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

Note: If an incremental metadata extraction job includes a view, it extracts the corresponding tables even if the tables were extracted in previous runs.

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

Note: File system assets can be source and target endpoint objects for other catalog sources, for example Business Intelligence or ETL sources. File system sources don't create any references to external assets. If you didn't assign them to any reference assets created by other catalog sources, the **Lineage** tab doesn't display the data flow.

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.

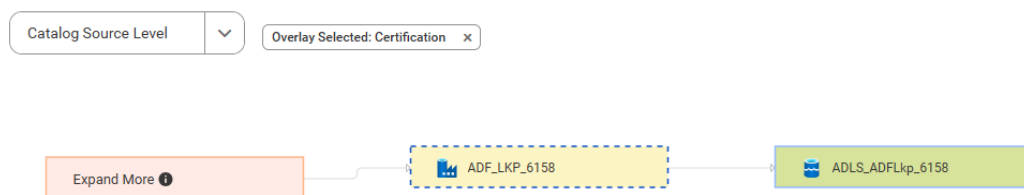
For information about linking catalog sources, see *Link catalog sources* in the Administration help.

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**.

The following image shows how the ADLS_ADFLkp_6158 catalog source gets data from the ADF_LKP_6158 catalog source:



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**.

The following image shows the data set level lineage for the CSVWithCommaDelimiter.csv file :



▼	AdlsNestedFolder_242	1
▼	Allsupportedfiles	1
⋮	CSVWithCommaDelimiter.csv	

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.

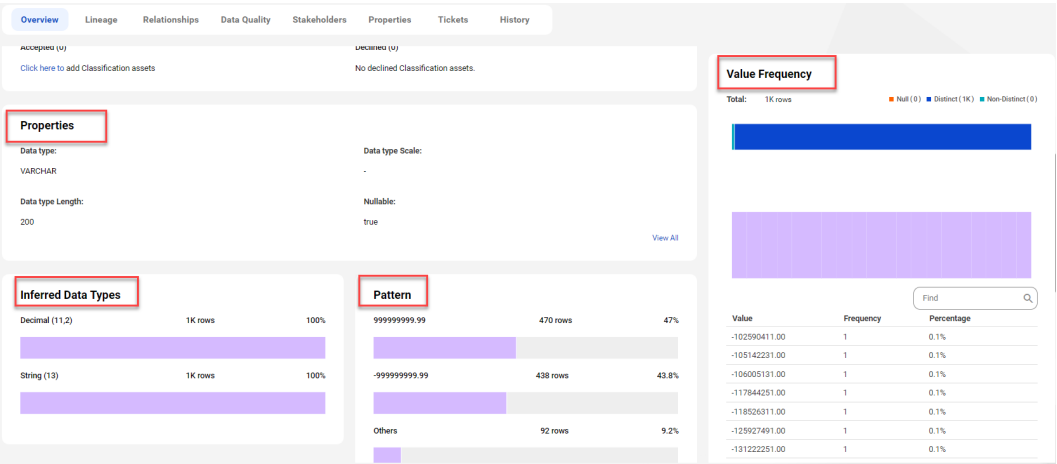
To view data lineage at the data element level, open a technical asset, click the **Lineage** tab, and then verify that the level is set to **Data Element Level**.

View data profiling results

When you enable the data profiling task for a catalog source in Metadata Command Center, the system runs a profile to evaluate the quality of the metadata extracted from the source system. The profiling statistics appear in Data Governance and Catalog when you open the technical assets.

The scope of profiling statistics that Data Governance and Catalog displays depends on the data profiling configuration parameters that you set when you configured the catalog source in Metadata Command Center.

