Cloudera Octopai Data Lineage 1.0.0

Administration

Date published: 2025-10-09
Date modified: 2025-10-20



https://docs.cloudera.com/

Legal Notice

© Cloudera Inc. 2026. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

Users and access.....4
Admin User - Creating users for Cloudera Data Lineage Portal.....4
Admin User - Octopai Client.....5

Admin Console.....12
Admin User - Cloudera Data Lineage Admin Console.....12
Augmented Links.....19
How to manage Inferred Objects.....21

Users and access

Administrators create users for the Cloudera Data Lineage portal and set up Octopai Client to extract metadata from source tools and upload it to the cloud.

Admin User - Creating users for Cloudera Data Lineage Portal

Learn about creating your user for the Cloudera Data Lineage (formerly Octopai) Portal using CSM.

About this task

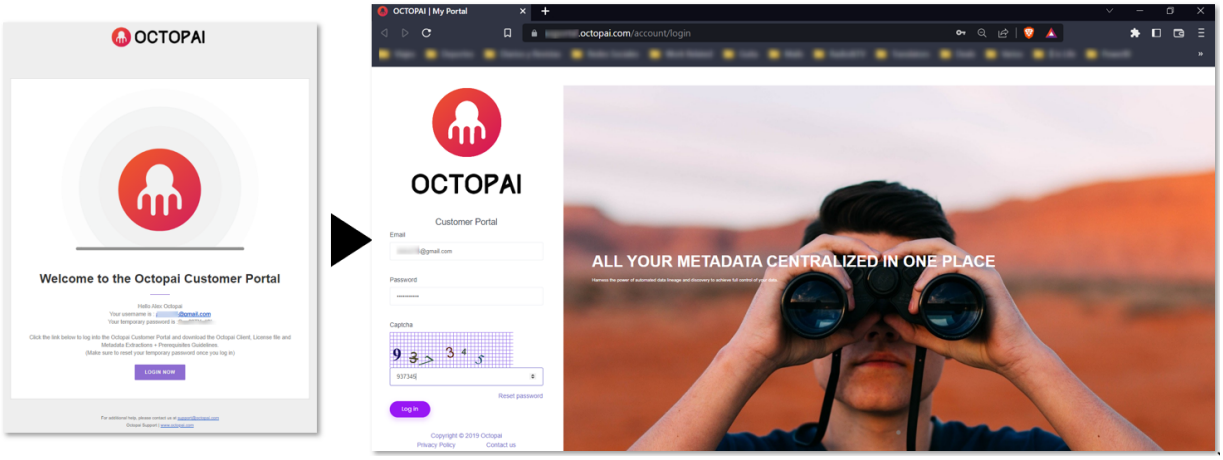
Note: In Cloudera Data Lineage Portal, you will find the Octopai Client, licenses, and files that have been last extracted. The Portal can also be used to exchange large files, always in a zip format.

For the CSM to create your user for the Cloudera Data Lineage Portal, perform the following steps:

Procedure

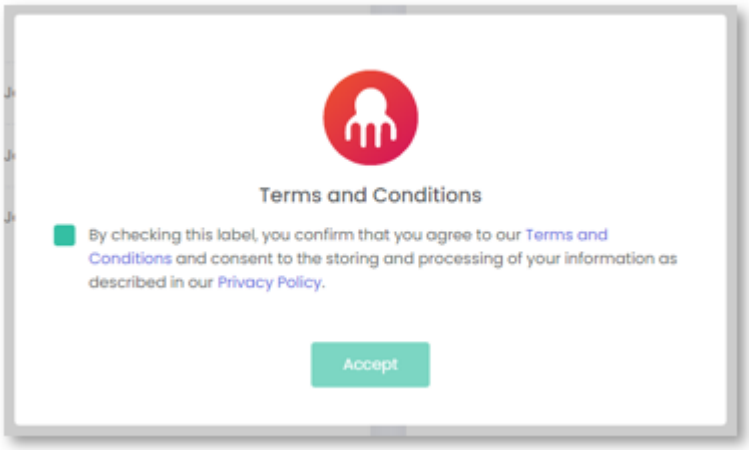
1. Click the Login Now link in the email you received from Cloudera Data Lineage.
2. Log in to the Cloudera Data Lineage Portal.

Figure 1: Cloudera Data Lineage Portal login page



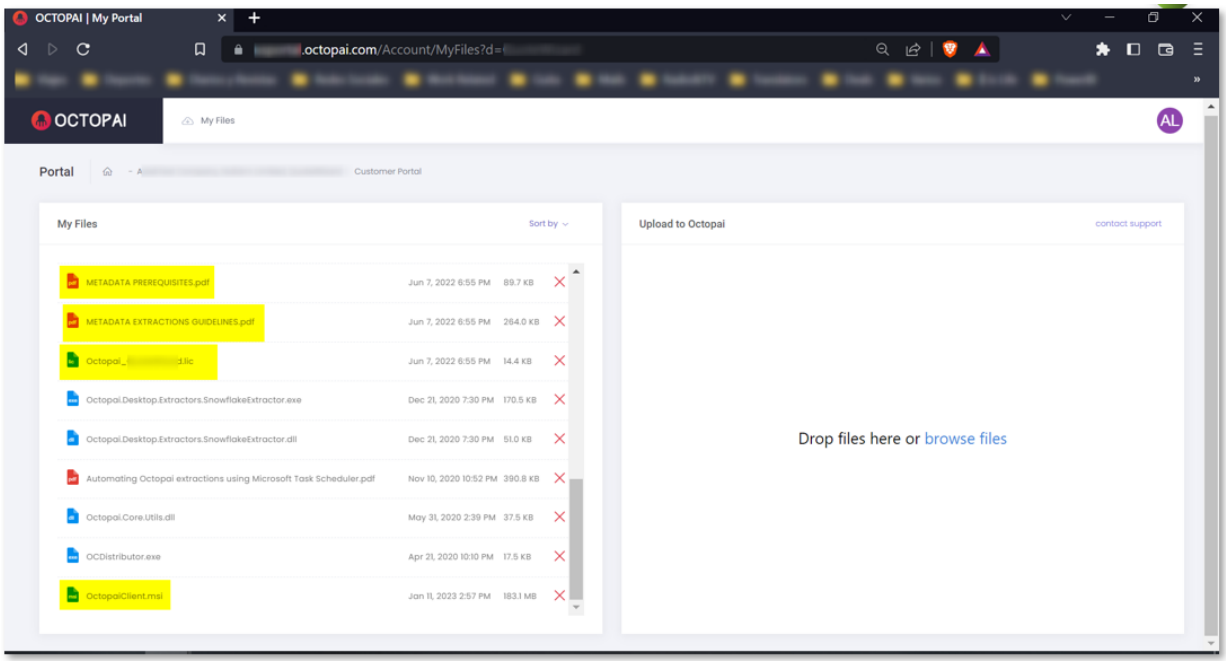
3. Accept the Terms and Conditions.

Figure 2: Terms and conditions modal window



4. Access you local files in the **My Files** section. The following files are your main files:
   - Metadata Prerequisites
   - Metadata Extractions Guidelines
   - Cloudera Data Lineage License - OC (*.lic)
   - Octopai Client (*.msi)

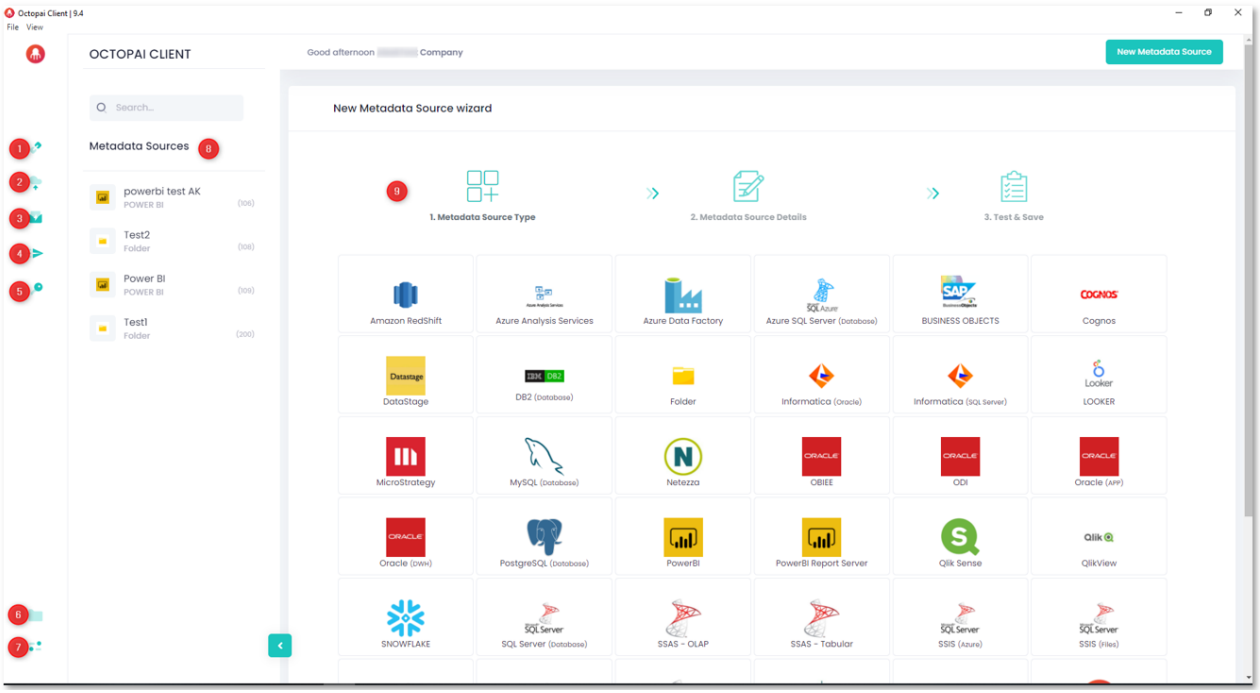
Figure 3: Cloudera Data Lineage Portal My Files section



5. Upload your files to Cloudera Data Lineage. Only files in zip format are accepted.

Admin User - Octopai Client

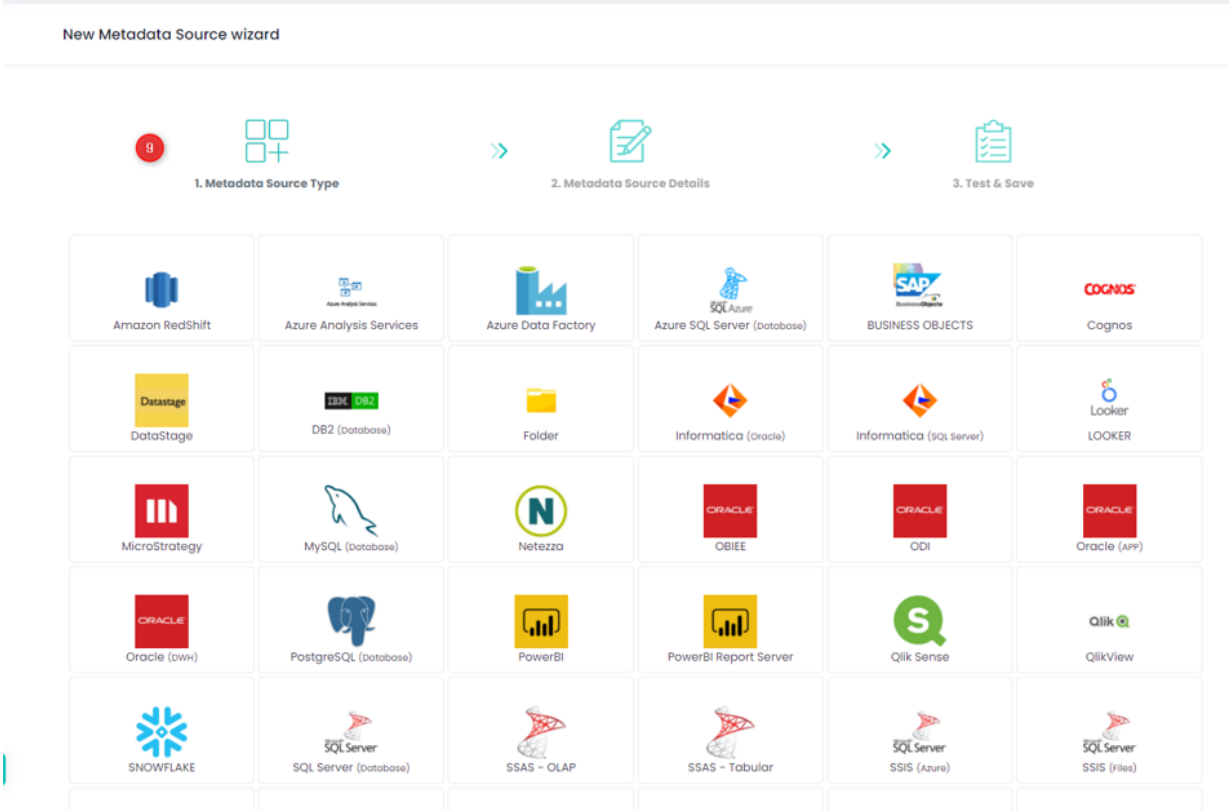
With the Octopai Client, you are able to extract metadata from your tools and upload them to the Cloudera Data Lineage (formerly Octopai) Cloud.



The Octopai Client application consists of the following sections:

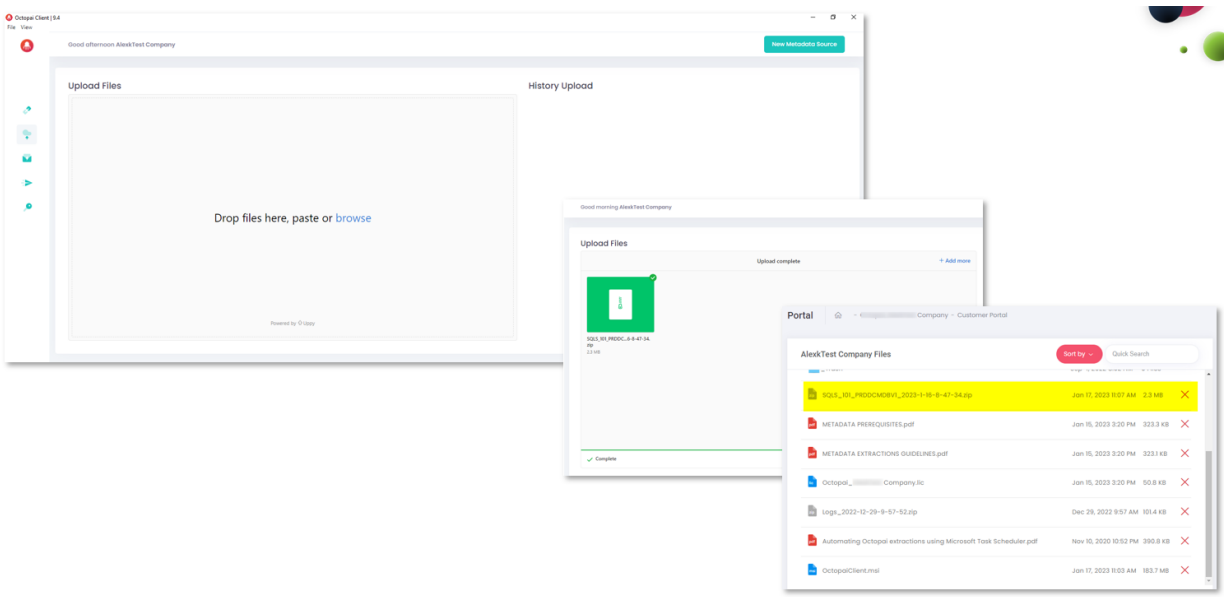
1. **Metadata Sources**

   The main panel contains the out-of-the-box connections in the purchased license.



2. **Upload Files to Cloudera Data Lineage**

   Upload files to Cloudera Data Lineage. Only files in zip format are accepted. When the upload is completed, the files are displayed on your Portal.

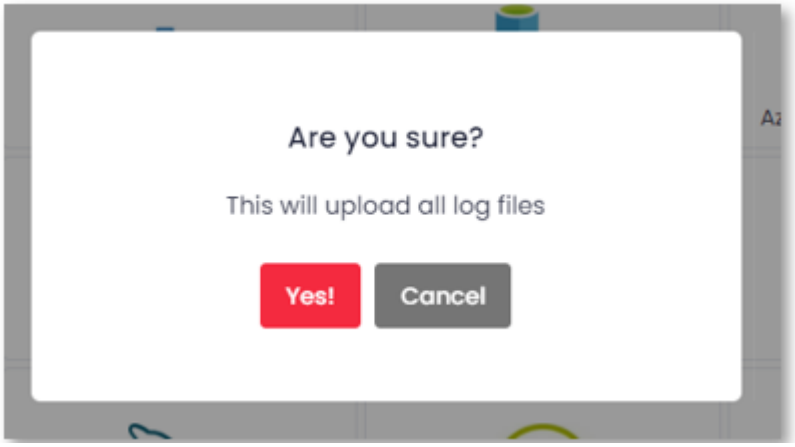


3. **Contact Support**

   If you need assistance, contact Cloudera Support.

4. **Send Logs to Cloudera**

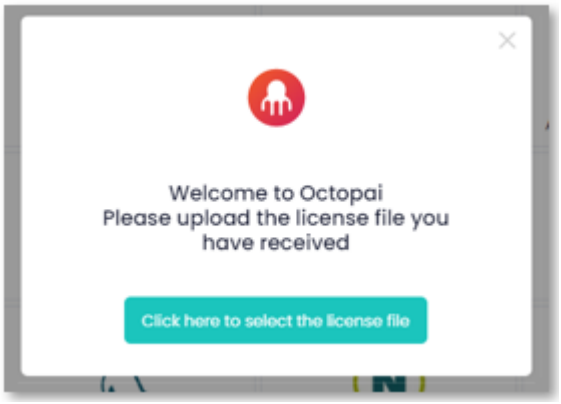
   Logs are helpful to solve issues. Click the icon to send the last logs to Cloudera to be analyzed by the Support team.



5. **Update License**

   After opening the Octopai Client for the first time, or after a Version Upgrade, a window to install the license pops up.

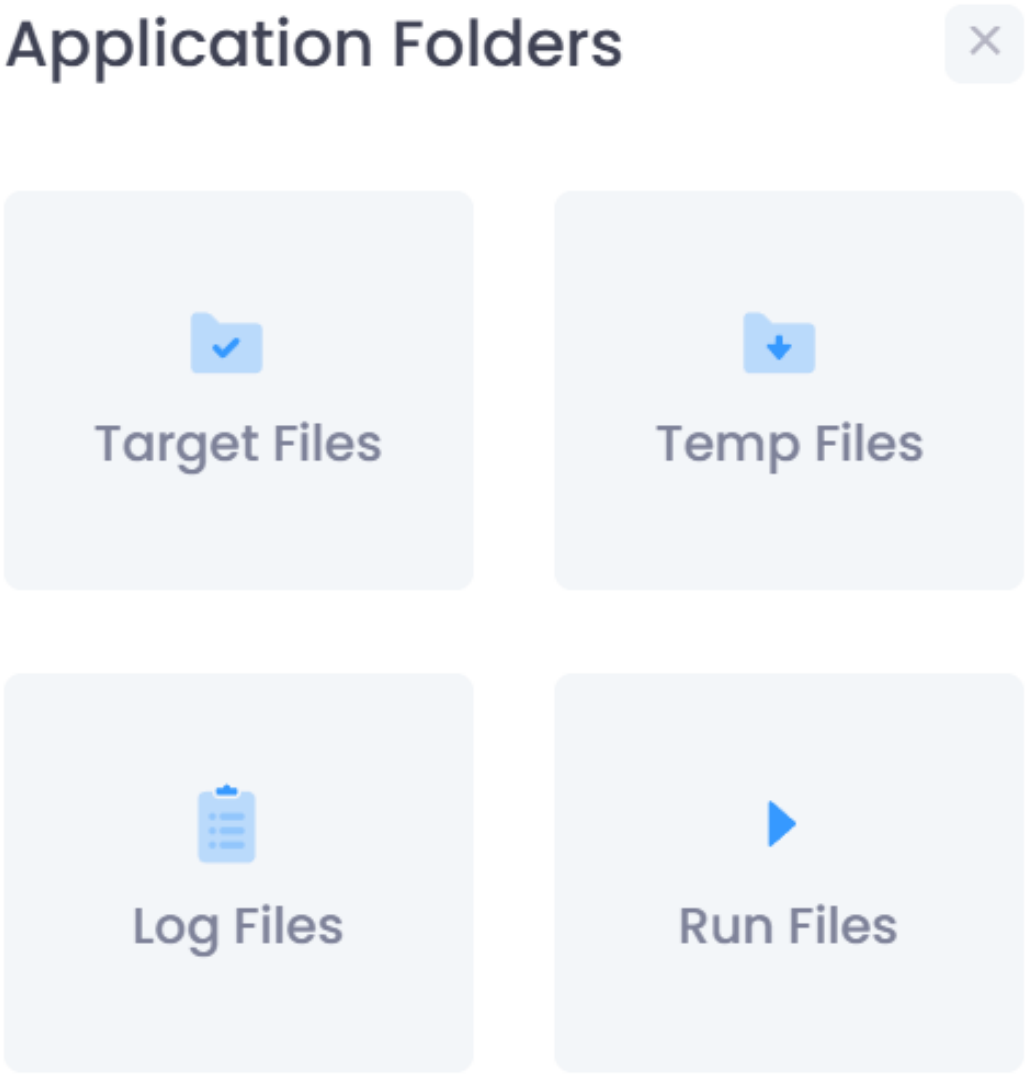
   Click Click to select the license file to browse for the .lic file and click Open to install the license.



