



Informatica® Metadata Command Center  
November 2025

# IBM InfoSphere DataStage 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](mailto: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. . . . .	4
<b>Chapter 1: Introduction to IBM InfoSphere DataStage catalog sources. . . . .</b>	<b>5</b>
Extraction and view process. . . . .	6
About the IBM InfoSphere DataStage catalog source. . . . .	6
Extracted metadata. . . . .	7
<b>Chapter 2: Before you begin. . . . .</b>	<b>10</b>
Export DataStage jobs. . . . .	10
Create endpoint catalog sources for connection assignment. . . . .	11
<b>Chapter 3: Create catalog sources in Metadata Command Center. . . . .</b>	<b>12</b>
Step 1. Register a catalog source. . . . .	12
Step 2. Configure capabilities. . . . .	14
Configure metadata extraction. . . . .	14
Configure lineage discovery. . . . .	17
Configure data classification. . . . .	18
Configure glossary association. . . . .	18
Step 3. Associate stakeholders and asset groups. . . . .	19
Step 4. Run or schedule the job. . . . .	21
Step 5. Assign reference catalog source connections to endpoint catalog source objects. . . . .	22
<b>Chapter 4: View results in Data Governance and Catalog. . . . .</b>	<b>24</b>
View metadata extraction results. . . . .	24
View referenced source systems. . . . .	26
View data lineage. . . . .	26
View lineage at the catalog source level. . . . .	26
View lineage at the data set level. . . . .	27
View lineage at the data element level. . . . .	27
View classified data. . . . .	28
View glossary associations. . . . .	29

# Preface

Read *IBM InfoSphere DataStageSources* to learn how to register and configure IBM InfoSphere DataStage 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 IBM InfoSphere DataStage 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, IBM InfoSphere DataStage is a source system from which you can extract metadata through an IBM InfoSphere DataStage 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.

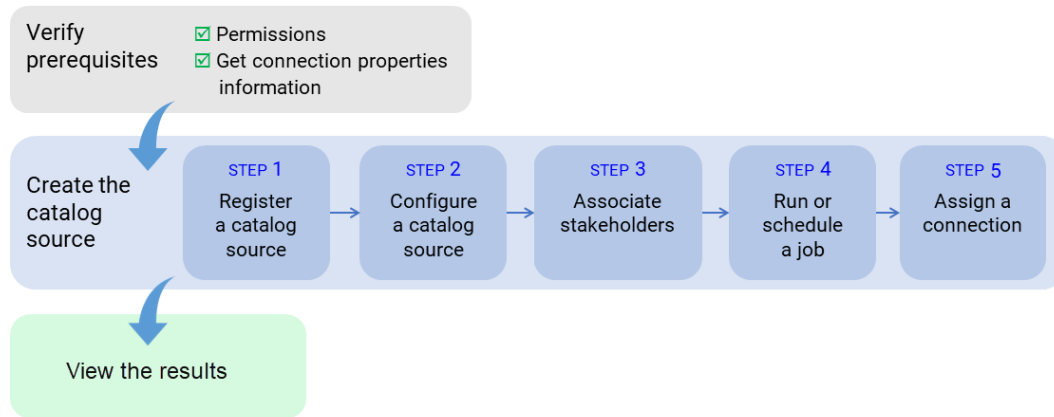
The following table describes the capabilities of the catalog source:

Capability	Description
Lineage Discovery	Builds the complete lineage of a catalog source by recommending endpoint catalog source objects to assign to reference catalog source connections. When you run the catalog source job, Metadata Command Center assigns the reference catalog source connections to CLAIRE recommended endpoint catalog source objects. You can then view the list of CLAIRE recommendations and accept or reject them.
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 an IBM InfoSphere DataStage source system:



After you verify prerequisites, perform the following tasks to extract metadata from IBM InfoSphere DataStage:

1. Register a catalog source. Create a catalog source object, select IBM InfoSphere DataStage, and specify the path to the files that contain exported DataStage jobs for metadata extraction.
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.

Run the catalog source again after you assign connections to referenced source system assets.

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

## About the IBM InfoSphere DataStage catalog source

You can use the IBM InfoSphere DataStage catalog source to extract metadata from a IBM InfoSphere DataStage source system.

IBM InfoSphere DataStage is an ETL tool to extract, transform, and load data from multiple sources to target databases or files.

## Extracted metadata

You can extract metadata from an IBM InfoSphere DataStage source system.