The following image shows the data profiling statistics that appear on a column asset page in Data Governance and Catalog:



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

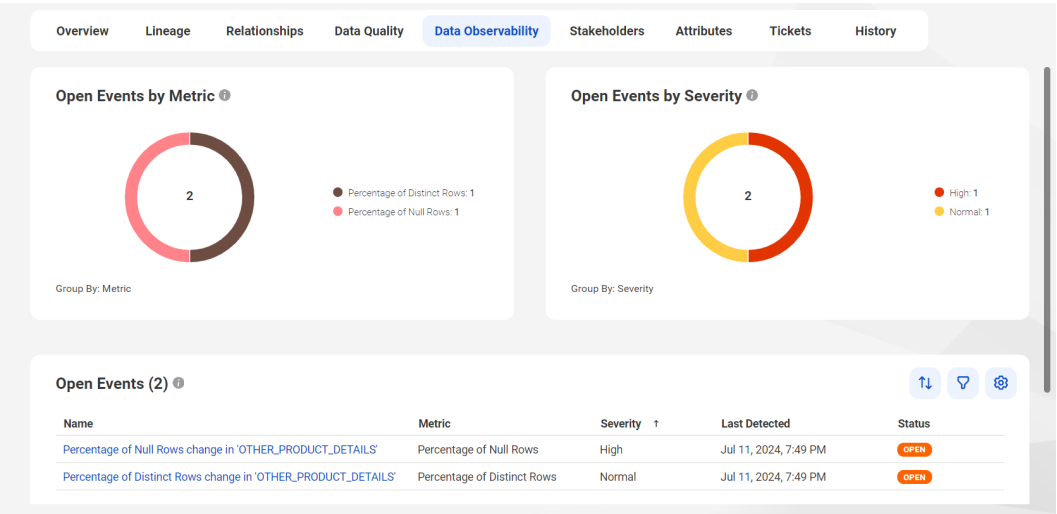
View data observability results

When you enable data observability for a catalog source in Metadata Command Center, you can view and evaluate the events that it generates in Data Governance and Catalog. These events indicate anomalies identified in the characteristics of the profiled data in your source system.

You can view the events that data observability generates for anomalies identified for catalog sources, technical data sets, and data elements. You can then take appropriate actions for the generated events.

Note: The administrator of the catalog source might have applied filters to the data to narrow down the data elements that are applicable for business users in Data Governance and Catalog. The data for which users receive anomaly notifications depend on the filters that are configured for the catalog source.

The following image shows the open events for a column asset in Data Governance and Catalog:

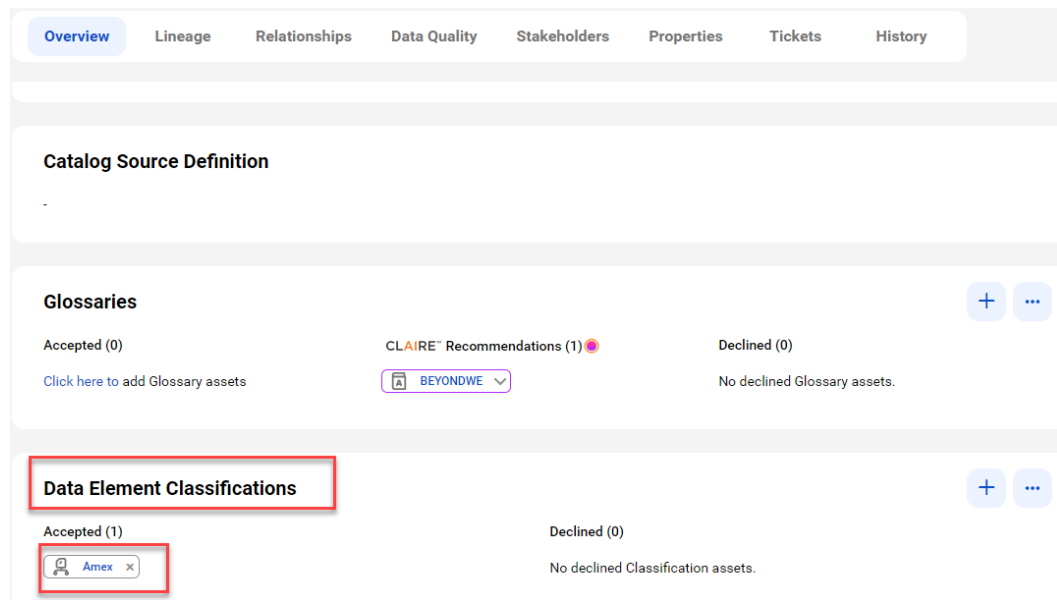


For more information about data observability results, see *Working With Assets* in the Data Governance and Catalog help.

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:

