



Informatica® Metadata Command Center  
November 2025

# AWS Glue Sources

© Copyright Informatica LLC 2022, 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

<b>Preface .....</b>	<b>4</b>
<b>Chapter 1: Introduction to AWS Glue catalog sources.....</b>	<b>5</b>
Extraction and view process. ....	6
About the AWS Glue catalog source. ....	6
Extracted metadata. ....	7
Compatible functionalities. ....	7
<b>Chapter 2: Before you begin.....</b>	<b>11</b>
Verify permissions and privileges. ....	11
Create a connection. ....	12
Permanent IAM credentials. ....	12
EC2 instance profile. ....	13
Get AWS Glue source information. ....	13
<b>Chapter 3: Create catalog sources in Metadata Command Center.....</b>	<b>16</b>
Step 1. Register a catalog source. ....	16
Step 2. Configure capabilities. ....	18
Configure metadata extraction. ....	18
Configure lineage discovery. ....	20
Step 3. Associate stakeholders and asset groups. ....	21
Step 4. Run or schedule the job. ....	23
Step 5. Assign reference catalog source connections to endpoint catalog source objects. ....	24
<b>Chapter 4: View results in Data Governance and Catalog.....</b>	<b>26</b>
View metadata extraction results. ....	26
View data lineage. ....	28
View lineage at the catalog source level. ....	29
View lineage at the data set level. ....	29
View lineage at the data element level. ....	30

# Preface

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

## CHAPTER 1

# Introduction to AWS Glue 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, AWS Glue is a source system from which you can extract metadata through an AWS Glue 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.

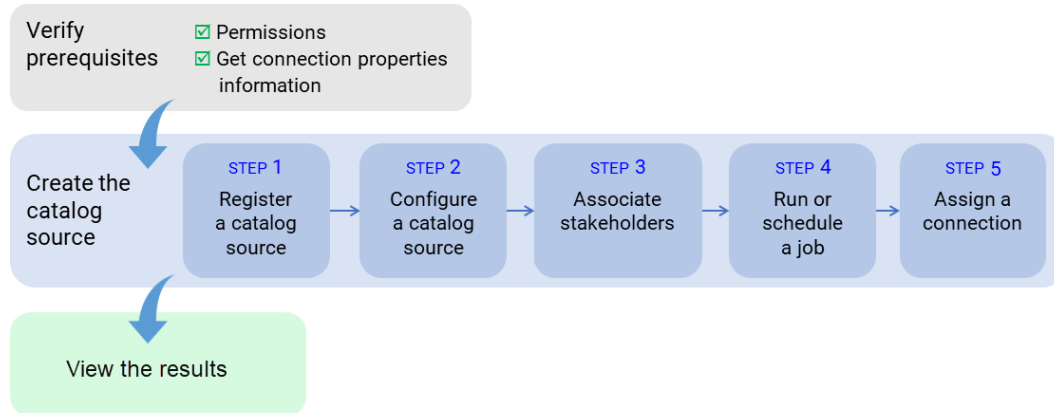
The following table describes the capabilities of the catalog source:

Capability	Description
Advanced Programming Language Parsing	Advanced Programming Language Parsing parses the source system code in addition to extracting objects from the source system.
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.

# 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 AWS Glue:

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

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

## About the AWS Glue catalog source

You can use the AWS Glue catalog source to extract metadata from an AWS Glue source system.

AWS Glue is a serverless ETL (extract, transform, and load) service that helps discover, prepare, and integrate data from multiple sources for analysis, machine learning, and application development.

You use Amazon Athena to query databases and tables created in AWS Glue. You can also use Amazon Athena to create schemas to use in AWS Glue.

# Extracted metadata

You can use the AWS Glue catalog source to extract metadata from an AWS Glue source.

Metadata Command Center extracts the following metadata from the AWS Glue source system:

- Calculation
- Job
- Job instance

# Compatible functionalities

AWS Glue offers integration with a diverse range of modules and the Python programming language.

You can use AWS Glue with the following Python functionalities:

- Standard language constructions
- Standard built-in functions
- Partially-compatible modules:

**Note:** Data Governance and Catalog processes only a subset of library functions of partially-compatible modules.

- abs
- adal
- argparse
- array
- ast
- awsglue
- azure
- base64
- binascii
- calendar
- codecs
- collections
- concurrent
- contextlib
- contextvars
- copy
- copyreg
- csv
- dataclasses
- datetime
- decimal

- delta
- difflib
- distutils
- email
- enum
- errno
- fnmatch
- fractions
- functools
- gc
- genericpath
- gettext
- glob
- graphframes
- hashlib
- heapq
- hmac
- importlib
- inspect
- io
- itertools
- json
- keyword
- locale
- logging
- math
- matplotlib
- nt
- numbers
- numpy
- operator
- os
- pandas
- pathlib
- pickle
- pkgutil
- posix
- posixpath
- pprint



- py4j
- pyodbc
- pyspark
- pytz
- random
- re
- reprlib
- requests
- seaborn
- secrets
- shutil
- simplejson
- six
- sklearn
- smtplib
- socket
- ssl
- stat
- string
- struct
- subprocess
- sys
- teradatasql
- textwrap
- threading
- time
- traceback
- types
- typing
- urllib
- urllib3
- uuid
- warnings
- weakref
- xml
- yaml
- zipfile
- zlib
- Custom libraries

**Note:** Custom libraries are libraries created by a user. You can also use a WHL file for your custom library.

If the catalog source detects an incompatible function or library, it can't process the statement. It skips the statement and continues to process the next one.

