



Informatica® Metadata Command Center
November 2025

Apache HiveQL Script Sources

© Copyright Informatica LLC 2023, 2025

This software and documentation are provided only under a separate license agreement containing restrictions on use and disclosure. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording or otherwise) without prior consent of Informatica LLC.

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

Informatica, Informatica Cloud, Informatica Intelligent Cloud Services, PowerCenter, PowerExchange, and the Informatica logo are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners.

Portions of this software and/or documentation are subject to copyright held by third parties. Required third party notices are included with the product.

The information in this documentation is subject to change without notice. If you find any problems in this documentation, report them to us at infa_documentation@informatica.com.

Informatica products are warranted according to the terms and conditions of the agreements under which they are provided. INFORMATICA PROVIDES THE INFORMATION IN THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2025-11-20

Table of Contents

Preface	4
Chapter 1: Introduction to Apache HiveQL Script catalog sources.....	5
Extraction and view process.	6
About the Apache HiveQL Script catalog source.	6
Extracted metadata.	7
Compatible connectors.	7
Chapter 2: Before you begin.....	8
Verify permissions.	8
Permissions for metadata extraction.	8
Create a connection.	9
Create endpoint catalog sources for connection assignment.	10
Chapter 3: Create catalog sources in Metadata Command Center.....	11
Step 1. Register a catalog source.	11
Step 2. Configure capabilities.	13
Configure metadata extraction.	13
Configure lineage discovery.	15
Step 3. Associate stakeholders and asset groups.	16
Step 4. Run or schedule the job.	18
Step 5. Assign reference catalog source connections to endpoint catalog source objects.	19
Chapter 4: View results in Data Governance and Catalog.....	21
View metadata extraction results.	21
View data lineage.	23
View lineage at the catalog source level.	23
View lineage at the data set level.	23
View lineage at the data element level.	24

Preface

Read *Apache HiveQL Script Sources* to learn how to register and configure Apache HiveQL Script 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 Apache HiveQL Script 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, Apache Hive is a source system from which you can extract metadata through an Apache HiveQL Script 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.

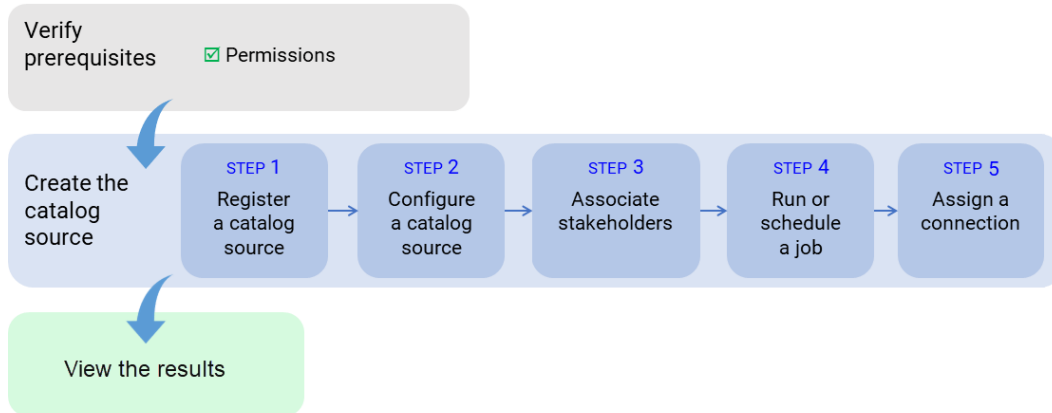
The following table describes the capabilities of the catalog source:

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.

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 Apache HiveQL Script:

1. Register a catalog source. Create a catalog source object, select Apache HiveQL Script, and select the connection.
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 Apache HiveQL Script catalog source

You can use the Apache HiveQL Script catalog source to extract metadata from Apache Hive scripts.

Apache HiveQL Script is a set of Apache HiveQL statements stored in files that you can use to run sequential scripts.

Extracted metadata

You can use the Apache HiveQL Script catalog source to extract metadata from Apache Hive scripts.

Objects extracted

Metadata Command Center extracts the following metadata from Apache Hive scripts:

- Calculation
- Folder
- Script
- Statements

Compatible connectors

Before you configure an Apache HiveQL Script catalog source, you must connect to the Apache Hive source system.

Use the Hive connector to connect to the Apache Hive source system.

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

CHAPTER 2

Before you begin

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

Perform the following tasks:

- Assign the required permissions.
- Configure a connection to the Apache Hive source system in Administrator.
- Save the Apache HiveQL Script files on the runtime environment from which you want to extract metadata.
- Create endpoint catalog sources for connection assignment.

Verify permissions

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

Permissions for metadata extraction

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

Verify that the Cloudera CDP, Amazon EMR, Google Cloud Dataproc, or Azure HDInsight cluster user has read permission on the source.

Grant permissions that allow you to perform the following operations:

- show schemas
- show tables
- show views
- show materialized views

Create a connection

Create a Hive Connector connection object in Administrator with the connection details of the Apache Hive source system.

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 Hive Connector connection type.
5. In the **Hive Connector Properties** section, in the **Runtime Environment** list, select the runtime environment where you want to run the tasks.
6. In the **Connection** section, enter the following connection details:

Connection property	Description
Authentication Type	You can select one of the following authentication types: <ul style="list-style-type: none">- Kerberos. Select Kerberos for a Kerberos cluster.- LDAP. Select LDAP for an LDAP-enabled cluster. Note: LDAP is not applicable to mappings in advanced mode. <ul style="list-style-type: none">- None. Select None for a Hadoop cluster that is not secure or not LDAP-enabled.
JDBC URL *	The JDBC URL to connect to Hive. Specify the following format based on your requirement: <ul style="list-style-type: none">- To view and import tables from a single database, use the following format: jdbc:hive2://<host>:<port>/<database name>- To view and import tables from multiple databases, do not enter the database name. Use the following JDBC URL format: jdbc:hive2://<host>:<port>/ Note: After the port number, enter a slash. <ul style="list-style-type: none">- To access Hive on a Hadoop cluster enabled for TLS, specify the details in the JDBC URL in the following format: jdbc:hive2://<host>:<port>/<database name>;ssl=true;sslTrustStore=<TrustStore_path>;trustStorePassword=<TrustStore_password>, where the truststore path is the directory path of the truststore file that contains the TLS certificate on the agent machine.
JDBC Driver *	The JDBC driver class to connect to Hive.
Username	The user name to connect to Hive in LDAP or None mode.

Connection property	Description
Password	The password to connect to Hive in LDAP or None mode.
Principal Name	The principal name to connect to Hive through Kerberos authentication.
Keytab Location	The path and file name to the Keytab file for Kerberos login.
Configuration Files Path *	<p>The directory that contains the Hadoop configuration files for the client.</p> <p>Copy the site.xml files from the Hadoop cluster and add them to a folder in the Linux box. Specify the path in this field before you use the connection in a mapping to access Hive on a Hadoop cluster:</p> <ul style="list-style-type: none"> - For mappings, you require the core-site.xml, hdfs-site.xml, and hive-site.xml files. - For mappings in advanced mode, you require the core-site.xml, hdfs-site.xml, hive-site.xml, mapred-site.xml, and yarn-site.xml files.

* These fields are mandatory parameters.

7. Click **Test Connection**.
8. Click **Save**.

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 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 Apache HiveQL Script 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

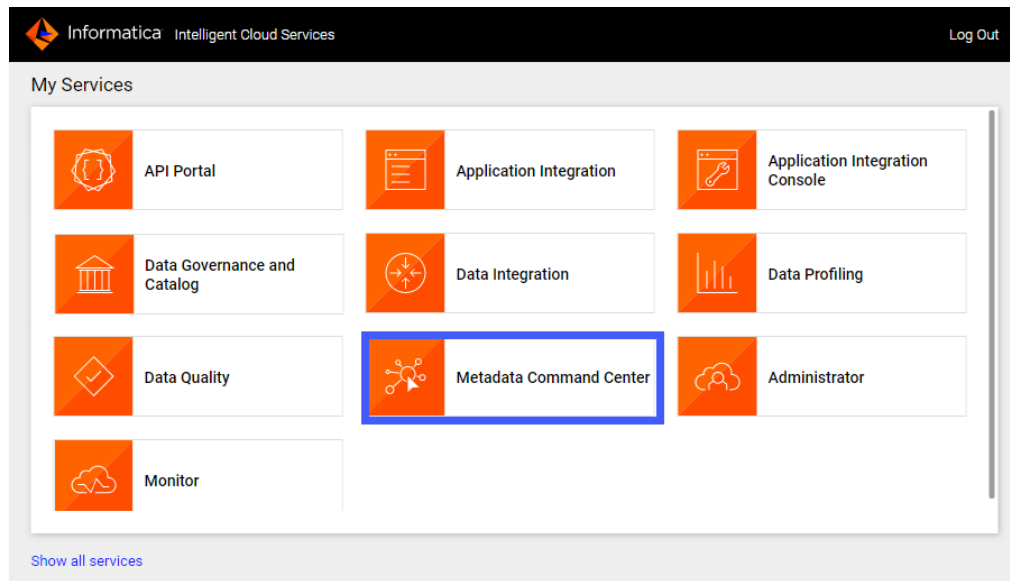
When you register a catalog source, provide general information and connection values.

1. Log in to Informatica Intelligent Cloud Services.

The **My Services** page appears.

2. Click **Metadata Command Center**.

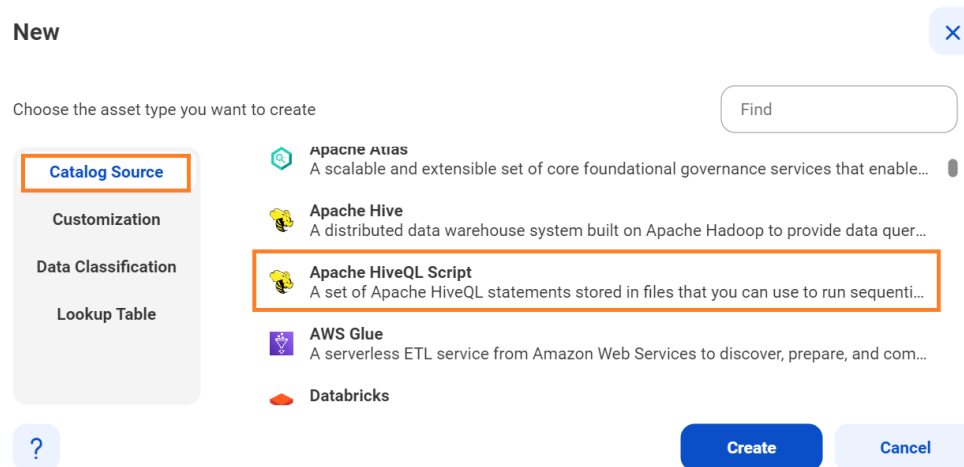
The following image shows the Metadata Command Center box on the **My Services** page:



The Metadata Command Center home page appears.

3. Click **New**.
4. Select **Catalog Source** from the list of asset types.
5. Select Apache HiveQL Script from the list of catalog source types.

The following image shows the Apache HiveQL Script catalog source type:



6. Click **Create**.
The **New Catalog Source** page opens.
7. In the **General Information** section, enter a name and an optional description for the catalog source.
Note: You can rename a catalog source after you create it, but to apply the change to all associated objects you must rerun the metadata extraction job.
After you save the catalog source, you can update the description in Metadata Command Center and Data Governance and Catalog. The update appears only in the service in which you update it.
8. In the **Connection Information** area, select the connection that you created in Administrator.

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

9. Click **Connection Properties** to expand and view the connection properties for the selected connection.
10. Click **Next**.

The **Configuration** page appears.

Step 2. Configure capabilities

When you configure the Apache HiveQL Script 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 Apache HiveQL Script 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 All.
 - c. Enter a value to specify the object location.

Filters are not case-sensitive.

Filters can contain the following wildcards:

- Question mark. Represents a single character.
- Asterisk. Represents multiple characters or empty text.
- Double asterisk. Represents one or more levels of folders. Patterns with a double asterisk are recursive. Recursive filter patterns might take longer to process as the data volume increases.

The following image shows the filter condition options:

Filters

Specify metadata filters: ☐ No ☒ Yes

[Show supported wildcards and examples](#)

Include or exclud...

Select the object ...