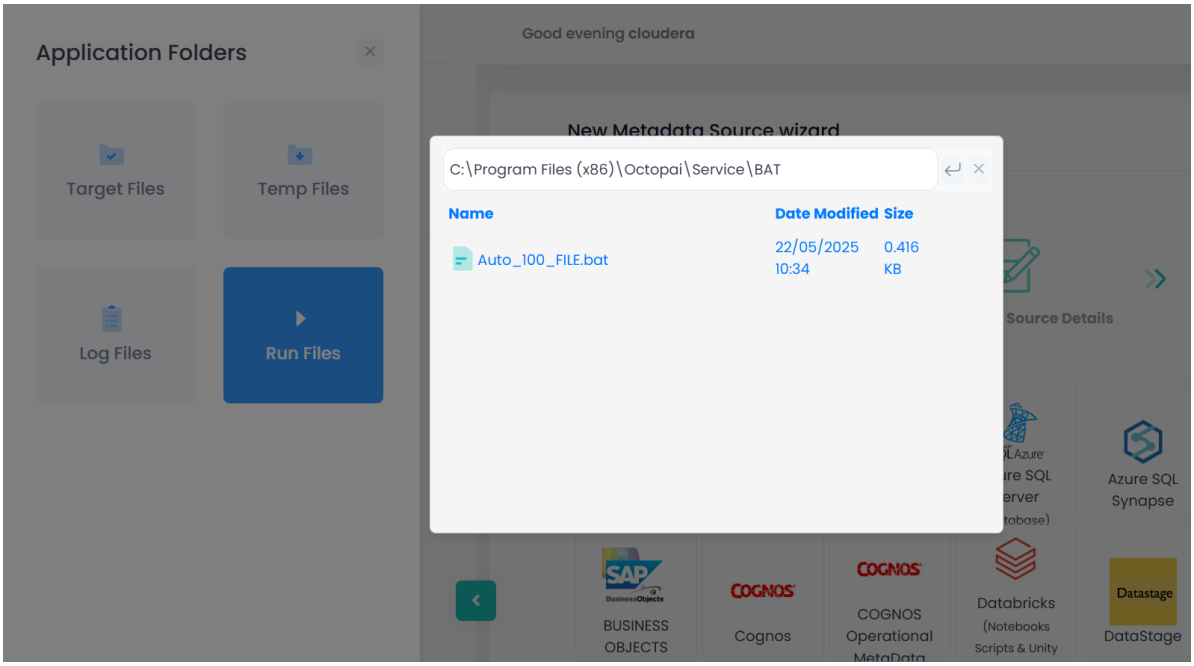
6. **Application Folders**

   The Cloudera Data Lineage application folders play a crucial role in managing metadata extraction processes and ensuring the quality and integrity of metadata. By providing a structured storage environment, these folders enable users to verify metadata accuracy and address any potential issues. You can find the following folders in Octopai Client:

   - Target Files – The Target Files folder serves as a repository for the zip files uploaded to Cloudera Data Lineage.
   - Temp Files – During the metadata extraction process, Cloudera Data Lineage creates temporary files to facilitate data processing and transformation.
   - Log Files – Cloudera Data Lineage generates log files during the metadata extraction process to capture valuable information about the extraction activities.
   - Run Files – To automate the metadata extraction process, Cloudera Data Lineage provides Run (bat/sh) files that can be used in conjunction with Microsoft Task Scheduler. For instructions on using Microsoft Task Scheduler, see Automating Cloudera Data Lineage metadata extractions with Microsoft Task Scheduler.



a. Click Run Files.
b. Select the respective Batch or shell script file, depending on your operating system, to download to your computer for automation activity. Point your scheduler of choice.



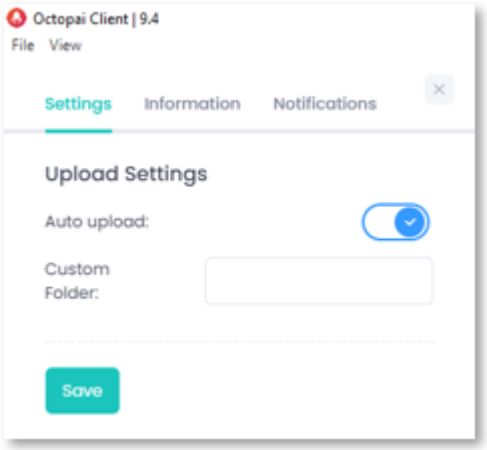
7. **Information and Settings**

   You can see the following tabs in the Information and Settings section:

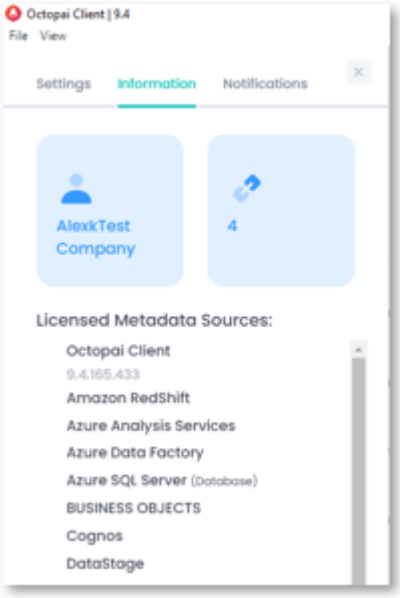
   - The **Settings** tab defines if the extractions are automatically uploaded to the Cloudera Data Lineage Cloud or are done manually.

     Set the **Auto Upload** to ON to enable the extractions to be automatically transferred to the Cloudera Data Lineage Cloud.

     Set the **Auto Upload** to OFF to review the files before they are being transferred to the Cloudera Data Lineage Cloud.

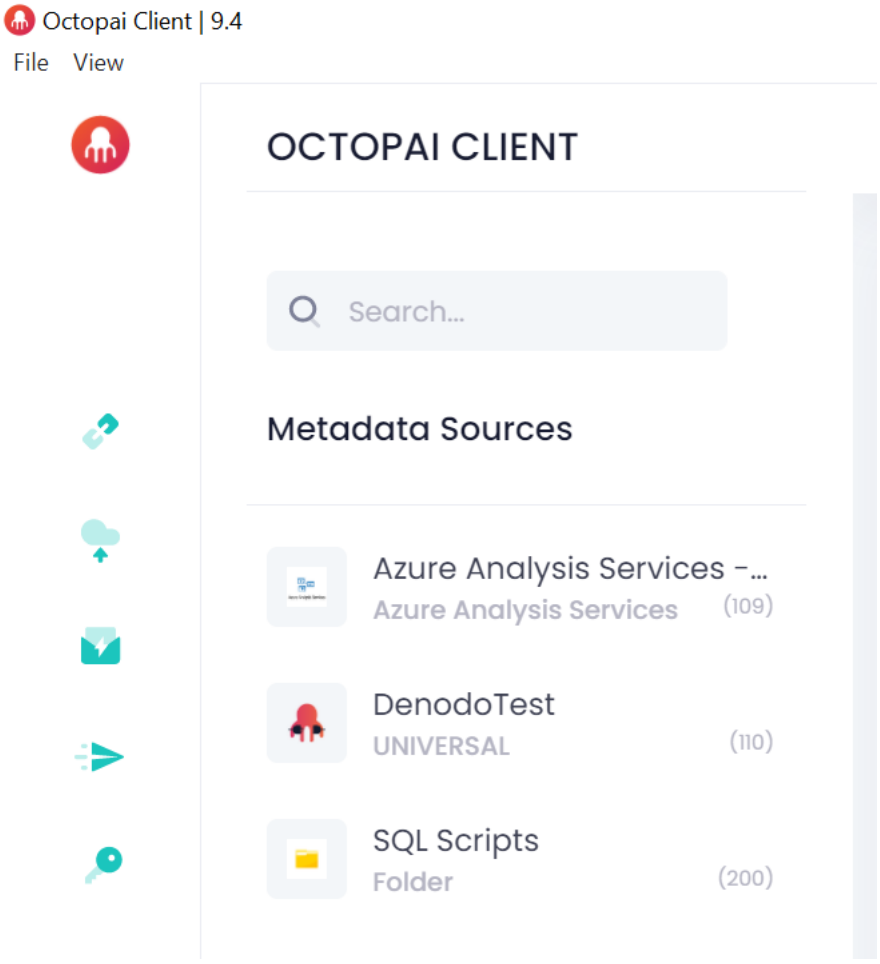


   - The **Information** tab describes the Octopai Client version installed and the list of tools in the purchased license.



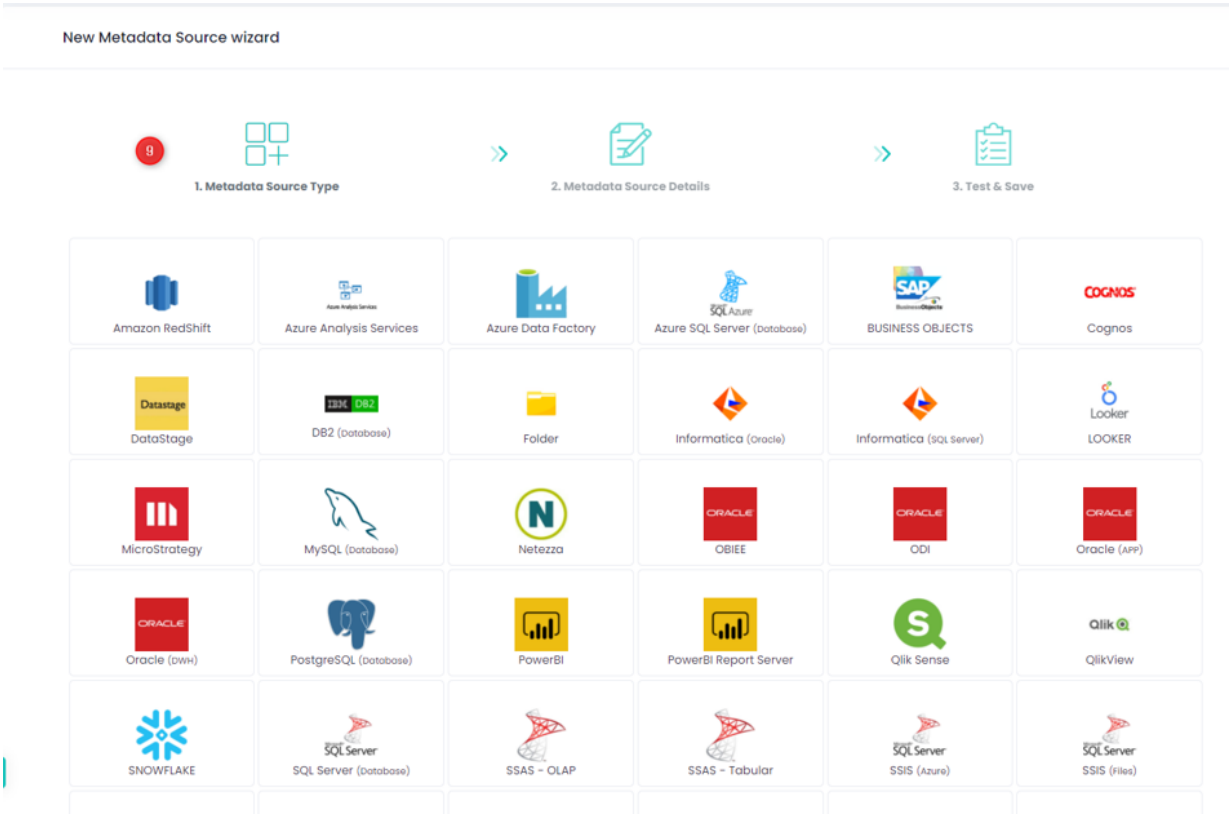
8. **List of Connectors**

   If you create a new connector, it is shown on the **Metadata Sources** list or List of Connectors. The new source set gets a Connector Number, a common denomination between you and Cloudera Data Lineage Support.



9. **Tools List**

   All the tools purchased to create connectors to extract metadata are displayed in the main pane.



Admin Console

The Admin Console is the central place where administrators configure platform settings, create augmented lineage links, and manage inferred objects.

Admin User - Cloudera Data Lineage Admin Console

Admin Console is the central location where administrators can configure settings and manage the platform.

You can access the Admin Console by clicking on the avatar icon. You can leverage Admin Console features to streamline user management, configure system settings, manage metadata, and gain valuable insights.

Figure 4: Access Admin Console

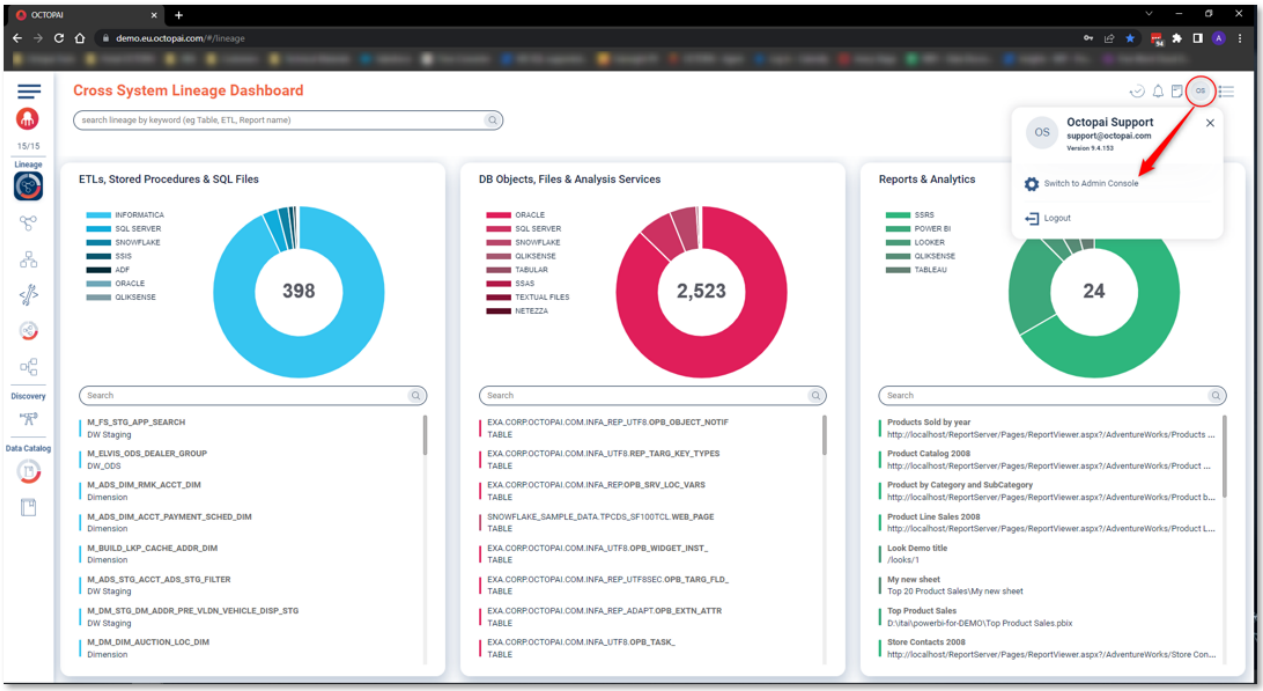
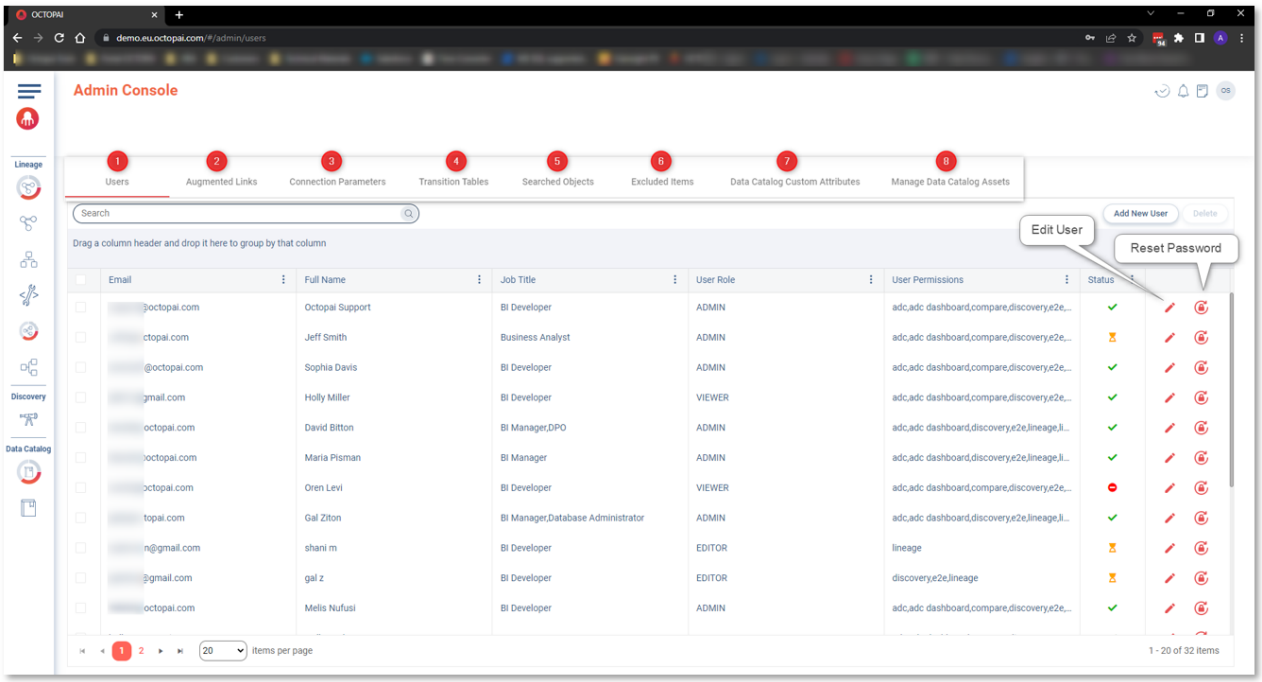


Figure 5: Admin Console sections



Admin Console consist of the following sections:

1. **User Management**

   You can create users manually by performing the following steps:

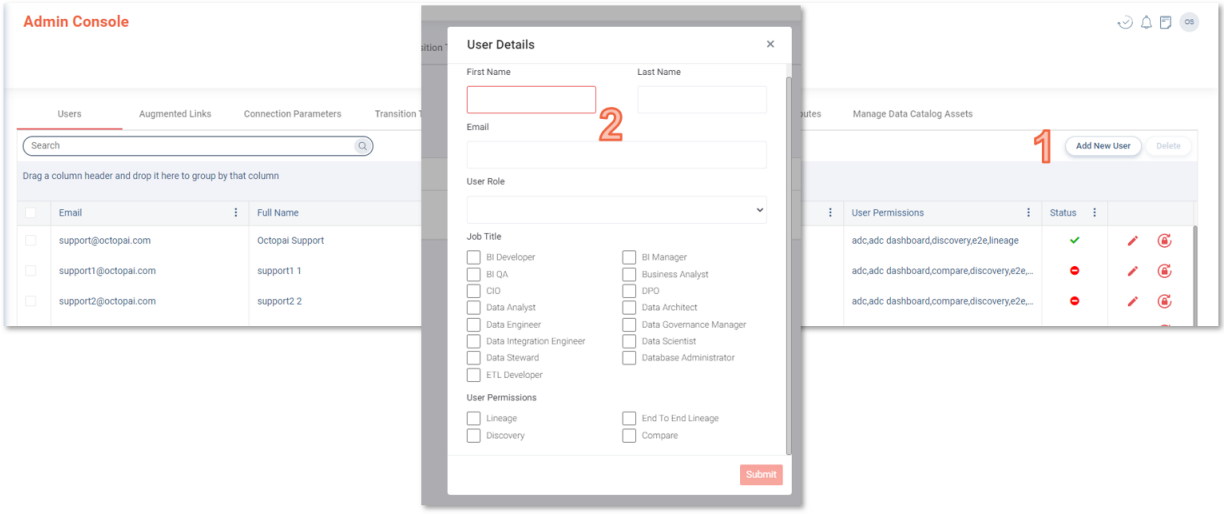
   a. Click Add a New User.

   b. Fill in the fields in the **User Details** model window. The following fields are mandatory:

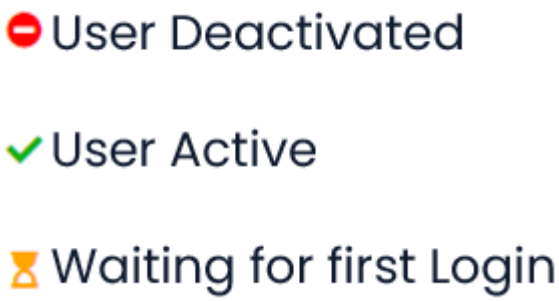
   - **First Name**
   - **Last Name**

- **Email**
- **User Role** – Can be Admin, Editor, or Viewer.
- **Job Title** – You can choose only one.

c. Click Submit.



Users can have the following statuses:



The new user will receive an automatic notification email.

2. Augmented Links

Add Database metadata from your systems into Cloudera Data Lineage (formerly Octopai) to get the full picture, completing the cross system lineage.

Figure 6: Augmented Links tab in Admin Console

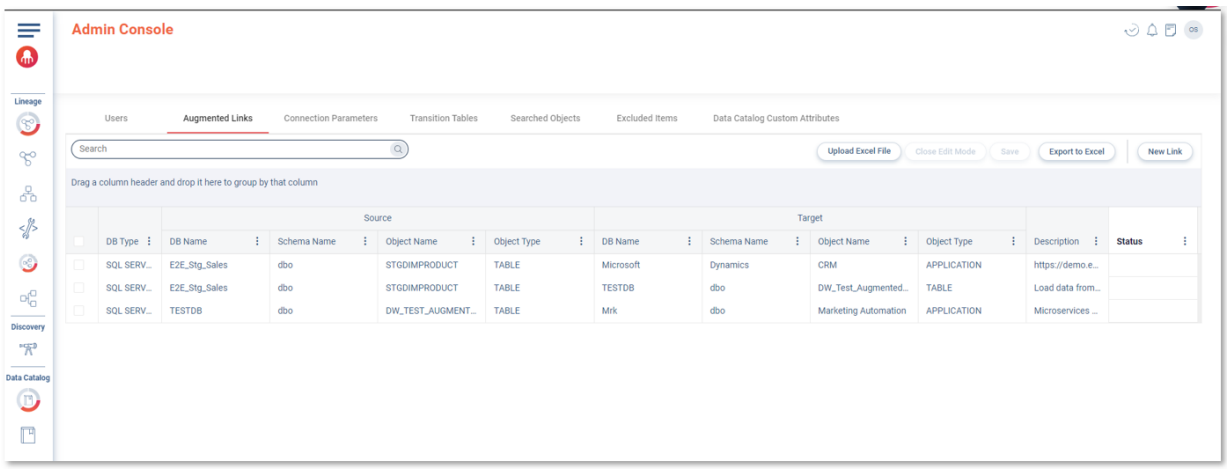
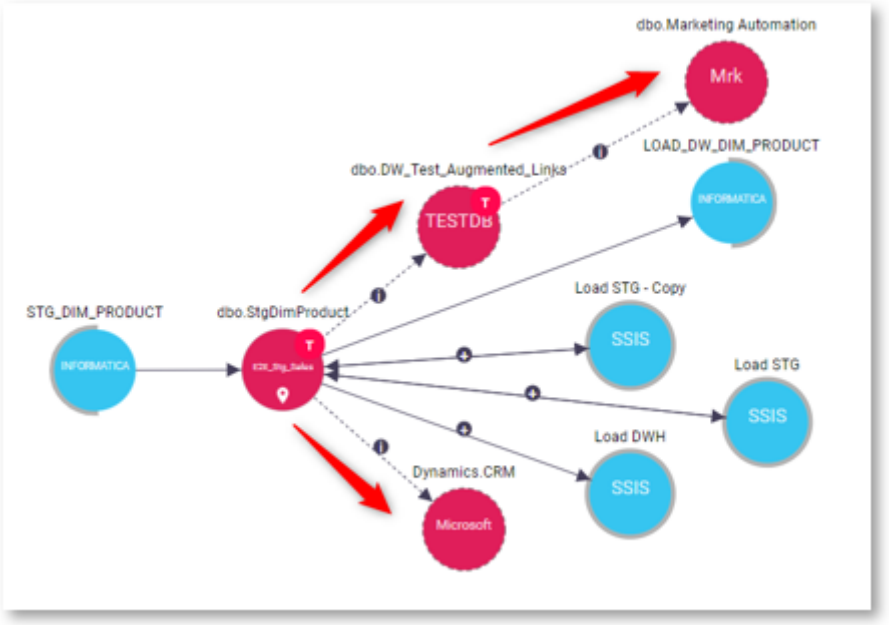


Figure 7: Cross system lineage example



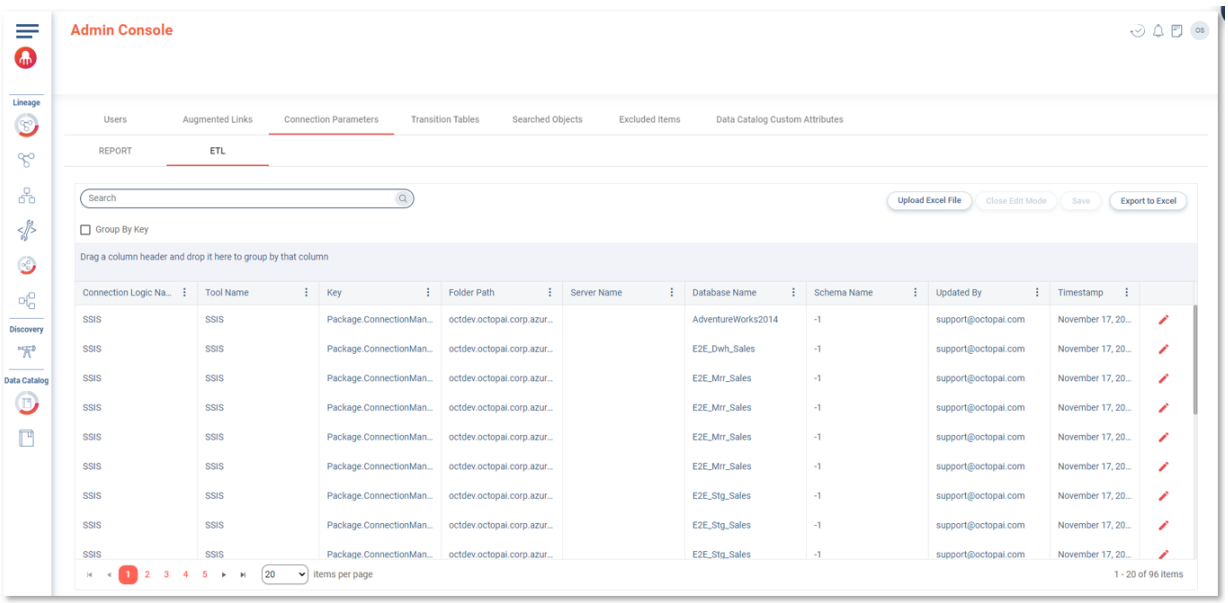
For more information about this capability, see Augmented Links.

3. **Connection Parameters**

   Cloudera Data Lineage helps address broken lineages by providing tools and functionalities to analyze and repair them. By leveraging metadata relationships and lineage information, Cloudera Data Lineage identifies and resolves gaps in data lineages. It assists users in tracing the missing connections, identifying the causes of lineage breaks, and restoring the lineage flow.

   Through a combination of automated processes, manual interventions, and intelligent algorithms, Cloudera Data Lineage helps reconstruct accurate data lineages. This enables data professionals to regain visibility into data flow, understand dependencies, and ensure data integrity throughout their systems.

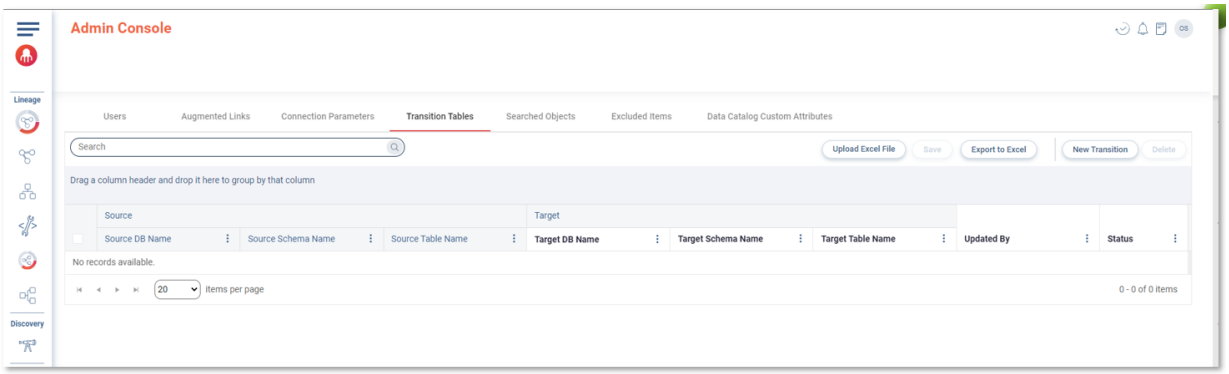
Figure 8: Connection Parameters tab in Admin Console



4. **Transition Tables**

   Cloudera Data Lineage facilitates the use of the `EXCHANGE PARTITION` command to swap tables and implement bulk loads for partitioned tables. It also establishes connections between two tables, recognizing ETL processes that write to the source table and reflecting the same connection in the **Transition Table** tab.

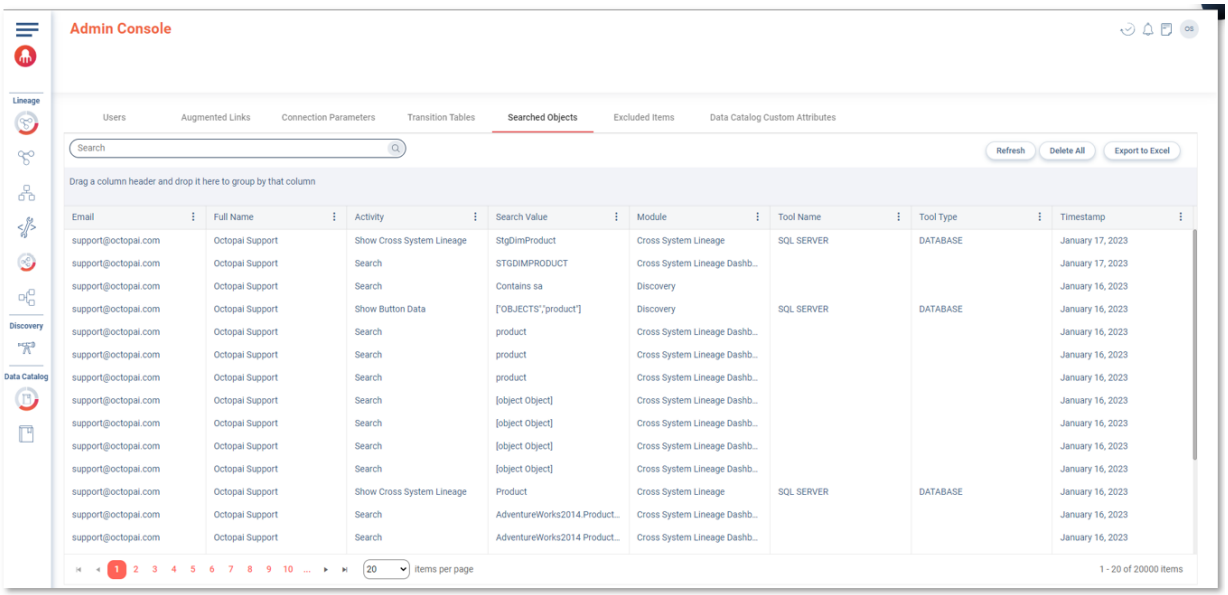
   Figure 9: Transition Tables tab in Admin Console



5. **Searched Objects**

   Cloudera Data Lineage collects user search statistics for analysis and insights.

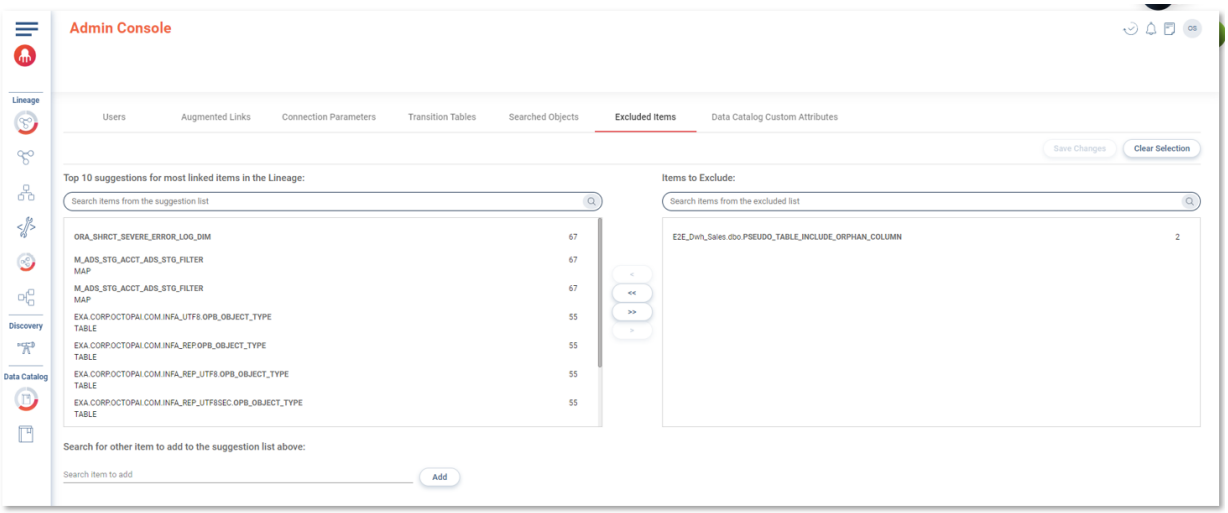
   Figure 10: Searched Objects tab in Admin Console



6. **Excluded Items**

Cloudera Data Lineage allows for excluding specific data objects from lineages. Objects such as logs and synonyms can be excluded from lineage display, simplifying the visualization and providing a cleaner representation. The aim is to streamline the lineage and enhance its visual clarity.

Figure 11: Excluded Items tab in Admin Console



7. **Knowledge Hub Custom Attributes**

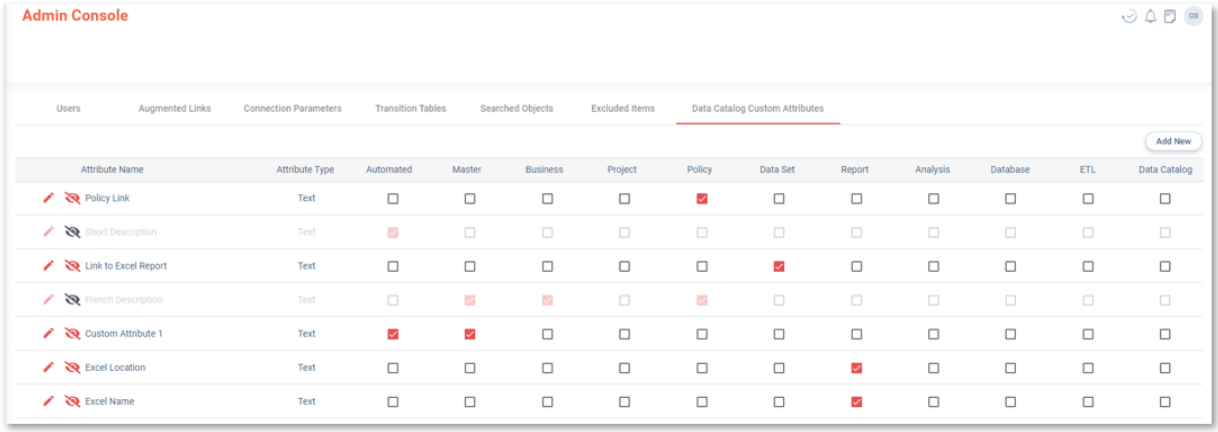
Administrators can create and manage additional attributes in the Knowledge Hub Custom Attributes tab. Administrators can perform the following actions:

a. Create a new attribute.
b. Select the applicable asset types by checking the relevant checkboxes. Assets generated through automation are grouped in the Automated column.
c. Edit the attribute name.
d. Sort the attributes using a simple drag-and-drop interface.
e. Hide attributes.

Note: Data already entered for hidden attributes will be preserved, but the attribute will no longer be displayed for any asset type.

Users can access and view the details of the additional attributes in the **Overview** tab of the **Catalog**, within the **Properties** section.

Figure 12: Data Catalog Custom Attributes tab in Admin Console



Important: Custom attributes are limited to 40 per Asset. Each custom attribute definition can contain a maximum of 500 characters.

8. **Manage Knowledge Hub Assets**

Administrators can update assets in bulk by importing the same template spreadsheet that was exported from the **Total Assets** section. This functionality allows for efficient and streamlined updates to multiple assets simultaneously.

Figure 13: Total Assets section

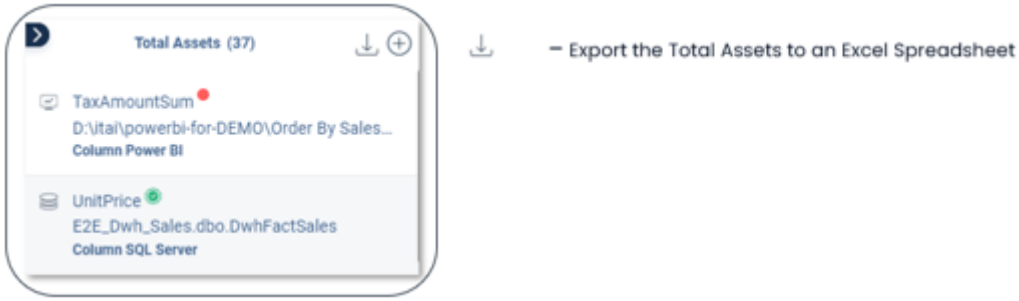
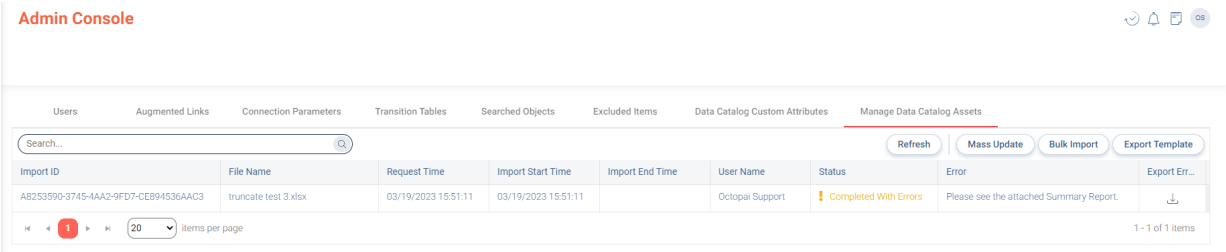


Figure 14: Manage Data Catalog Assets tab in Admin Console

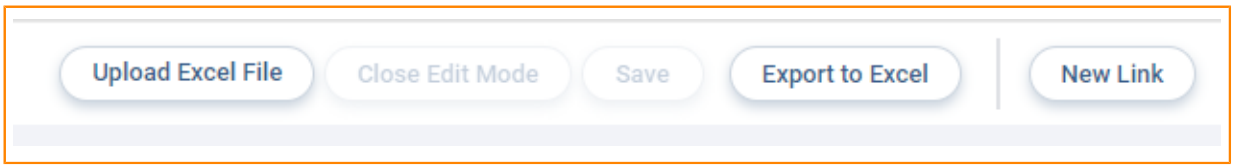


Augmented Links

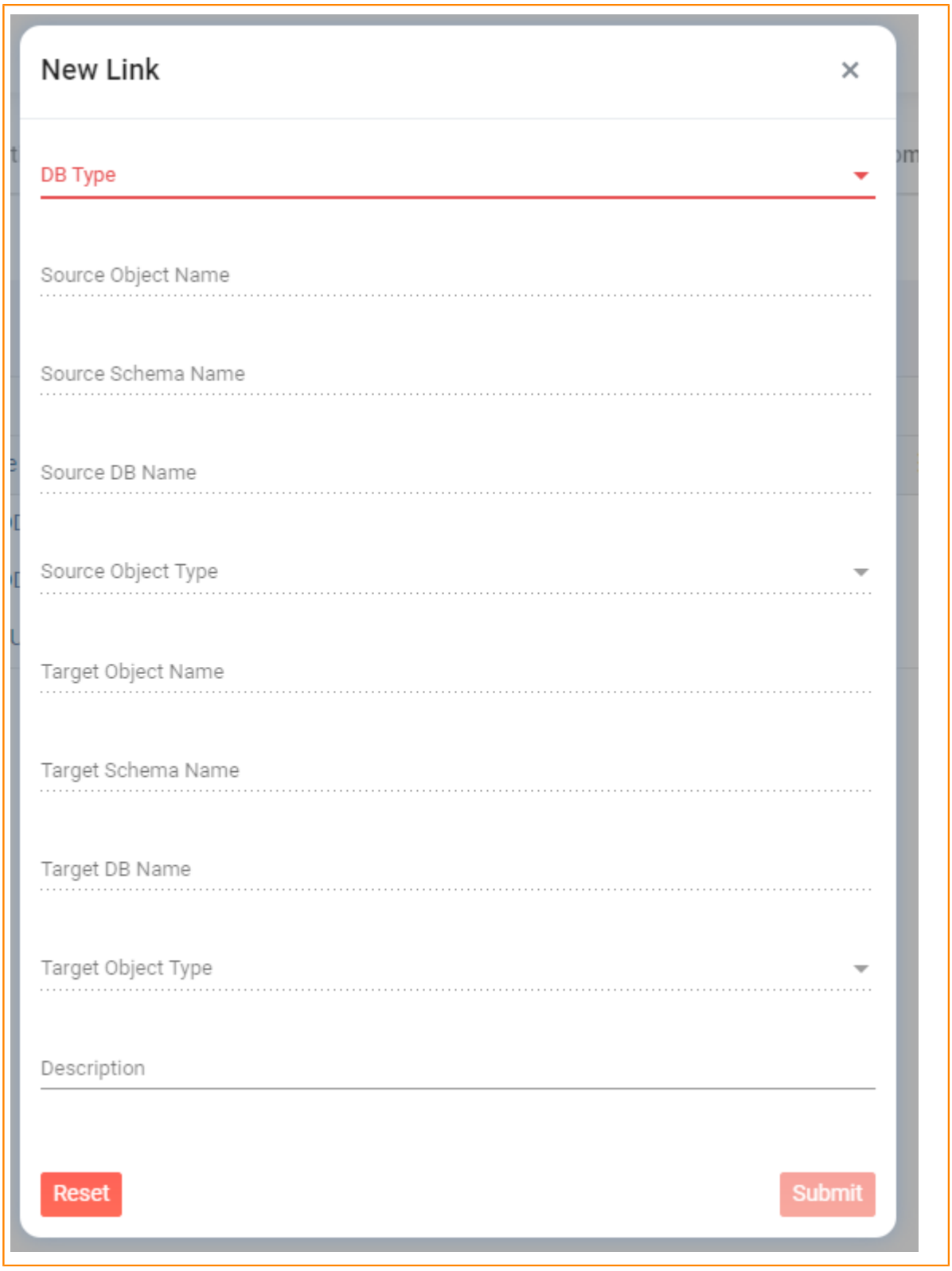
Create manual or bulk augmented lineage links in the Admin Console to connect database objects in Cross System Lineage.

How to Create an Augmented Link

- Go to Admin Console --> Augmented Links



- **Manual Creation**
  - Click on "New Link"



- Fill in the Source To Target information
- Click on Submit

- **Bulk Creation**
  - Click on "Export to Excel" (Template spreadsheet)
  - Fill in the Source To Target **with a limit of 2000 rows**

- Upload the file

Important: **IMPORTANT** : Augmented Links will work only for Database Objects and will be displayed as red bubbles in Cross System Lineage

How to manage Inferred Objects

An Inferred Object (INF) is an object that Cloudera Data Lineage detected by reference — for example, from an SSIS data flow or a Power BI report, but could not match to a specific item in your catalog. Either the object was not extracted, or key details such as the Database Name or Schema Name are missing from the reference. The object appears in the lineage graph flagged as "INF" until the reference can be resolved.

About this task

When viewing lineage in Cloudera Data Lineage, you might come across Data Objects that are displayed as Inferred Objects. This happens when Cloudera Data Lineage lacks sufficient information to resolve the reference — typically because the associated Database Name and Schema Name were not captured for that connection, or because the object itself was not extracted. Use the following diagnostic flow to identify the cause and apply the correct fix.

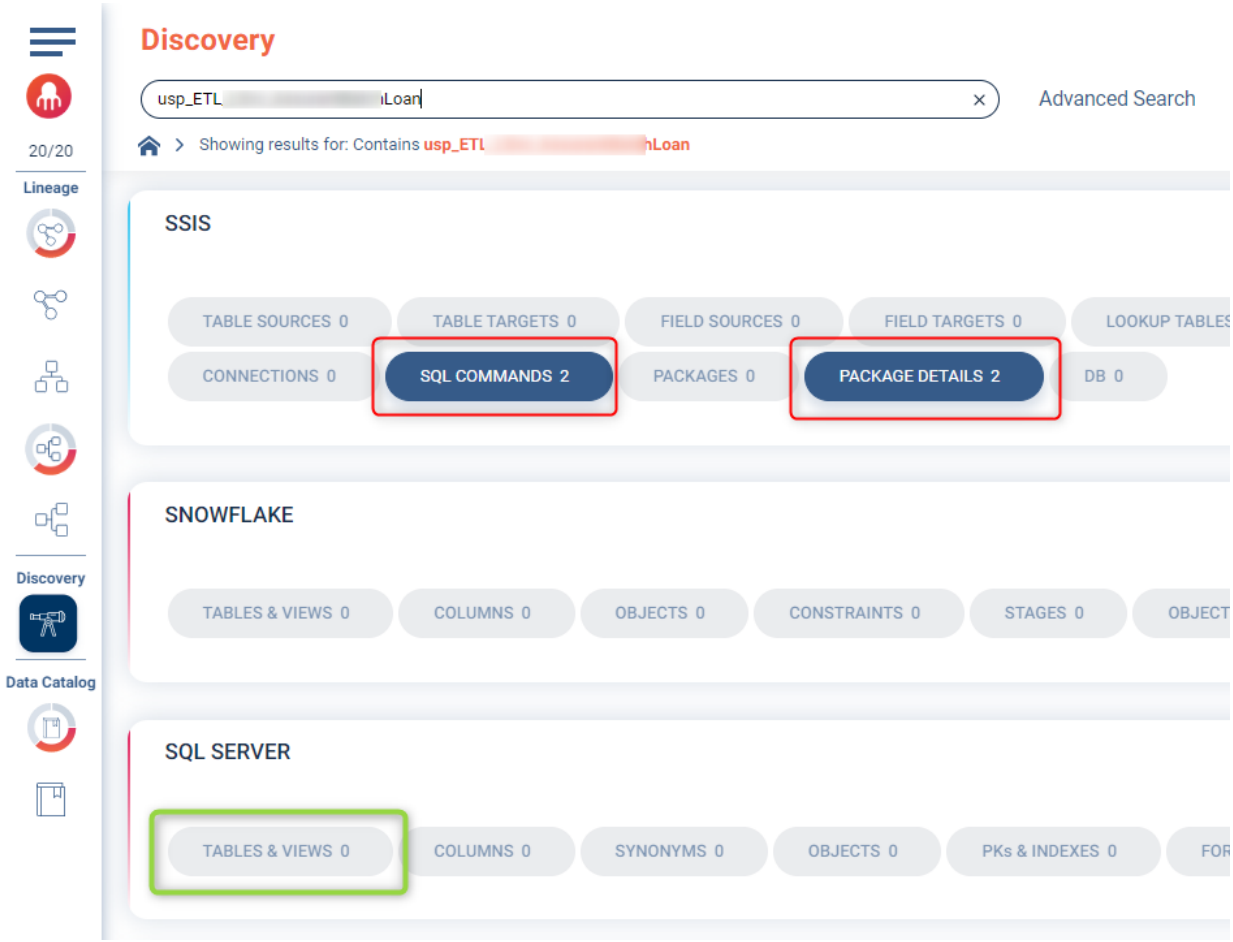
Procedure

1. From the lineage graph, click the Information i icon on the object's radial menu, copy the object name from the **Properties** panel, and paste it into the Cloudera Data Lineage **Discovery** space.

Figure 15: Inferred Object Properties



Figure 16: Discovery space



- **Found, but still flagged as INF** — the object was extracted, but Cloudera Data Lineage cannot resolve the reference back to it, typically because the Database Name or Schema Name were not captured for that connection. Proceed to the next step to add the missing connection parameters.
- **Not found** — the object either was not in scope for extraction, or the source was extracted but this specific object did not come through. Confirm with the source owner whether the object should have been in scope. If yes, check the permissions of the user or service account Cloudera Data Lineage uses to extract metadata from that source, and fix as needed. If the object is intentionally out of scope, this is an expected behavior and no action is required.

2. Fix the unresolved reference in Admin Console Connection Parameters by adding the missing Database Name and Schema Name.

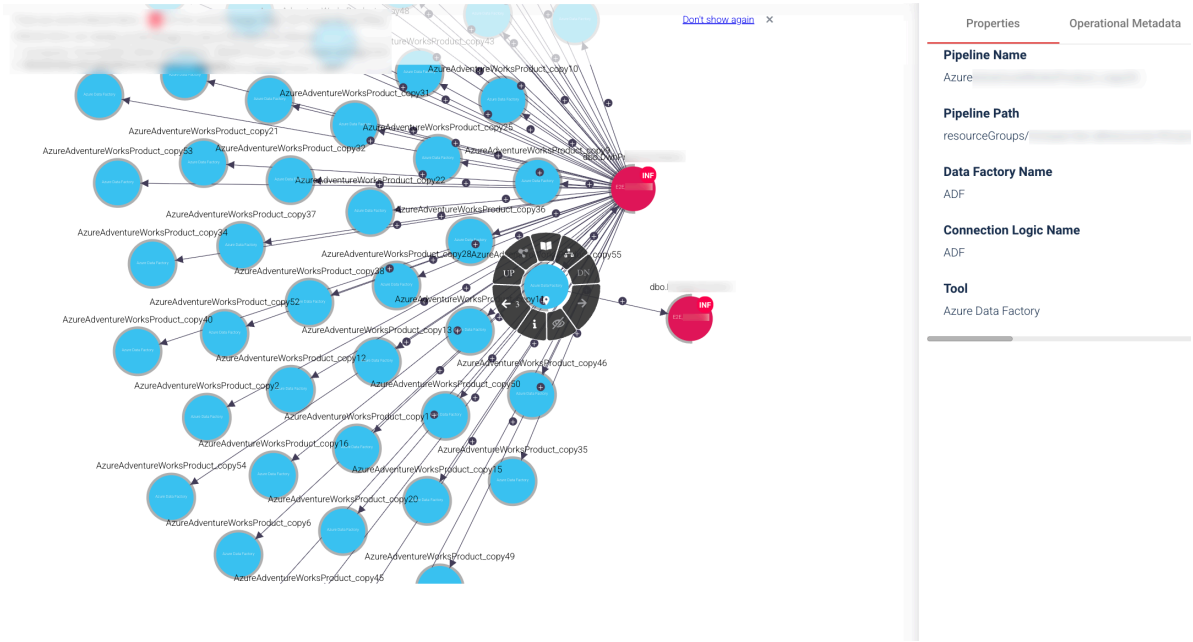
   Note: The same fix applies whether the referencing source is an ETL tool or a BI or reporting tool. Only the Admin Console tab differs. Use the REPORT tab for BI or reporting objects and the ETL tab for ETL objects.

   Use the single-object procedure to fix one reference at a time, or the bulk procedure to fix many at once.

   Single-object

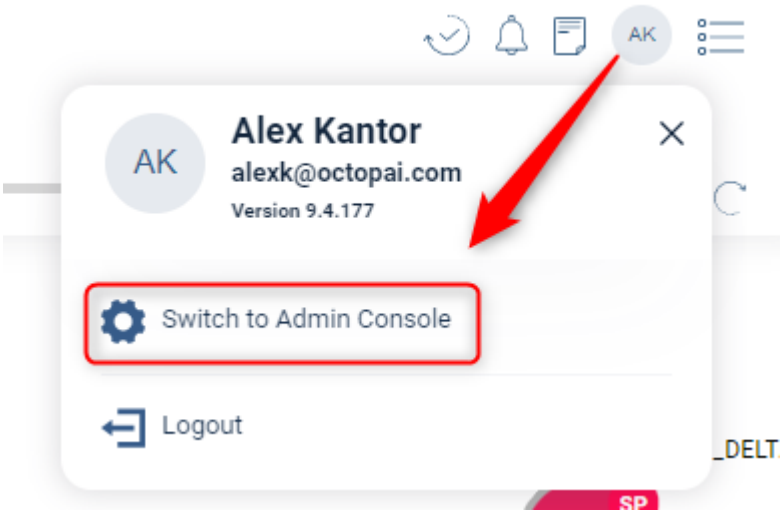
   a) From the **Properties** panel of the Data Object, note the Report or ETL path.

   Figure 17: Path in the Properties panel



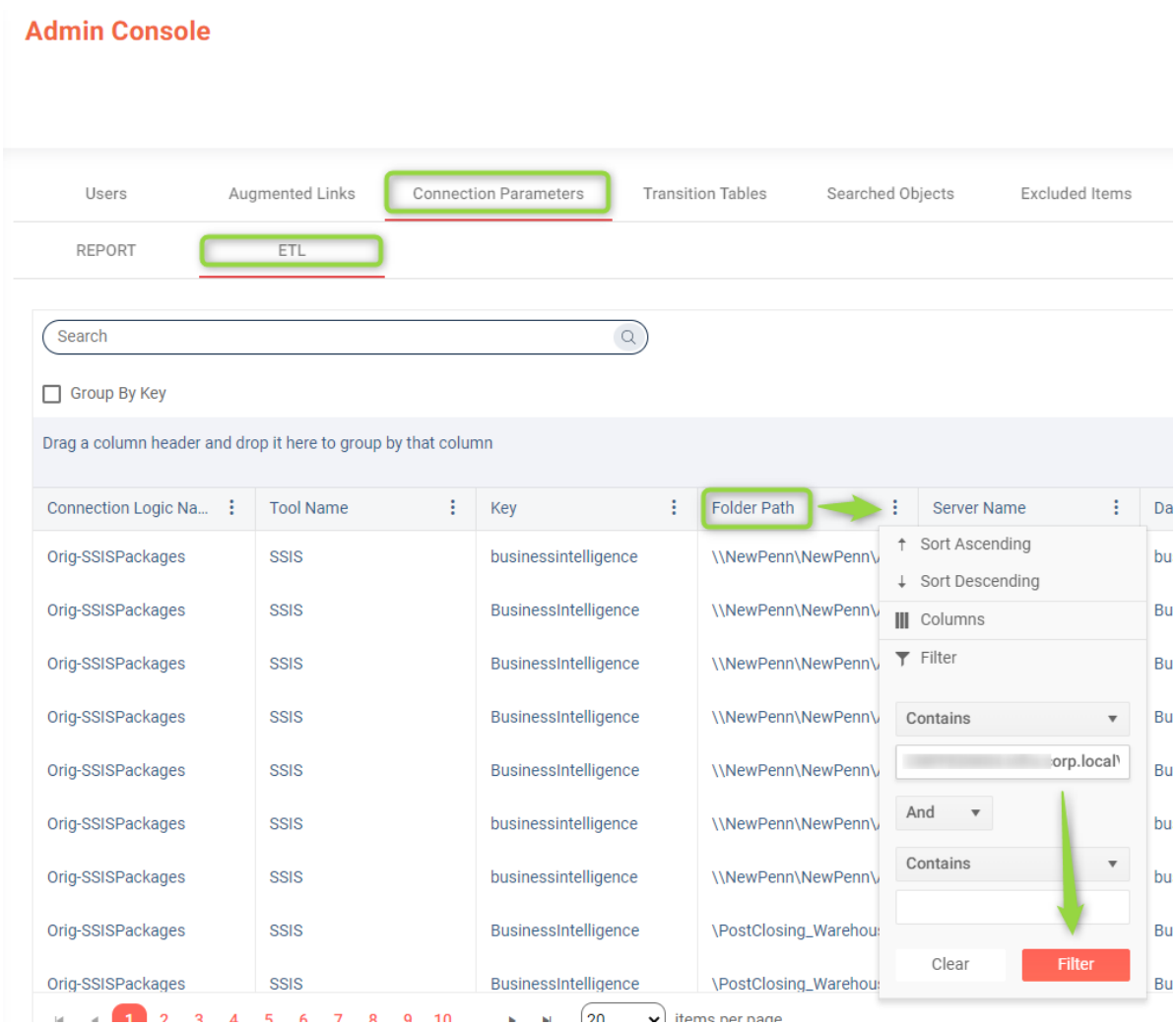
b) Switch to Admin Console from the user menu at the top right.

Figure 18: Switch to Admin Console



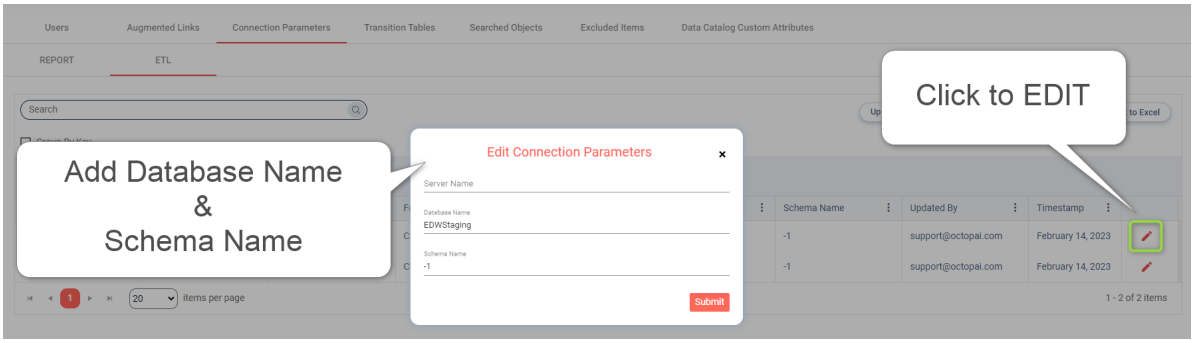
c) In the **Admin Console**, click Connection Parameters, select the REPORT or ETL tab depending on the object type, and find the row matching the path you noted.

Figure 19: Admin Console — Connection Parameters



d) Click the Edit icon, fill in the missing Database Name and Schema Name, and click Submit.

Figure 20: Add the connection parameters

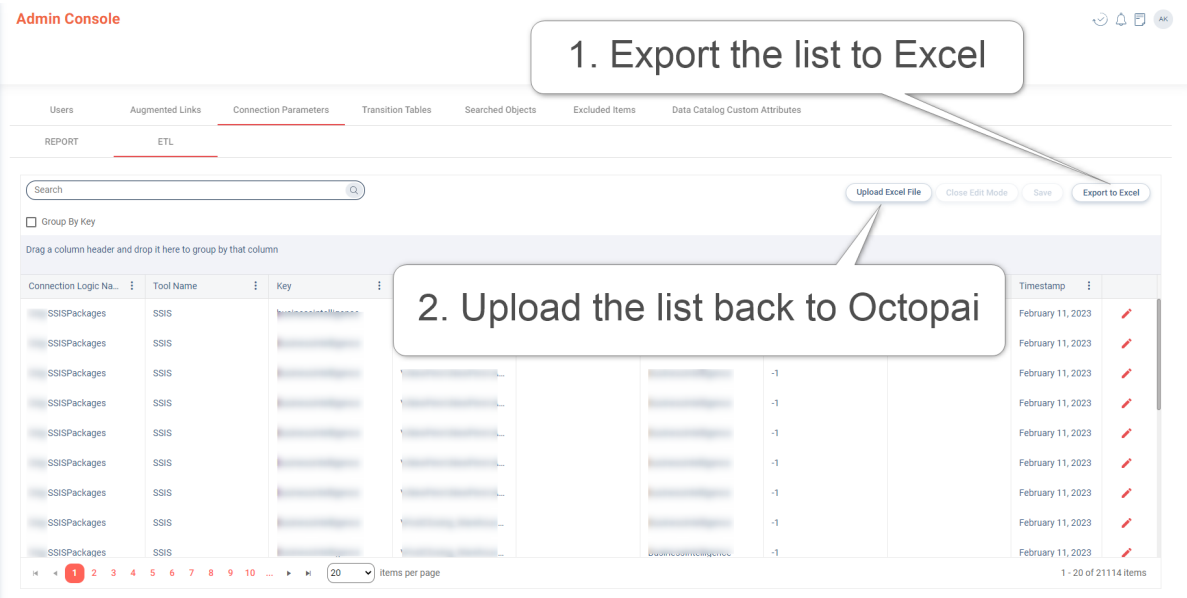


The next Cloudera Data Lineage nightly job picks up the changes and re-resolves affected objects.

Bulk (many objects at once)

a) From Admin Console Connection Parameters , click Export to Excel.

Figure 21: Bulk edit — Export to Excel



b) In the exported file, enter the missing "Database Name" and "Schema Name" fields for Reports and the "Database Name" field for ETL rows.

c) Upload the completed file using the Upload Excel File option. The next Cloudera Data Lineage nightly job picks up the changes and re-resolves affected objects.

What to do next

If an object is still shown as Inferred after a nightly run and you have ruled out scope and permissions issues, contact Cloudera Support. At that point it is likely an extraction or parser issue rather than something you can fix from the Admin Console.