## CHAPTER 2

# Before you begin

Before you can extract catalog source metadata, get information from the AWS Glue administrator.

Perform the following prerequisite tasks:

- Verify permissions and privileges.
- Get AWS Glue source information.
- Create an Amazon Athena connection.
- Install the Secure Agent on an Amazon Elastic Compute Cloud (EC2) system.

## Verify permissions and privileges

To extract AWS Glue metadata, you need account access and permissions to the AWS Glue and Amazon Athena source systems.

Verify that the administrator performs the following tasks:

- Creates a user account for the Informatica user to access the AWS Glue source system.
- Configures the read permission on the AWS Glue source for the user account.
- Grants the following Identity and Access Management (IAM) permissions to the user to access all user-defined databases:

```
glue:BatchGetJobs
glue:GetDatabases
glue:GetJobRuns
glue:GetCatalogImportStatus
glue:GetJobs
glue:ListJobs
```

- Grants the following IAM permissions to the user to perform operations on Amazon S3 buckets:

```
s3:GetObject
s3:ListBucket
```

- Grants the required permissions to extract metadata from the Amazon Athena source. See *Amazon Athena* in the Catalog Source Configuration help.

# Create a connection

You use the Amazon Athena connection to connect to the Amazon Athena source system and create schema to use in AWS Glue. Create an Amazon Athena connection object in Administrator.

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

Connection 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.

4. Select the Amazon Athena connection type.
5. Enter properties specific to the Amazon Athena connection:

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. <b>Note:</b> If you're using this connection to apply data access policies through pushdown or proxy services, you cannot use the Secret Vault configuration option. For information about how to configure and use a secrets manager, see "Secrets manager configuration" in the Administrator help.
Runtime Environment	The name of the runtime environment where you want to run tasks. For more information about how to configure and use the runtime environments, see <i>Runtime Environments</i> in the Administrator help.

6. Select the authentication type to connect to Amazon Athena and enter the required properties.  
You can select one of the following authentication types:
  - Permanent IAM credentials
  - EC2 instance profile
7. Click **Test Connection**.
8. Click **Save**.

## Permanent IAM credentials

Permanent IAM credentials authentication is the default type that requires the access key and secret key of the IAM user to connect to Amazon Athena.

The following table describes the basic connection properties for permanent IAM credentials authentication:

Property	Description
Access Key	The access key of the IAM user to connect to Amazon Athena.
Secret Key	The secret key of the IAM user to connect to Amazon Athena.
JDBC URL	<p>The URL to connect to Amazon Athena.</p> <p>Enter the JDBC URL in the following format:</p> <pre>jdbc:awsathena:// AwsRegion=&lt;region_name&gt;;S3OutputLocation=&lt;S3_Output_Location&gt;;Workgroup=&lt;Workgroup_Name&gt;;</pre> <p><b>Note:</b> If you use a workgroup with customer managed query results, specify at least one of the two parameters in the JDBC URL, either the S3 output location or the workgroup name. For a workgroup with Athena managed query results, specify only the workgroup name and do not include the S3 output location in the JDBC URL.</p>

## EC2 instance profile

You can configure AWS Identity and Access Management (IAM) authentication to connect to Amazon Athena when the Secure Agent is installed on an Amazon Elastic Compute Cloud (EC2) system.

The following table describes the basic connection properties for EC2 instance profile authentication:

Property	Description
JDBC URL	<p>The URL of the Amazon Athena connection.</p> <p>Enter the JDBC URL in the following format:</p> <pre>jdbc:awsathena:// AwsRegion=&lt;region_name&gt;;S3OutputLocation=&lt;S3_Output_Location&gt;;Workgroup=&lt;Workgroup_Name&gt;;</pre> <p><b>Note:</b> If you use a workgroup with customer managed query results, specify at least one of the two parameters in the JDBC URL, either the S3 output location or the workgroup name. For a workgroup with Athena managed query results, specify only the workgroup name and do not include the S3 output location in the JDBC URL.</p>

## Get AWS Glue source information

Get the connection properties that you need to configure from the AWS Glue administrator.

**Note:** You don't need to create a connection object for AWS Glue. You provide this information when you configure the catalog source.

The following table describes the properties that you need:

Property	Description
Athena Connection	The Amazon Athena connection object.
Region	The Amazon Web Services region from where you want to run the catalog source job.
Authentication mode	Select the authentication type to connect to Amazon Web Services account. You can select one of the following authentication types: <ul style="list-style-type: none"><li>- Basic</li><li>- Assume Role.</li><li>- IAM Roles Anywhere</li></ul>

#### Basic authentication

This is the default method of authentication. Provide an access key and security key to access the Amazon Web Services account.

The following table describes the connection properties for basic authentication:

Property	Description
Access Key	The access key of the Amazon Web Services account.
Security Key	The security key of the Amazon Web Services account.

#### Assume Role authentication

Assume Role authentication allows a user or service temporarily inherit permissions from another role. Instead of using permanent credentials, you assume an IAM role to get temporary security credentials. This allows you to access AWS resources securely without sharing credentials.

Provide the IAM Role ARN and, optionally, provide the access key and security key to access the Amazon Web Services account.

**Note:** Verify that the administrator granted the minimum user permission to access the AWS Glue and Amazon Athena source systems.

The following table describes the connection properties for Assume Role authentication:

Property	Description
IAM Role ARN	The Amazon Resource Name (ARN) of the AWS Identity and Access Management (IAM) role that the user assumes to use. The user uses the dynamically generated temporary security credentials. For more information about how to get the ARN of an IAM role, see the AWS documentation.
Access Key	Optional. The access key of the Amazon Web Services account.
Security Key	Optional. The security key of the Amazon Web Services account.

#### IAM Roles Anywhere authentication

IAM Roles Anywhere authentication allows an external application, user, or system of AWS securely access AWS resources with the X.509 certificates instead of AWS login credentials. This makes it easier and safer to manage access across different environments.

You can provide the credential file path and profile name to access the Amazon Web Services account. The following table describes the connection properties for IAM Roles Anywhere authentication:

Property	Description
Credential file path	The location of the file containing the credentials used to authenticate the user. For more information about how to get the AWS credential file path, see the AWS documentation.
Profile name	The profile name that you defined in the credential file for user authentication. If you don't provide the profile name, the authentication process uses the default profile.

## CHAPTER 3

# Create catalog sources in Metadata Command Center

Use Metadata Command Center to configure a catalog source for AWS Glue and run the catalog source job.

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

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

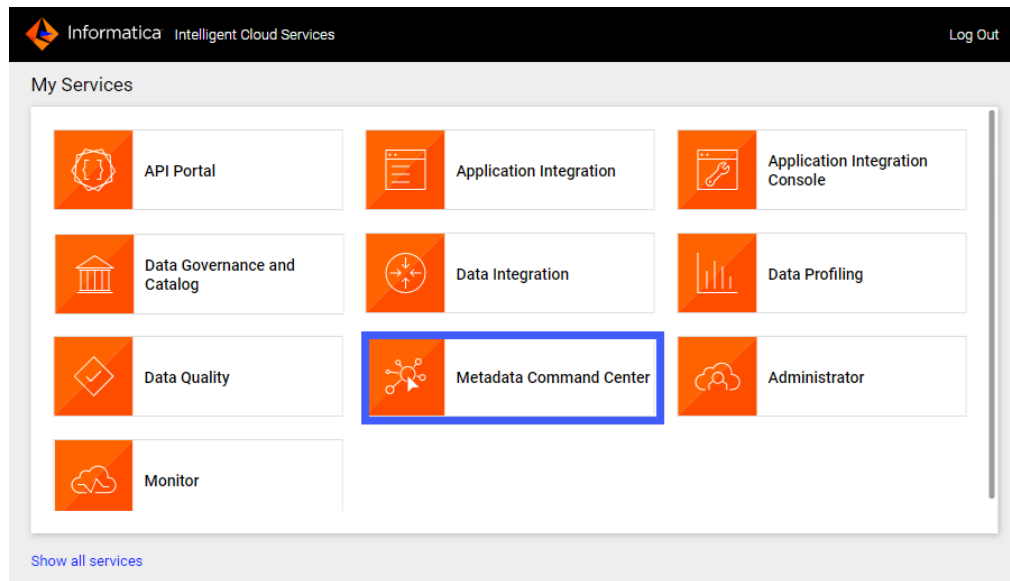
## Step 1. Register a catalog source

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

1. 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:

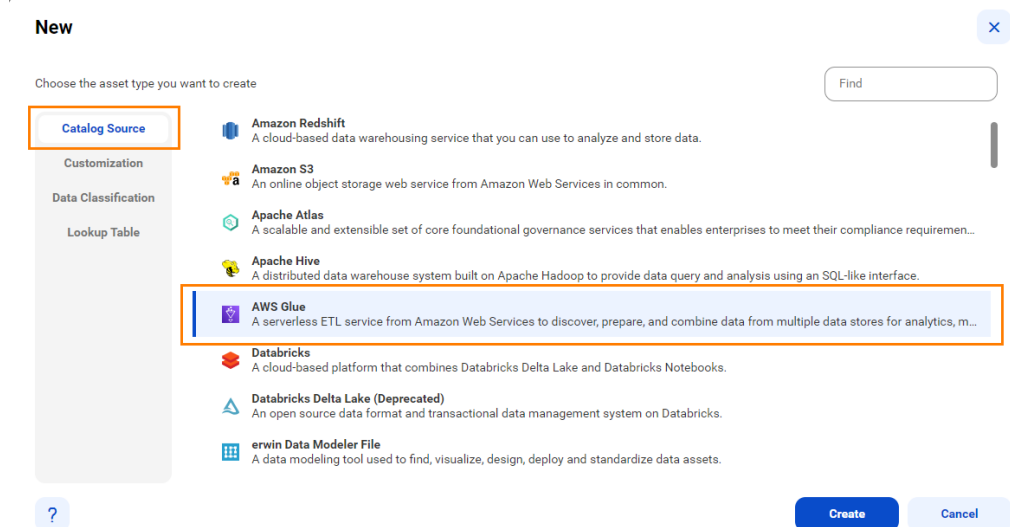




The Metadata Command Center home page appears.

3. Click **New**.
4. Select **Catalog Source** from the list of asset types.
5. Select AWS Glue from the list of source systems.

The following image shows the page where you select the AWS Glue source system:



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, enter the AWS Glue connection information based on the connection values that you got from the administrator.

The following table describes the properties to configure:

Property	Description
Athena Connection	The Amazon Athena connection object.
Region	The Amazon Web Services region from where you want to run the catalog source job.
Authentication mode	Select the authentication type to connect to Amazon Web Services account. You can select one of the following authentication types: <ul style="list-style-type: none"><li>- Basic</li><li>- Assume Role.</li><li>- IAM Roles Anywhere</li></ul>

9. Click **Next**.

The **Configuration** page appears.

## Step 2. Configure capabilities

When you configure the AWS Glue 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 AWS Glue 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. From the **Include or exclude metadata** list, choose to include or exclude metadata based on the filter parameters.
  - b. From the **Object type** list, select JobName to filter the metadata based on the job names of the AWS Glue jobs.
  - c. Enter a value to specify the object location.  
Filter values can contain the following wildcards:
    - Question mark. Represents a single character.
    - Asterisk. Represents multiple characters or empty text.
 The following image shows the filter options:

**Filters**

Specify metadata filters: ☐ No ☒ Yes

[Show supported wildcards and examples](#)

Include Metadata	JobName	Enter a value to specify the object location	+	🗑️
------------------	---------	--	---	----

- d. Optionally, to define an additional filter with an OR condition, click the **Add** icon.
- The following image shows the filter conditions for an AWS Glue catalog source:

**Filters**

Specify metadata filters: ☐ No ☒ Yes

[Show supported wildcards and examples](#)

Include Metadata	JobName	CreateTableFromS?	+	🗑️
Include Metadata	JobName	Integration.Job*	+	🗑️

The filter includes metadata from AWS Glue jobs with names that start with CreateTableFromS followed by a single character in addition to jobs with names that start with IntegrationJob.

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

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

Property	Description
Runs Per Job	The number of job runs to fetch. Default is 1. If you don't specify a value, the catalog source job fetches the last job run. The catalog source job only extracts unique job instances that run successfully and differ in lineage.

The following table describes the properties that you enter for Amazon Athena Catalog Preload Filters:

Property	Description
Include filter	<p>A list of filters to preload Amazon Athena catalog assets. Use the include filter to load a limited set of assets and optimize job time. A job processes an asset when it matches at least one include filter.</p> <p>A filter value contains segments separated by periods. You can enter two wildcards in each segment. Use a question mark to represent a single character and an asterisk to represent multiple characters.</p> <p>The filter segments contain the Amazon Athena database name and the asset name, such as <code>&lt;Database name&gt;.&lt;Asset name&gt;</code></p> <p>To configure a filter, click the <b>Add</b> icon and provide a value in the Value field.</p>
Exclude filter	<p>A list of filters to exclude Amazon Athena catalog assets. A job doesn't process an asset when it matches any of the exclude filters.</p> <p>A filter value contains segments separated by periods. You can enter two wildcards in each segment. Use a question mark to represent a single character and an asterisk to represent multiple characters.</p> <p>The filter segments contain the Amazon Athena database name and the asset name, such as <code>&lt;Database name&gt;.&lt;Asset name&gt;</code></p> <p>To configure a filter, click the <b>Add</b> icon and provide a value in the Value field.</p>

- Optional. In the **Configuration Parameters** area, enter additional settings.

The following table describes the property that you 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>

- 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.

- Click the **Lineage Discovery** tab.
- Select **Enable Lineage Discovery**.
- 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:

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.

## 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.

	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.