You can extract metadata from the following DataStage assets:

- Project
- Folder
- Sequence Job
- Parallel Job
- Server Job
- Parallel Shared Container
- Server Shared Container

IBM InfoSphere DataStage jobs consist of stages that you can add to perform specific tasks. Metadata Command Center supports the following stages:

- Parallel job stages:
  - Aggregator
  - Change Capture
  - Checksum
  - Complex Flat File
  - Copy
  - DB2 Connector
  - DRS Connector
  - Data Set
  - FTP
  - FTP Plug In
  - File Connector
  - File Set
  - Filter
  - Funnel
  - Google BigQuery Connector
  - Hierarchical Data
  - Note:** Partially compatible. You can use it as an internal component, but not as the job input or output.
  - Hive Connector
  - JDBC Connector
  - Join
  - Lookup
  - Merge
  - Modify
  - Netezza Connector
  - ODBC Connector

- Oracle Connector
- Peek
- Pivot
- Pivot Enterprise
- Remove Duplicates
- Row Generator
- Sequential File
- Slowly Changing Dimensions
- Sort
- Stored Procedure Connector
- Surrogate Key Generator
- Switch
- Teradata Connector
- Transformer
- Unstructured Data
- Xml Output
- Local Container
- Shared Container
- Snowflake Connector
- Server job stages:
  - Aggregator
  - COBOLStage
  - DB2 Connector
  - FTP
  - Hashed File
  - Netezza Connector
  - ODBC Connector
  - Oracle Connector
  - OracleOCI
  - Sequential File
  - Sort
  - Stored Procedure Connector
  - Teradata Connector
  - Transformer
  - Local Container
  - Shared Container
- Sequence job stages:
  - End Loop Activity
  - Execute Command



- Job Activity
- Start Loop Activity
- User Variables

The extracted metadata appear in Data Governance and Catalog as the following assets:

- Project
- Folder
- Job Instance
- Calculation
- Sequence Job
- Parallel Job
- Server Job

**Note:** Parallel Shared Container and Server Shared Container appear as Calculation under Job Instance in Data Governance and Catalog.

## 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:

- Export the objects that you want to extract, such as parallel jobs, to ISX files.
- Optionally, download schema files used in Sequential File stages or external SQL query files used in Database stages to the same location where you exported ISX files.
- Create endpoint catalog sources for connection assignment.

**Note:** You don't need to create a connection in Informatica Intelligent Cloud Services Administrator for IBM InfoSphere DataStage.

## Export DataStage jobs

Before you configure the catalog source, you need to export the DataStage jobs that you want to use, such as parallel jobs, to ISX files.

When you configure the catalog source, you specify the path to the files that contain exported DataStage jobs for metadata extraction.

You can use one of the following tools to export the jobs to ISX files:

- IBM InfoSphere Information Server Manager  
When you export the jobs into ISX files in IBM InfoSphere Information Server Manager, select **Include dependent items** and **Include designs** but don't select **Include executables** to reduce the size of the file and to speed up the export process.
- istool command line  
Use the `includedependent` option to include all the dependent items.

The following command exports all the jobs and dependent assets from `Jobs/Folder1` path:

```
istool.sh export -domain domainName:9443 -username user1 -password pass1 -archive  
"export.isx" -datastage '-base="engineName/projectName"' "Jobs/Folder1/*.*)" -  
includedependent'
```

**Note:** DataStage jobs that you export to ISX files can contain designs and files with extensions such as `pjb`, `sjb`, and `qjb`. They cannot contain executables or files with `bin` extension.

If the jobs that you want to export contain schema files used in Sequential File stages or external SQL query files used in Database stages, ensure that the files are included with the exported ISX file. You need to manually download the files from the DataStage Engine server into the same location where you exported the ISX file.

You can export jobs into one or more ISX files, but all dependencies for the given job must be in the same ISX file.

**Note:** For more information, see the IBM InfoSphere DataStage documentation.

## Create endpoint catalog sources for connection assignment

An endpoint catalog source represents a source system that the catalog source references. Before you perform connection assignment, create endpoint catalog sources and run the catalog source jobs.

You can then perform connection assignment to reference source systems and run connection-aware scans to view complete lineage with source system objects.

## CHAPTER 3

# Create catalog sources in Metadata Command Center

Use Metadata Command Center to configure a catalog source for IBM InfoSphere DataStage and extract metadata.