Enter a value to specify the object location

+

If the scripts root directory path is /users/opt/input, use the following examples to create filter conditions:

- To include or exclude metadata from the script named script1.sql located in the path /users/opt/input/folder1/, enter: `folder1/script1.sql`
 - To include or exclude metadata from all scripts with SQL extensions from folder2 as well as any subfolders within folder2, enter: `folder1/folder2/**/*.*.sql`
 - To include or exclude metadata from all scripts from folder2 as well as any subfolders within folder2, enter: `folder1/folder2/*`, `folder1/folder2/`, or `folder1/folder2`
4. 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 enter for Catalog Source Configuration Options:

Property	Description
Script Root Directory Path	Path to the HiveQL script root directory on the Secure Agent.
Default Schema	Default schema to process HiveQL scripts.

Property	Description
Default Variable Values	Values for the variables used in the HiveQL scripts .
MetaTables Include Filter	<p>Advanced parameter. Use this parameter if you want Metadata Command Center to read the content from tables or views when processing dynamic HiveQL statements.</p> <p>When you process HiveQL statements, Metadata Command Center does not read the content in tables or views by default.</p> <p>This parameter prompts the database for the required metadata. Verify that the user has SELECT permissions for metatables.</p> <p>Note: Don't use this option to specify filters for tables that you want to include or exclude during the metadata extraction run.</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>Caution: 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.

 - Select **Yes** to view filter options.
 - From the Include/Exclude list, choose to include or exclude catalog sources for lineage discovery based on the filter parameters.
 - From the filter type list, select catalog source type, catalog source name, or asset group.
 - 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.

Step 3. Associate stakeholders and asset groups

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

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

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

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

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

- e. Select one or more users or user groups to assign as stakeholders for the technical assets, and click **OK**.
Only the selected users and user groups belonging to the specified stakeholder role are granted the permissions to technical assets.
- f. To assign users or user groups from another stakeholder role, click **Add** and then repeat the steps.

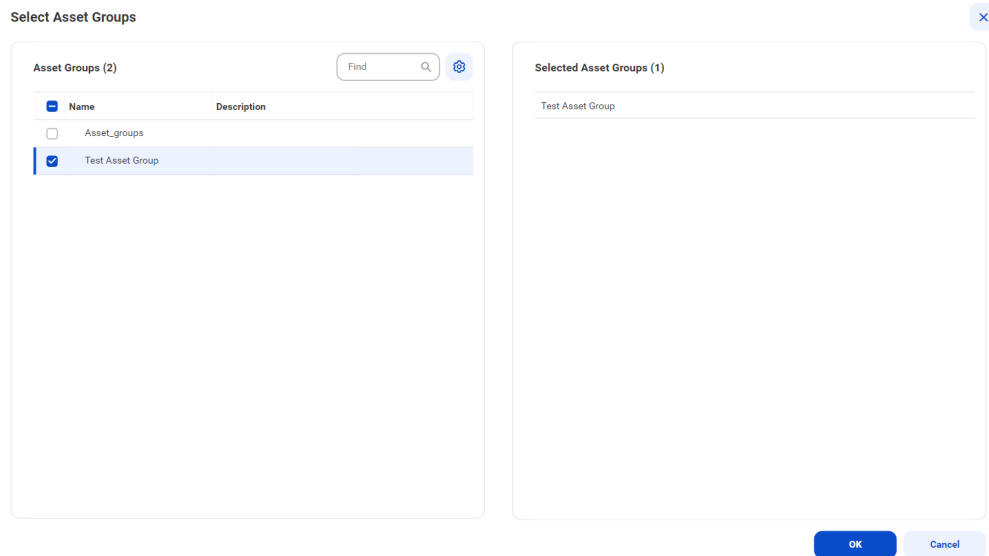
2. To assign asset groups to technical assets extracted from the catalog source, perform the following steps:

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

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

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

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



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

Step 4. Run or schedule the job

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

Note: You can't run multiple jobs simultaneously.

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

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

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

Run the job manually

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

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

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

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

Run the job on a schedule

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

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

For example, this can happen in the following scenarios:

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

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

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

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

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

Monitor job status

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

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

Step 5. 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 a data

warehouse, such as Apache Hive. 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 assign an Apache Hive source system as an endpoint catalog source. The objects must be of the Schema 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.

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

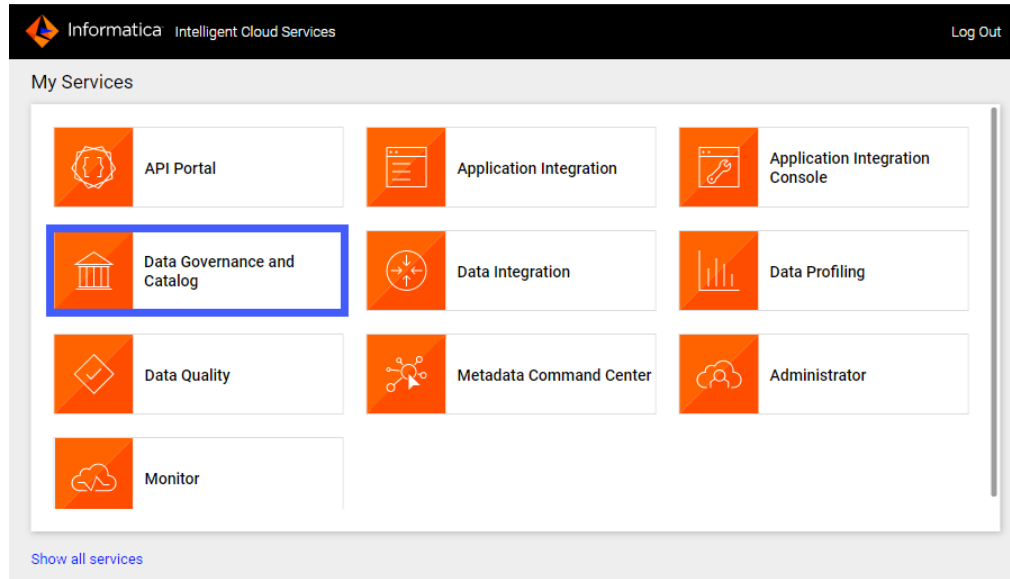
After a job runs in Metadata Command Center, view the results in Data Governance and Catalog. You can view details about source system contents in a hierarchical structure and trace data lineage.

1. Log in to Informatica Intelligent Cloud Services.

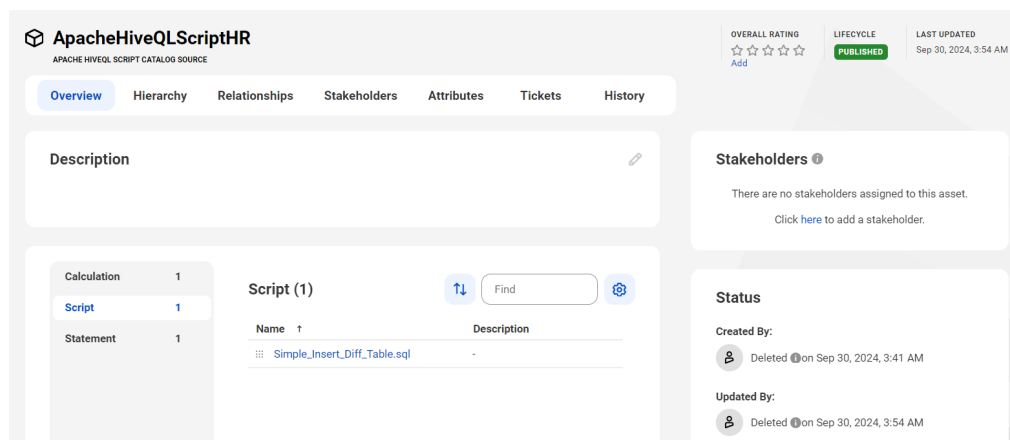
The **My Services** page appears.

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.

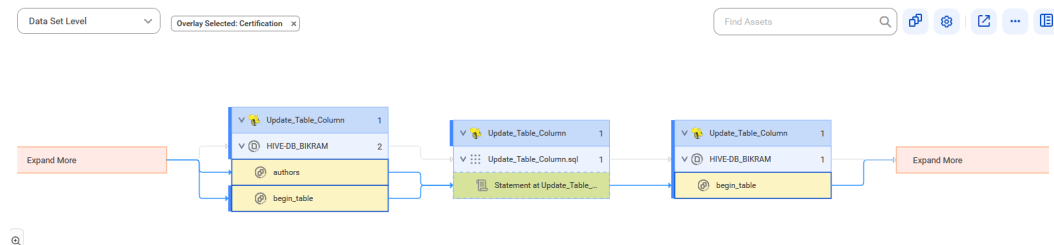
To view data lineage at the catalog source level, open a technical asset, click the **Lineage** tab, and then verify that the level is set to **Catalog Source Level**.

View lineage at the data set level

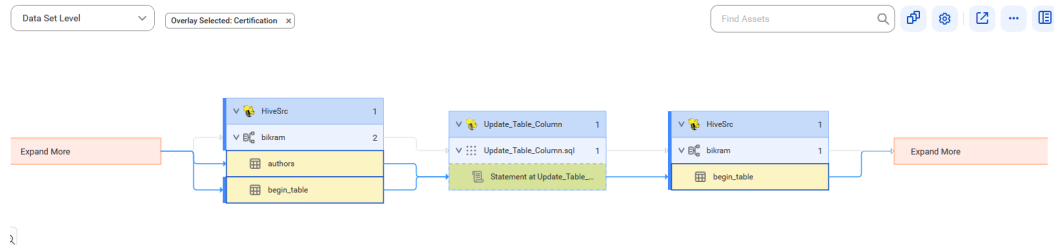
The data set level displays individual sets of data in the data flow.

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

The following image shows data set level lineage where Statement at Update_Table statement catalog source gets data from the authors and begin_table tables of the source system before connection assignment:



The following image shows data set level lineage where Statement at Update_Table statement gets data from the authors and begin_table tables of the source system after connection assignment:



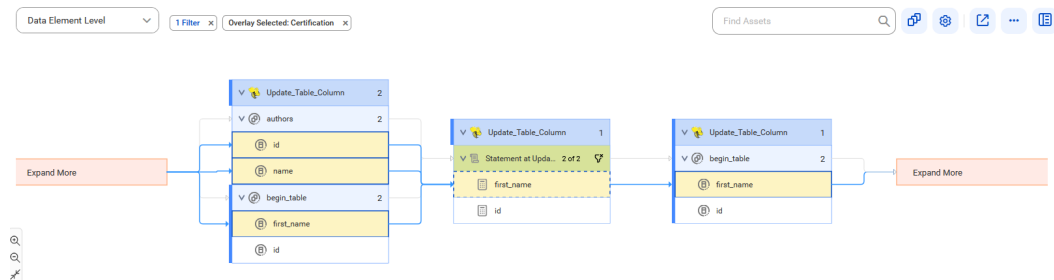
After connection assignment, the referenced object icons change to specific object icons.

View lineage at the data element level

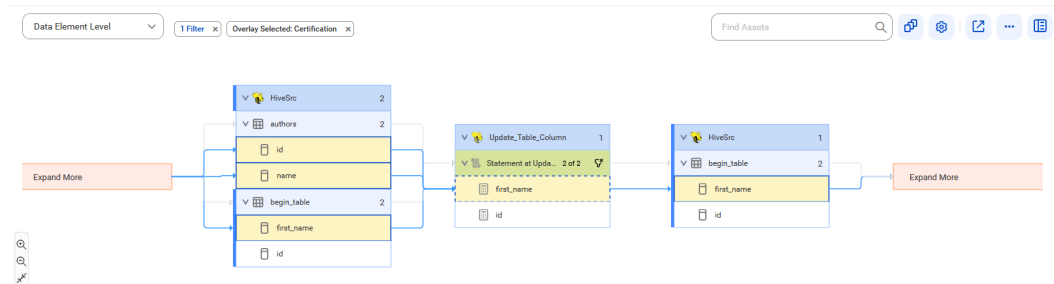
The data element level displays details of the data set level. At the data element level, you can see the input sources for expressions or commands and calculations or transformations on the data.

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 data element level lineage where the first_name calculation of the Statement at Update_Table statement gets data from id, name, and first_name reference data elements of the authors reference data set before connection assignment:



The following image shows data element level lineage where the first_name calculation of the Statement at Update_Table statement gets data from id, name, and first_name data elements of the authors table data set after connection assignment:



After connection assignment, the referenced object icons change to specific object icons.