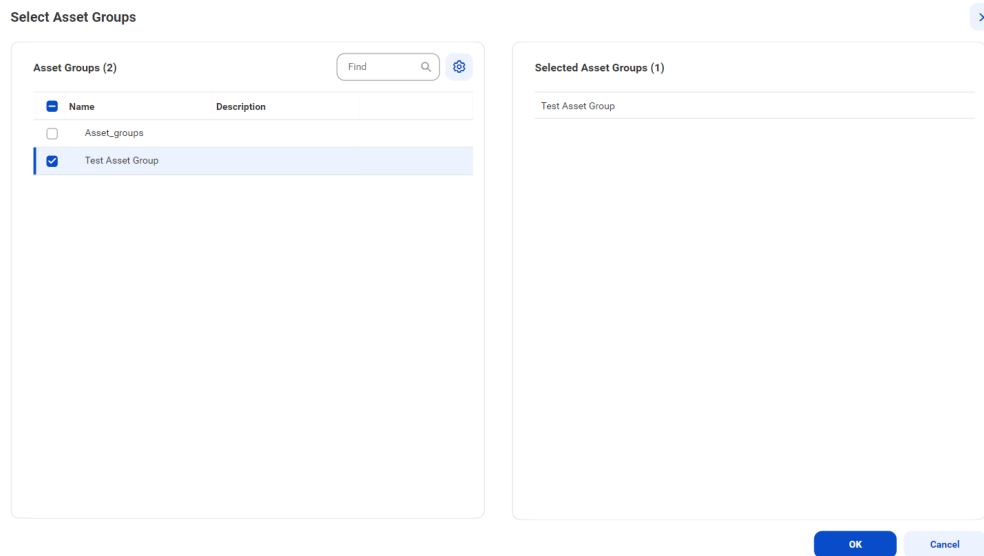
- f. To assign users or user groups from another stakeholder role, click **Add** and then repeat the steps.
2. To assign asset groups to technical assets extracted from the catalog source, perform the following steps:

- 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*.

## 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 referenced source system might be a database, such as Oracle.

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 catalog sources to assign to the selected connection and click **Assign**.

You can assign an Amazon Athena source system as a referenced source system. To create a connection assignment to Amazon Athena catalog sources, the referenced catalog source must belong to the Database class type.

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.

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.

## 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

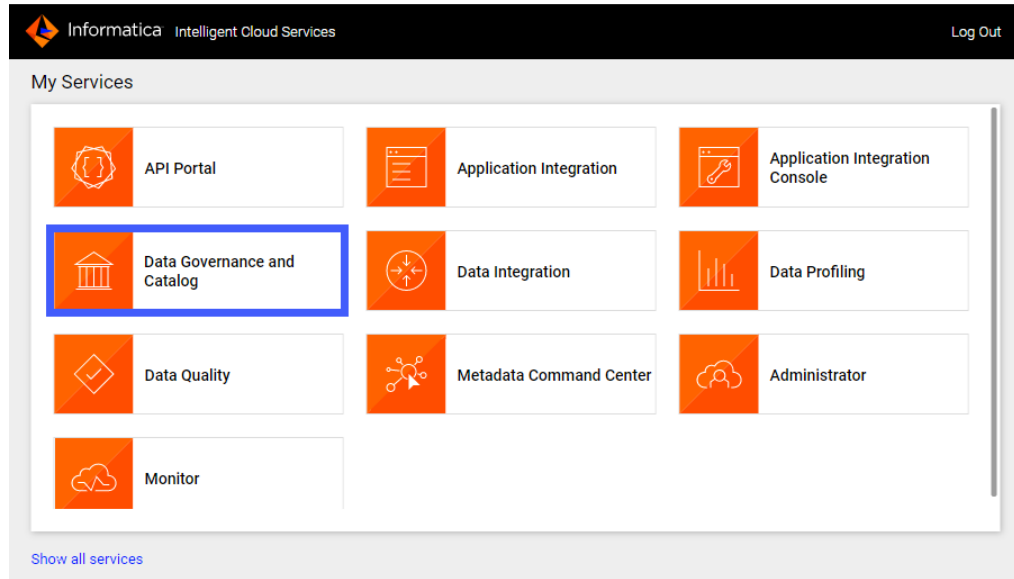
After a job runs in Metadata Command Center, view the results in Data Governance and Catalog. You can view details about source system contents in a hierarchical structure and trace data lineage.

1. Log in to Informatica Intelligent Cloud Services.

The **My Services** page appears.

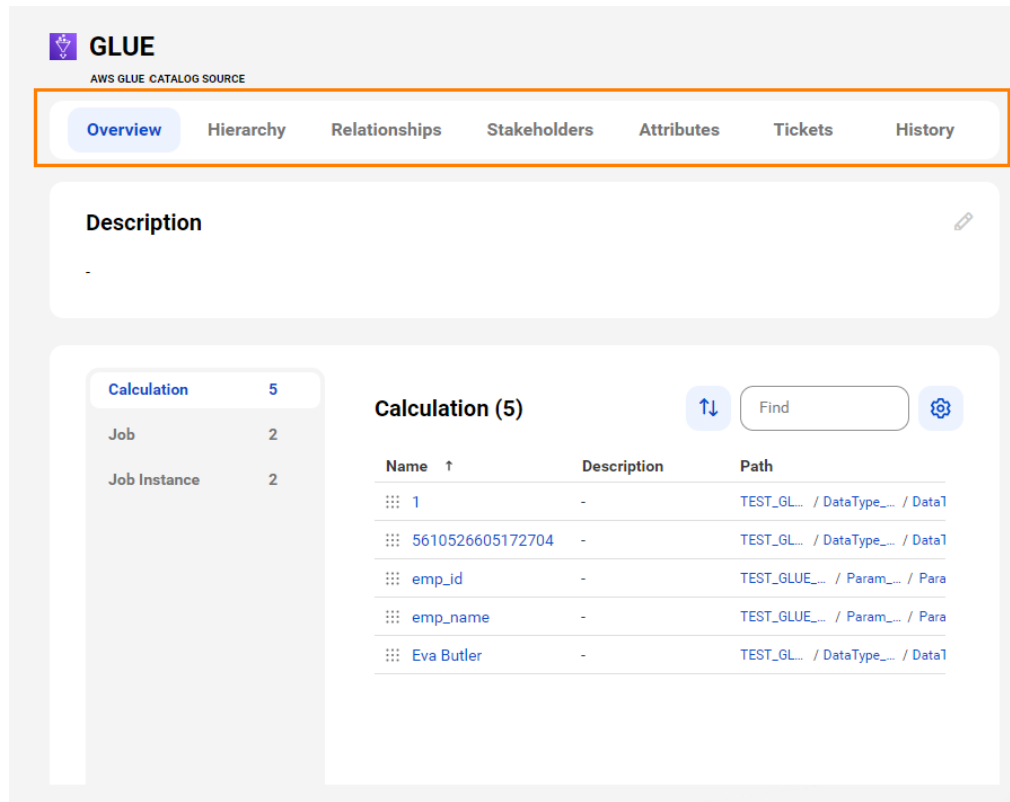
2. Click Data Governance and Catalog.

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 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.

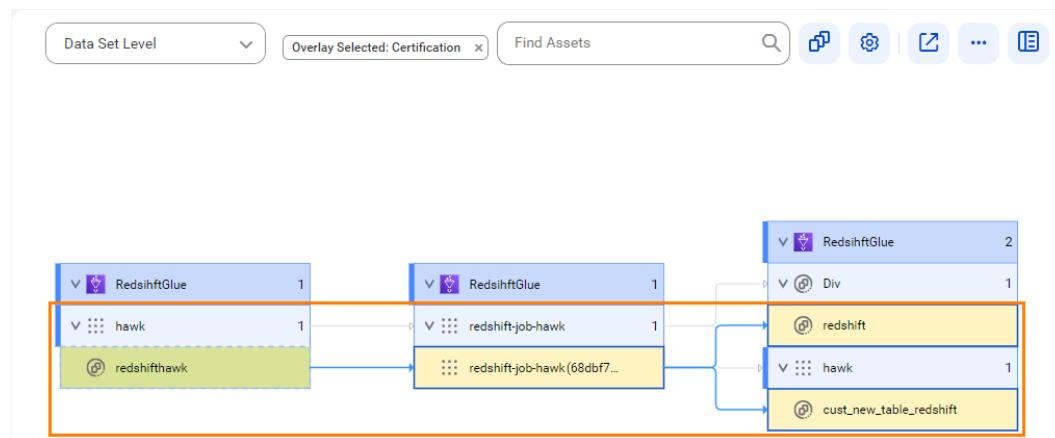
To view data lineage at the catalog source level, open a technical asset, click the **Lineage** tab, and then verify that the level is set to **Catalog Source Level**.

## View lineage at the data set level

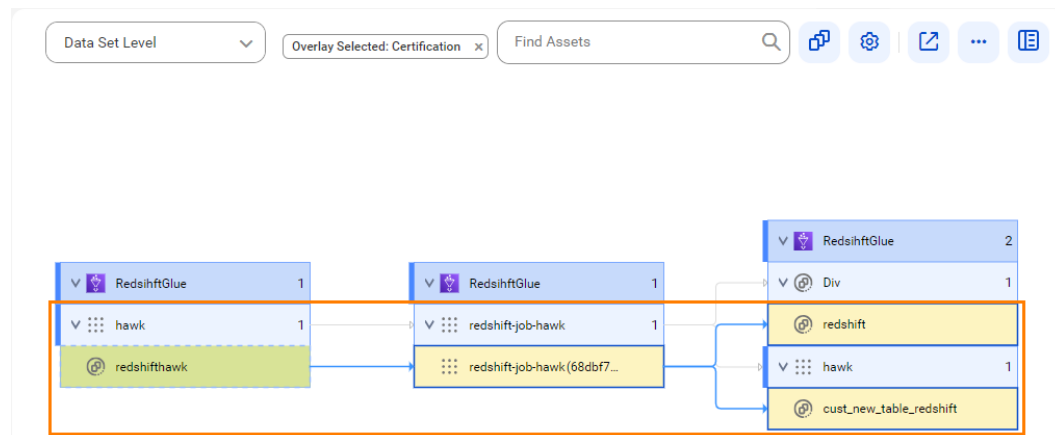
The data set level is a view that shows individual sets of data in the data flow.

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 how the redshift-job-hawk job instance uses data from the redshifthawk referenced data set to generate output to the target redshift and cust\_new\_table\_redshift referenced data sets before connection assignment:



The following image shows how the redshift-job-hawk job instance uses data from the redshifthawk actual data set to generate output to the target redshift and cust\_new\_table\_redshift actual data sets after connection assignment:



## View lineage at the data element level

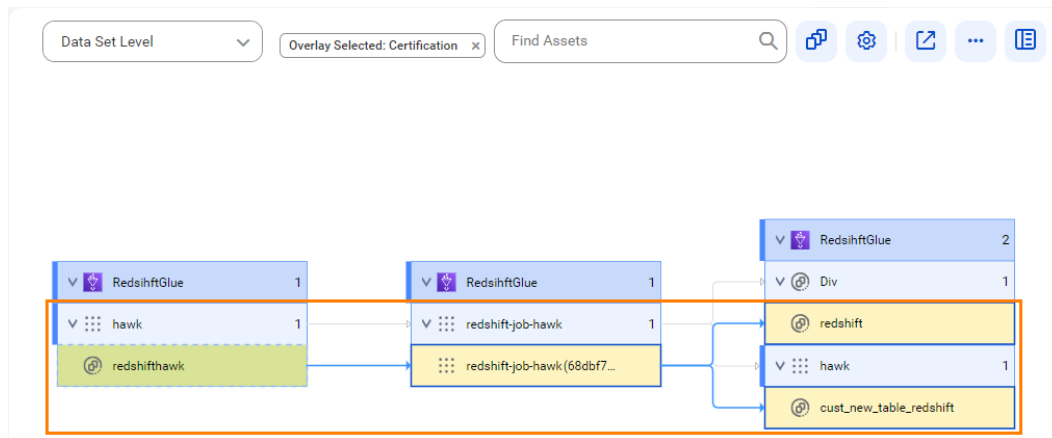
The lineage at the data set level and the data element level shows how technical assets such as files and commands contribute to the selected asset.

Data sets are technical assets that contain sets of data. Examples include files, databases, or temp files that hold the results of calculations. Data elements are objects upstream or downstream of a data set, and are accessible when you expand a data set to the data element level. For example, a table is a data set, and a column in a source object is a data element.

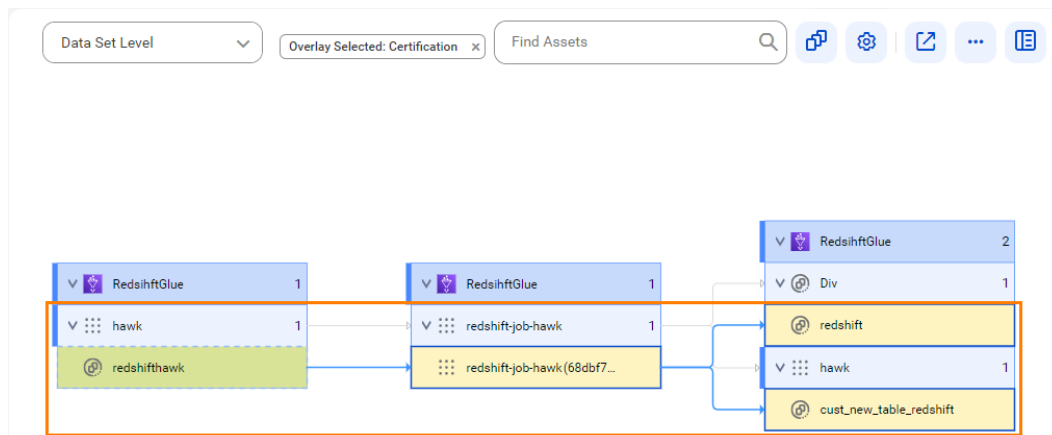
### View lineage at the data set level

The data set level is a view that shows 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 how the redshift-job-hawk job instance uses data from the redshifthawk referenced data set to generate output to the target redshift and cust\_new\_table\_redshift referenced data sets before connection assignment:



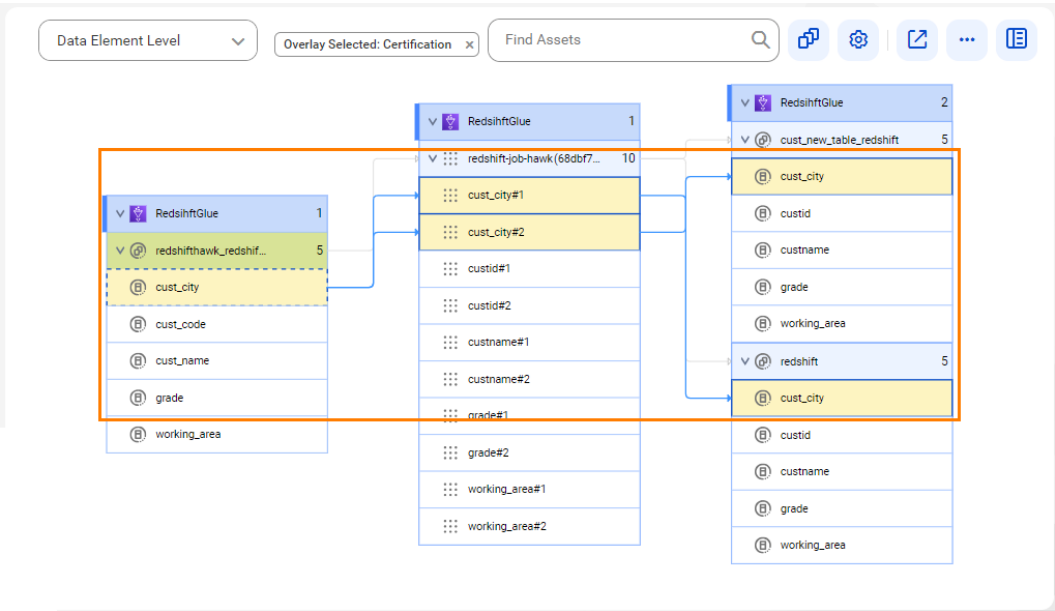
The following image shows how the redshift-job-hawk job instance uses data from the redshifthawk actual data set to generate output to the target redshift and cust\_new\_table\_redshift actual data sets after connection assignment:



### View lineage at the data element level

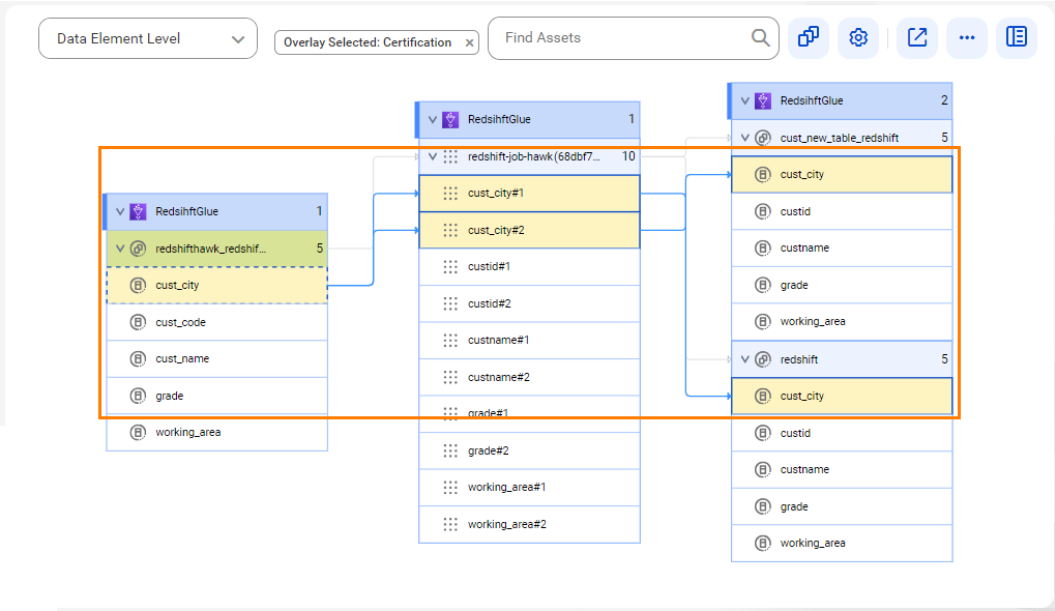
The data element level displays details of the data set level. At the data element level, you can see the input sources for expressions or commands and calculations or transformations on the data. To view data lineage at the data element level, open a technical asset, click the **Lineage** tab, and then verify that the level is set to **Data Element Level**.

The following image shows the lineage where the cust\_city referenced data elements of the cust\_new\_table\_redshift and redshift referenced data sets get processed data before connection assignment:



The source data from the cust\_city referenced data element of the redshifthawk\_redshift referenced data set is processed using the cust\_city#1 and cust\_city#2 calculations of the redshift-job-hawk job instance.

The following image shows the lineage where the cust\_city actual data elements of the cust\_new\_table\_redshift and redshift actual data sets get processed data after connection assignment:



The source data from the cust\_city actual data element of the redshifthawk\_redshift actual data set is processed using the cust\_city#1 and cust\_city#2 calculations of the redshift-job-hawk job instance.