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

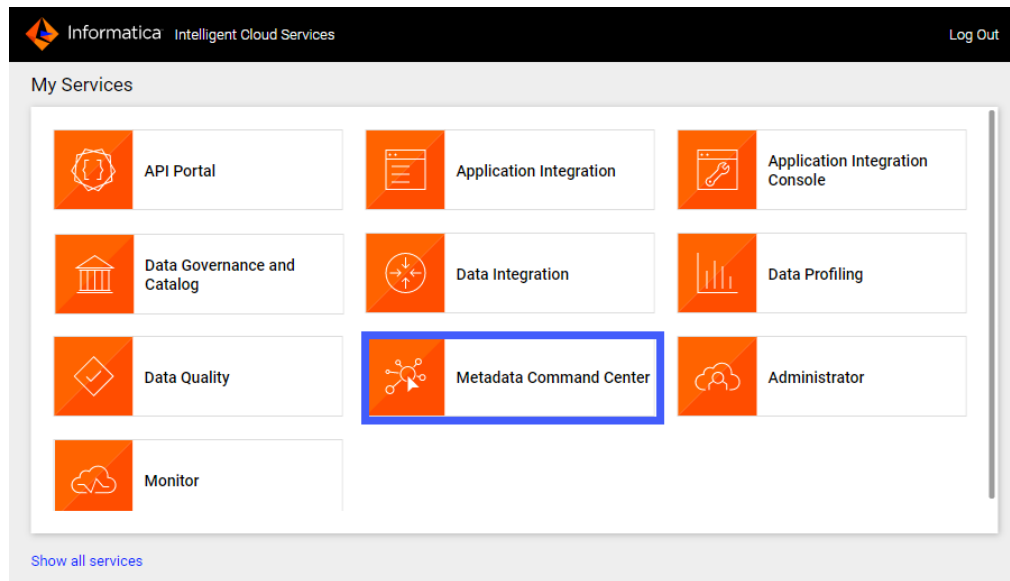
When you register a catalog source, provide general information and connection values.

1. Log in to Informatica Intelligent Cloud Services.

The **My Services** page appears.

2. Click **Metadata Command Center**.

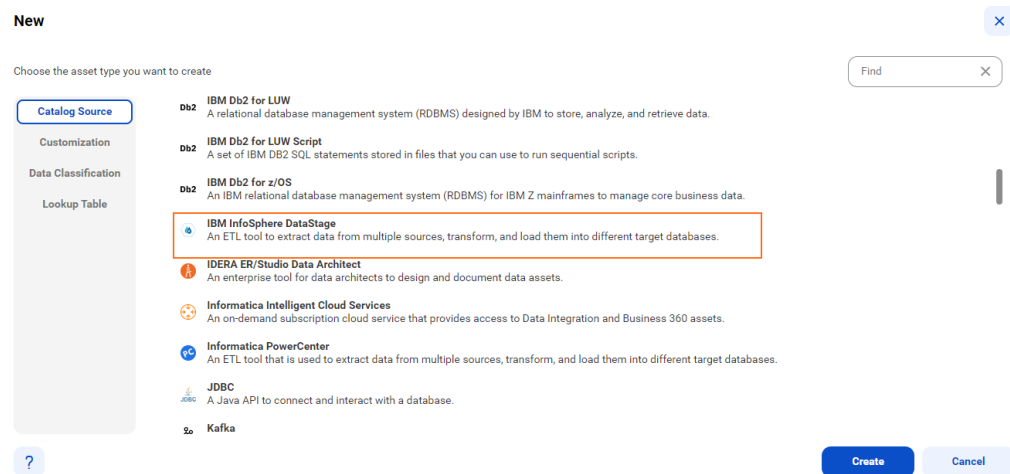
The following image shows the Metadata Command Center box on the **My Services** page:



The Metadata Command Center home page appears.

3. Click **New**.
4. Select **Catalog Source** from the list of asset types.
5. Select IBM InfoSphere DataStage from the list of catalog source types.

The following image shows the IBM InfoSphere DataStage catalog source:



6. Click **Create**.
7. 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, specify the path to the files that contain exported DataStage jobs.

Examples for Windows operating systems:

- `C:\Program Files\Informatica Cloud Secure Agent\apps\datastage file\*`: Extracts metadata from all files located in the "datastage file" folder including ISX and text files.
- `C:\Program Files\Informatica Cloud Secure Agent\apps\datastage file\*.isx`: Extracts metadata from all ISX files located in the "datastage file" folder.
- `C:\Program Files\Informatica Cloud Secure Agent\apps\datastage file*.isx`: Extracts metadata from all ISX files with names that start with "datastage file" located in the "apps" folder.

Examples for Linux operating systems:

- `${HOME}/Informatica Cloud Secure Agent/apps/datastage file/*`: Extracts metadata from all files located in the "datastage file" folder including ISX and text files.
- `${HOME}/Informatica Cloud Secure Agent/apps/datastage file/*.isx`: Extracts metadata from all ISX files located in the "datastage file" folder.
- `${HOME}/Informatica Cloud Secure Agent/apps/datastage file*.isx`: Extracts metadata from all ISX files with names that start with "datastage file" located in the "apps" folder.

9. Click **Next**.

The **Configuration** page appears.

## Step 2. Configure capabilities

When you configure the IBM InfoSphere DataStage catalog source, you define the settings for the metadata extraction capability.

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 IBM InfoSphere DataStage 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 or Exclude metadata list, choose to include or exclude metadata based on the filter parameters.
  - c. From the Object type list, select **All**, **Parallel Jobs**, **Sequence Jobs**, or **Server Jobs** depending on the object that you want to extract metadata from.
  - d. Enter the path to the object as the filter value.

Use the following rules to create filter conditions:


- Use a question mark to represents a single character.
- Use an asterisk to represent multiple characters or empty text.
- For path hierarchies, use '/' as a separator..
- When you enter values for filters, enclose them in double quotes if the path includes ":" or spaces.

The pattern for filter definition is: `Project_Name[:Directory_Path[:Job_Name]]`

Filters are case-sensitive.


The following image shows the filter condition options:


**Filters**

Specify metadata filters:  ☐ No ☒ Yes



[Show supported wildcards and examples](#)

---

Include or exclude m... 

Select the object type 

Enter a value to specify the object location

Examples:

- To include or exclude all jobs from the Jobs/Extract path, select **All** as the object type and enter `*:/Jobs/Extract:*` in the value field.
- To include or exclude all jobs that include 'backup' in the directory path, select **All** as the object type and enter `*:*backup*` in the value field.
- To include or exclude all parallel jobs with names that start with 'CopyOf' select **Parallel Jobs** as the object type and enter `*:*CopyOf*` in the value field.
- To include or exclude all parallel jobs with names from the Project1 project with names that end with 'LoadData', select **Parallel Jobs** as the object type and enter `Project1:*:*LoadData` in the value field.

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

4. Optionally, in the **Configuration Parameters** area, enter properties to override default context values and job parameters.

**Note:** Click **Show Advanced** to view all configuration parameters.

The following table describes the properties that you can enter:

Parameter	Description
Operational Metadata Directory	<p>Advanced parameter. The path to the operational metadata XML files directory. Use this parameter if you want to use operational metadata exported from IBM InfoSphere DataStage.</p> <p>For more information, see the following Informatica Knowledge Base article: <a href="#">HOW TO: Export Operational Metadata from IBM InfoSphere DataStage</a>.</p> <p><b>Note:</b> The IBM InfoSphere DataStage catalog source processes only XML files with the <code>flow</code> extension and ignores other file types in the directory.</p>
Input File Encoding	<p>Advanced parameter. The type of SQL file encoding used. Defines how the catalog source displays and processes characters. For example, UTF-8, UTF-16BE.</p> <p><b>Note:</b> Some IBM InfoSphere DataStage stages, such as the ODBC Connector, can use the SQL SELECT statements to read data. The files that contain the SQL SELECT statements can be stored in different encodings than standard UTF-8.</p> <p>Select the proper encoding to ensure that Metadata Command Center can read the input file. The default value is UTF-8.</p>
File Name Mapper	<p>Replace the DataStage file path to ensure that Metadata Command Center can resolve the paths to user files such as SQL statements and schema files in the Sequential File stage with the Schema File property.</p> <p>For example:</p> <pre>C:\RootDir\datastage=/export/home/RootDir/datastage</pre>
Parameter Override	<p>Specify the default values for job parameters and parameter sets.</p> <p>Specify parameter name using job name, parameter set name and parameter name separated by dot.</p> <p>You can use an asterisk in the job name.</p> <p>Examples:</p> <ul style="list-style-type: none"><li>- <code>jobName.parameterName=parameterValue</code></li><li>- <code>*.parameterName=parameterValue</code></li><li>- <code>*.*.parameterName=parameterValue</code></li><li>- <code>jobName.parameterSetName.parameterName=parameterValue</code></li><li>- <code>*.parameterSetName.parameterName=parameterValue</code></li></ul>

The following table describes the property that you can enter for additional settings:

**Note:** The **Additional Settings** section appears when you click **Show Advanced**.

Property	Description
Expert Parameters	<p>Enter additional configuration options to be passed at runtime. Required if you need to troubleshoot the catalog source job.</p> <p><b>Caution:</b> Use expert parameters when it is recommended by Informatica Global Customer Support.</p>

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



## Configure lineage discovery

Enable the lineage discovery capability and use CLAIRE to build complete lineage by recommending endpoint catalog source objects to assign to reference catalog source connections.

1. Click the **Lineage Discovery** tab.
2. Select **Enable Lineage Discovery**.
3. In the **Filters** area, define one or more filter conditions to apply for lineage discovery.

To define filters, you can choose to select catalog source types, asset groups, or enter a catalog source name or search from a list of catalog sources.

- a. Select **Yes** to view filter options.
- b. From the Include/Exclude list, choose to include or exclude catalog sources for lineage discovery based on the filter parameters.
- c. From the filter type list, select catalog source type, catalog source name, or asset group.
- d. In the filter value field, select the required catalog source types, or click the Search button and select catalog sources or asset groups.

Filters can contain the asterisk wildcard to represent multiple characters or empty text.

The following image shows the filter condition options:

Enable Lineage Discovery: ☒

**Filters**

Specify lineage discovery filters: ☒ No ☒ Yes

[Show supported wildcards and examples](#)

Include	Catalog Source Type	Select Catalog Source Types	+	✕
Exclude	Catalog Source Name	Select Catalog Sources	+	✕
Exclude	Asset Group	Select Asset Groups	+	✕

Examples:

- To include or exclude all Oracle catalog sources, select **Catalog Source Type** as the filter type and select `Oracle` in the filter value field.
- To include or exclude the 'Oracle\_Retail' catalog source, select **Catalog Source Name** as the filter type and search for the catalog source or enter `Oracle_Retail` in the filter value field.
- To include or exclude all catalog sources with names that start with 'Oracle', select **Catalog Source Name** as the filter type and search for the catalog source or enter `Oracle*` in the filter value field.
- To include or exclude all catalog sources with names that end with 'Retail', select **Catalog Source Name** as the filter type and search for the catalog source or enter `*Retail` in the filter value field.
- To include or exclude all catalog sources with names that contain 'Ret', select **Catalog Source Name** as the filter type and search for the catalog source or enter `*Ret*` in the filter value field.
- To include or exclude all catalog sources that are part of the 'Financial Group' asset group, select **Asset Group** as the filter type and search `Financial Group` in the filter value field.

**Note:** You can't add more than one include or exclude filter for the same filter type.

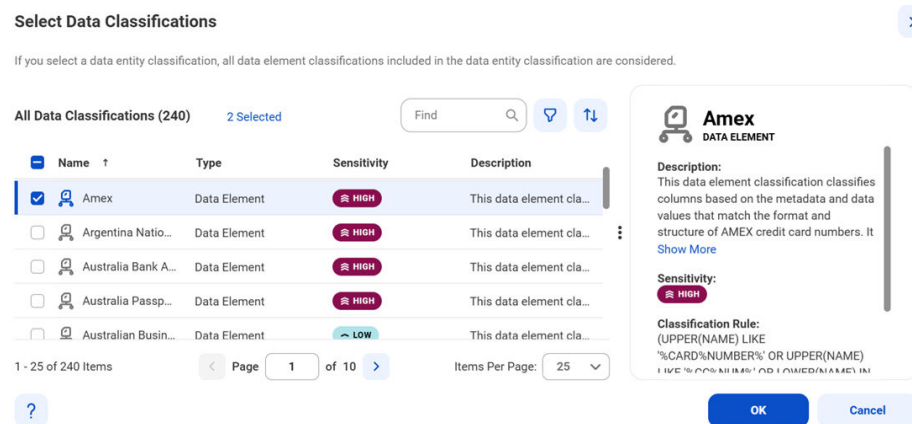
- e. Optionally, to define an additional filter with an AND condition, click the **Add** icon.

For more information about lineage discovery, see *Lineage discovery* in the *Administration* help.

## 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.

Add Users & User Groups

Users

User Groups

All Users (1)

Find

↕

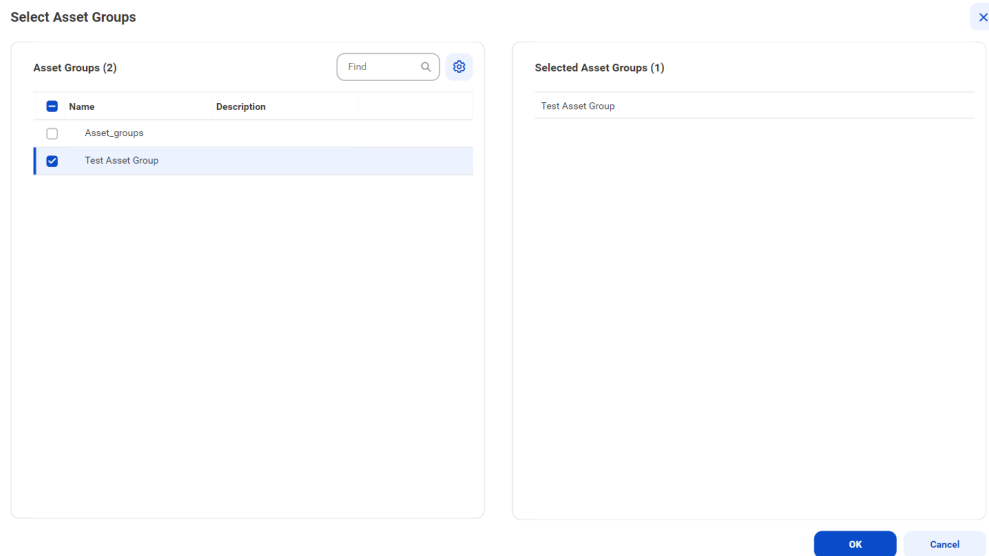
<input type="checkbox"/>	Full Name	Email	User Name ↑	Status
<input type="checkbox"/>	gov owner_09			Active

?

OK

Cancel

- 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. Assign reference catalog source connections to endpoint catalog source objects

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 an ETL tool

such as IBM InfoSphere DataStage. 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.

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 objects from the endpoint catalog sources and click **Assign**.

You can filter the list in the **Assign Connection** dialog box by name, type, or endpoint. You can connect to the following source systems:

- Oracle. The target catalog source must belong to the Database class type.
- Oracle Cloud Infrastructure GoldenGate. The target catalog source must belong to the Database class type.
- IBM Db2 for LUW. The target catalog source must belong to the Database class type.
- JDBC. The target catalog source must belong to the Database class type.
- SFTP File System. The target catalog source must belong to the File System class type.
- Teradata Database. The target catalog source must belong to the Database class type.
- Apache Hive. The target catalog source must belong to the Database class type.
- Microsoft SQL Server. The target catalog source must belong to the Database class type.

4. Run the catalog source job again. If you configured the catalog source job to run on a regular schedule, the next scheduled run picks up the updated details. If you didn't configure a schedule, run the catalog source job again to view complete lineage.

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 in a hierarchical structure and trace data lineage.

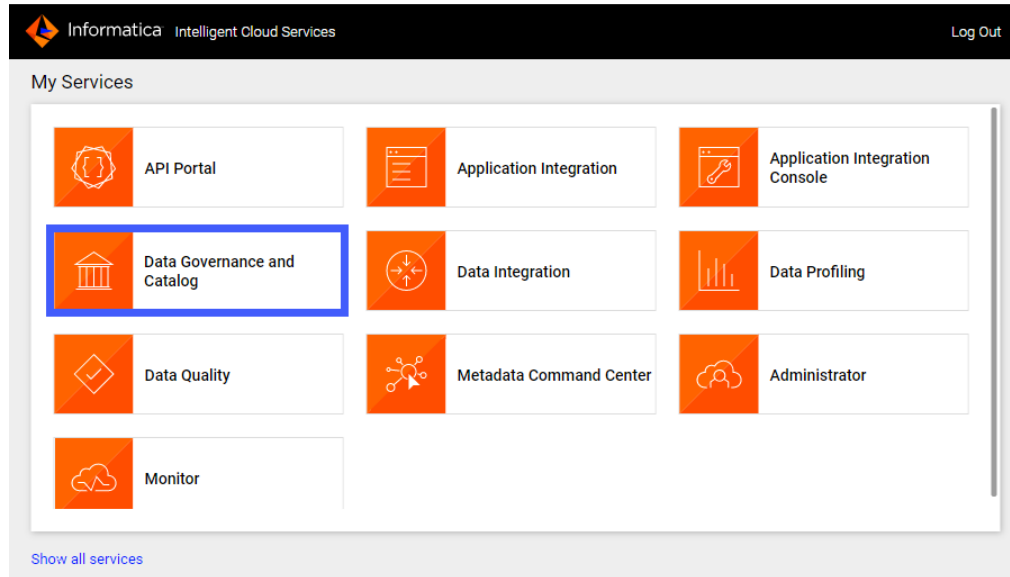
1. Log in to Informatica Intelligent Cloud Services.

The **My Services** page appears.

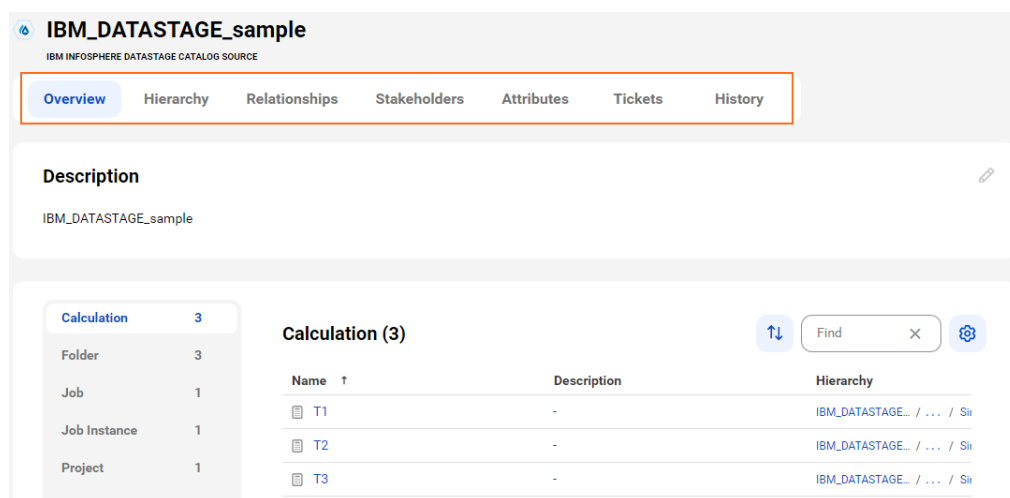
2. Click Data Governance and Catalog.

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 asset page:



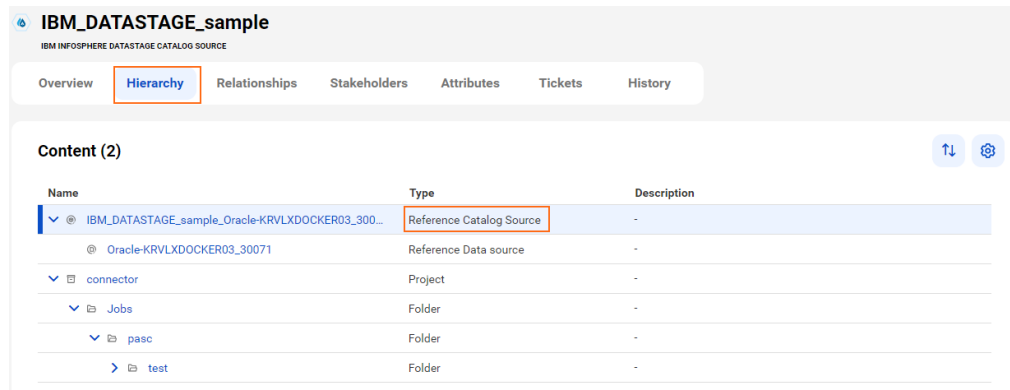
6. View the asset from different perspectives by clicking on the tabs. For more information about working with assets, see *Working with Assets* in *Data Governance and Catalog* help.

## View referenced source systems

Drill down into the referenced source systems that are referenced by catalog sources.

1. Browse to the catalog source that you want to view.
2. Click the **Hierarchy** tab to view the catalog source as a hierarchy.
3. Find an asset with the type **Reference Catalog Source**.

The following image shows an expanded view of a referenced catalog source:



4. Click to expand the referenced source to see its components.

## 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.

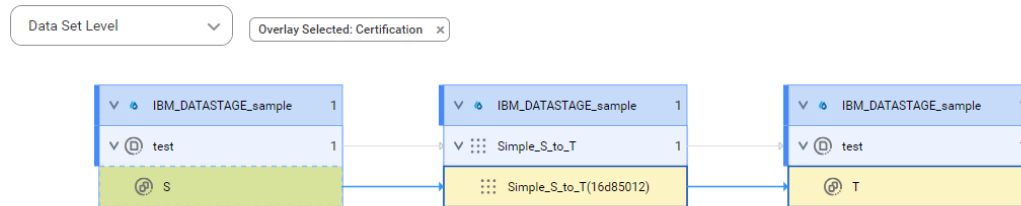
To view data lineage at the catalog source level, open a technical asset, click the **Lineage** tab, and then verify that the level is set to **Catalog Source Level**.

## View lineage at the data set level

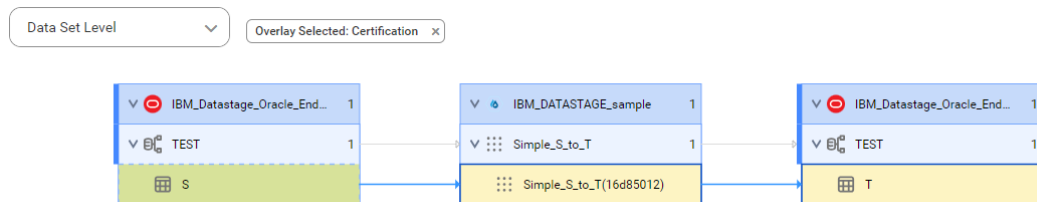
The data set level displays individual sets of data in the data flow.

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 target table gets data from the referenced source table before connection assignment:



The following image shows how the target table gets data from the actual source table after connection assignment:

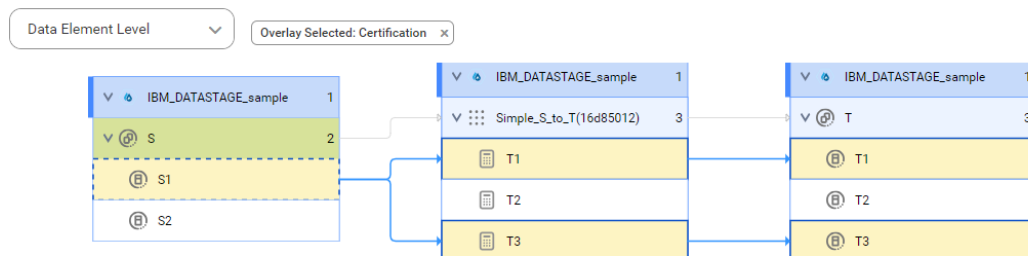


## View lineage at the data element level

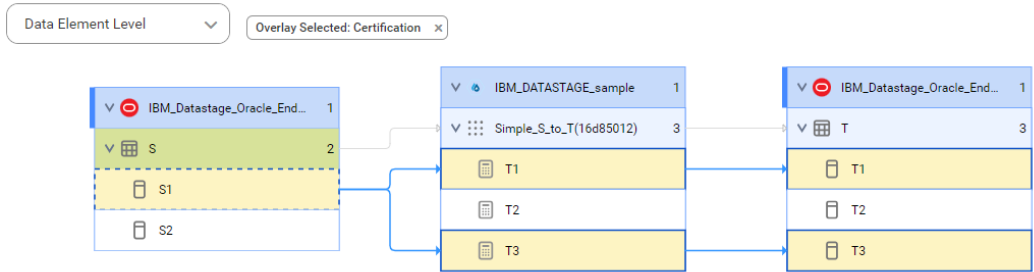
The data element level displays details of the data set level. At the data element level, you can see the input sources for expressions or commands and calculations or transformations on the data.

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**.

The following image shows how the target columns get data from one column of the referenced source table before connection assignment:



The following image shows how the target columns get data from one column of the actual source table after connection assignment:



# 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:

Overview

Lineage

Relationships

Data Quality

Stakeholders

Properties

Tickets

History

Catalog Source Definition

Glossaries

Accepted (0)

Click here to add Glossary assets

CLAIRE™ Recommendations (1)

BEYONDWE

Declined (0)

No declined Glossary assets.

Data Element Classifications

Accepted (1)

Amex

Declined (0)

No declined Classification assets.

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:

