Cloudera Octopai Data Lineage 1.0.0

How To

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

https://docs.cloudera.com/

Legal Notice

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.

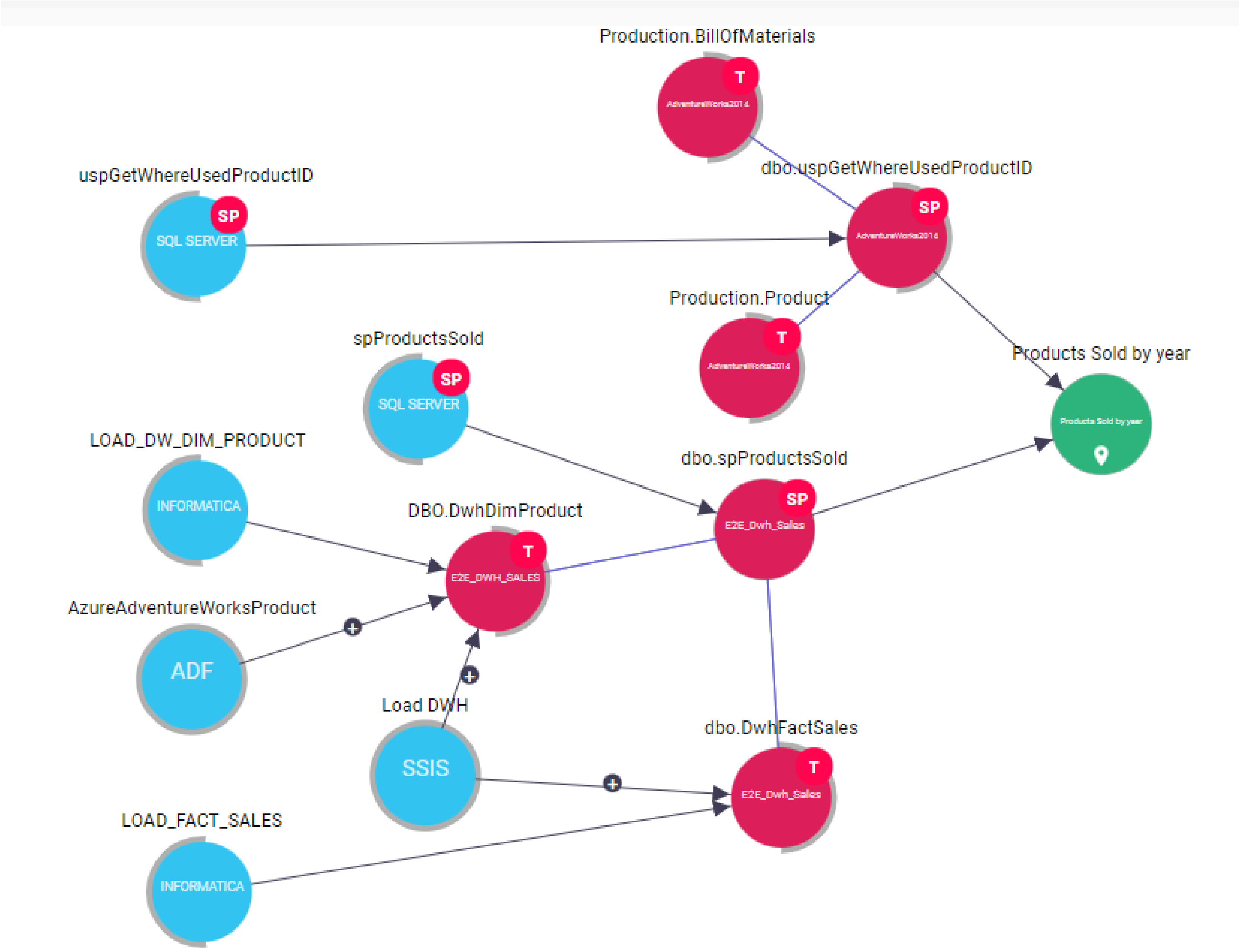
Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES 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

Cross system lineage

Cross System Lineage is a feature or capability provided by Cloudera Octopai, a data management and metadata management platform. Cloudera Octopai is designed to help organizations understand, govern, and optimize their data assets across various systems and platforms.

Figure 1: Visualized data flow



Cross System Lineage specifically focuses on tracking and visualizing the flow of data across different systems within an organization. It provides a comprehensive view of how data moves from its source to its destination, traversing through multiple systems, applications, and processes.

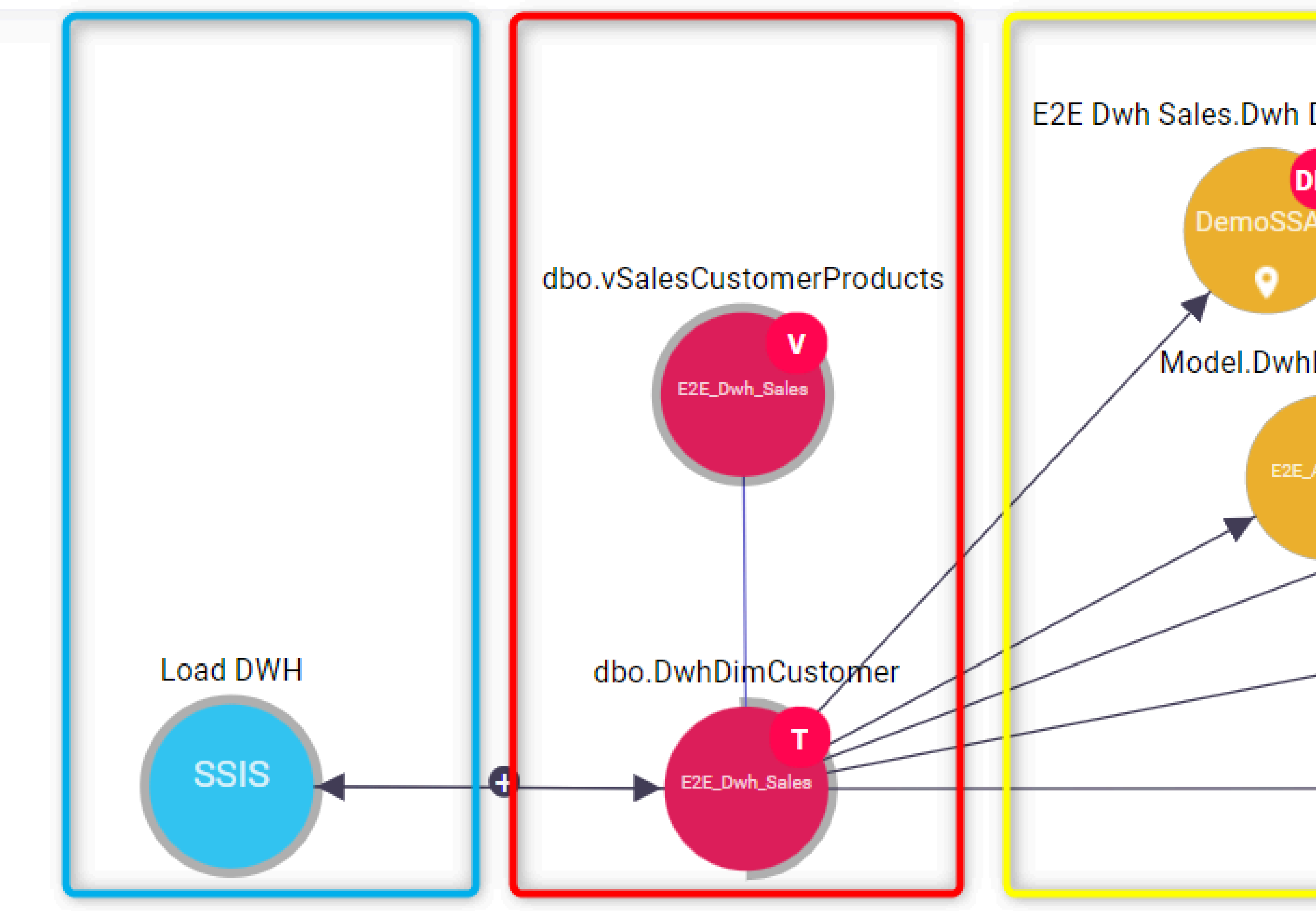
With Cross System Lineage, Cloudera Octopai enables users to gain insights into the end-to-end data lineage, regardless of the complexity of the data ecosystem. It allows users to trace the data path across systems such as databases, data warehouses, data lakes, Extract, Transform, Load (ETL) processes, Business Intelligence (BI) tools, and more.

Cross System Lineage has the following benefits:

- Understanding data flow – Users can track the flow of data from its origin to its final destination, providing a clear understanding of how data is transformed and used throughout the organization.

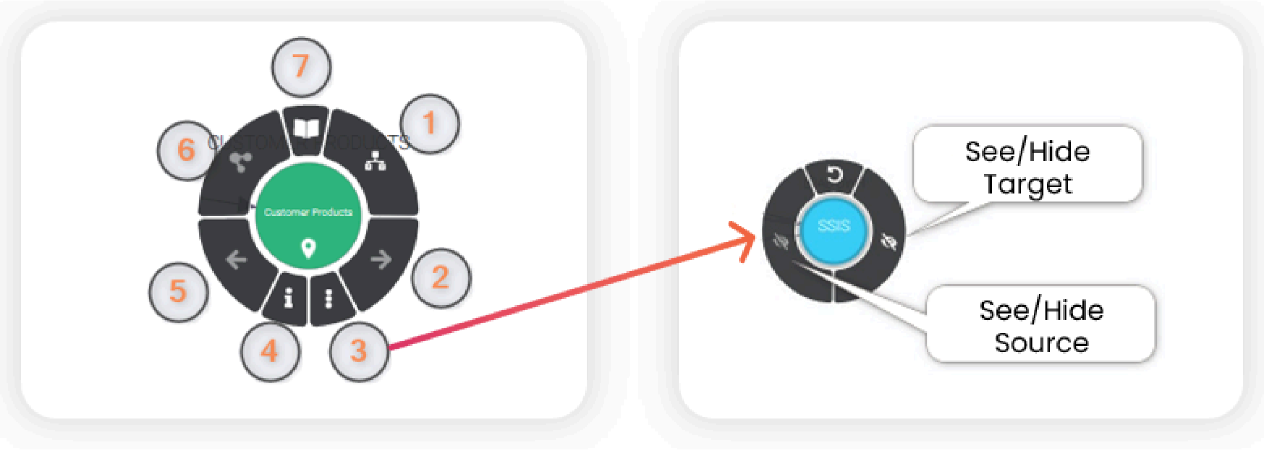
- Impact analysis – Cross System Lineage helps users identify the impact of changes or issues in one system on downstream systems. It allows organizations to assess the potential consequences of modifications, ensuring data integrity and minimizing risks.
- Compliance and governance – By providing visibility into the movement of data across systems, Cloudera Octopai Cross System Lineage assists in meeting compliance requirements and data governance initiatives. It helps organizations maintain data lineage documentation and ensure data accuracy, privacy, and security.
- Troubleshooting and root cause analysis – When data-related issues occur, Cross System Lineage aids in identifying the root causes and troubleshooting effectively. It enables users to pinpoint where problems arise within the data flow and take appropriate actions to resolve them.

Overall, Cross System Lineage offered by Cloudera Octopai enhances data understanding, enables efficient data management, and facilitates informed decision-making across complex data landscapes by visualizing the end-to-end flow of data across systems.



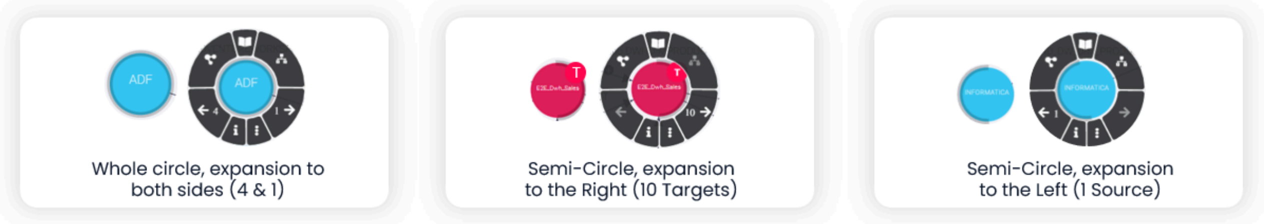
Clicking on each Data Object Bubble will show a Radial button with the following Cross System Lineage functionalities:

Figure 2: Data object bubble functionalities



1. **1.** Hop on to Inner View – Internal lineage view of the component
2. **2.** Lineage Expansion – Impact analysis
3. **3.** More Actions – See or hide target and see or hide source
4. **4.** Information – Component properties
5. **5.** Lineage Expansion – Root cause analysis
6. **6.** Lineage Focus – Change focus to this item
7. **7.** Hop to Catalog Module – Automatic Data Catalog, if available

Figure 3: Data object bubbles with full circle and semi-circle



Whole circle, expansion to both sides (4 & 1)

Semi-Circle, expansion to the Right (10 Targets)

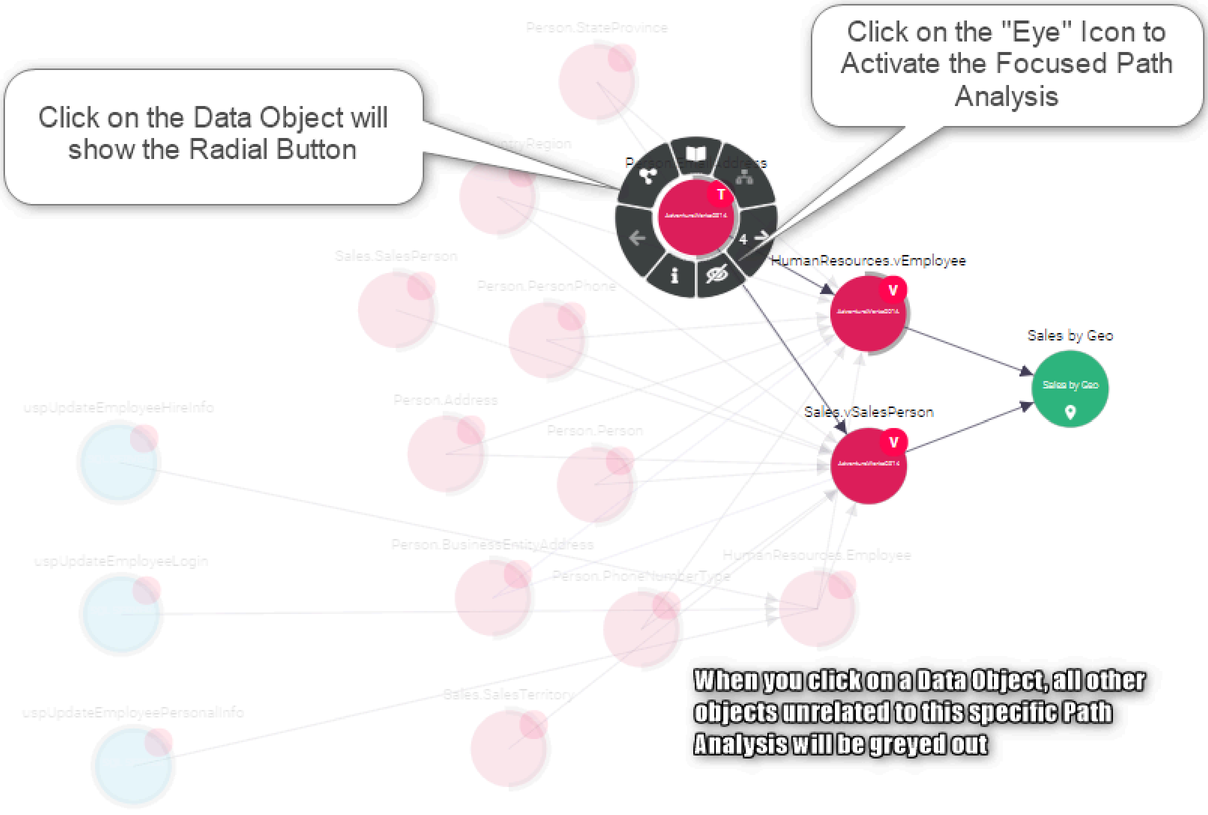
Semi-Circle, expansion to the Left (1 Source)

Enhanced Focused Path Analysis

The Cloudera Octopai focused path analysis tool offers better usability and clearer visual indications with the following enhancements:

- Visual Indicators for Selected Objects – When analyzing cross-system data flows, any object selected for focused path analysis displays a visual indication. This makes it easier for users to identify which objects are part of the focused path.
- Improved Object Selection – If an object cannot be selected for focused path analysis, it means the map is already reduced to that specific path. The map shows all objects going through the selected object and their connected objects.
- Stable Map Filters – If your analysis is focused on a specific path, filters cannot be activated. This ensures the stability of the map, as any filter changes would trigger a map recalculation. For optimal results, Cloudera Octopai recommends configuring filters before applying focused path analysis.

Figure 4: Active and inactive data objects

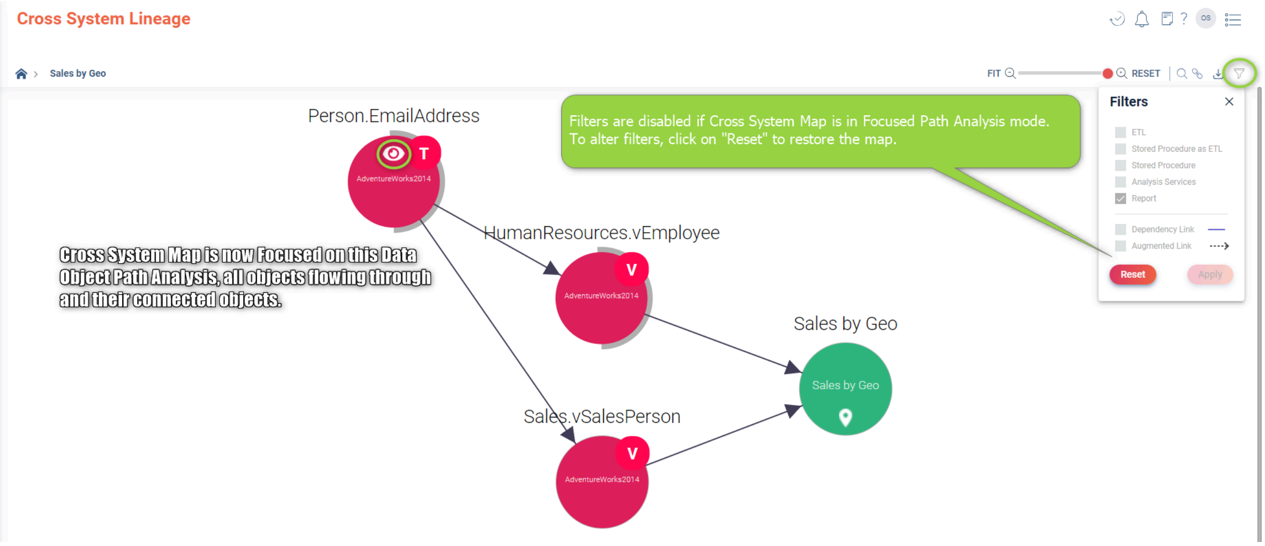


Click on the Data Object will show the Radial Button

Click on the "Eye" Icon to Activate the Focused Path Analysis

When you click on a Data Object, all other objects unrelated to this specific Path Analysis will be greyed out

Figure 5: Focused cross system map



Cross System Lineage

Sales by Geo

Filters are disabled if Cross System Map is in Focused Path Analysis mode. To alter filters, click on "Reset" to restore the map.

Cross System Map is now Focused on this Data Object Path Analysis, all objects flowing through and their connected objects.

Filters

□ ETL
□ Stored Procedure as ETL
□ Stored Procedure
□ Analysis Services
☑ Report
□ Dependency Link
□ Augmented Link

Inner System Lineage

The Cloudera Octopai connector for Apache Hive enables metadata extraction and lineage tracking with setup and troubleshooting guidance.

Inner System Lineage is a feature within Cloudera Octopai, a data management platform designed to assist organizations in effectively managing their data assets. The Inner System Lineage functionality provides users with a comprehensive understanding of the relationships and dependencies that exist between various data elements within their systems.

The Inner System Lineage feature within Cloudera Octopai enables users to track the flow of data across different stages of data processing, such as data extraction, transformation, and loading. It offers a visual representation of the data lineage, allowing users to navigate through the complex web of data connections and gain insights into the origin, transformations, and destinations of their data.

When using Inner System Lineage, users can identify the sources from which their data originates and the intermediate steps through which it passes before reaching its final destination. This knowledge is crucial for ensuring data accuracy, understanding data transformations, and troubleshooting issues that may arise during the data management process.

The Cloudera Octopai Inner System Lineage feature also provides users with the ability to view the lineage of specific data attributes or columns. This level of granularity allows users to trace the path of a particular data element, understand its transformations, and determine where it is used across different reports, dashboards, or downstream systems.

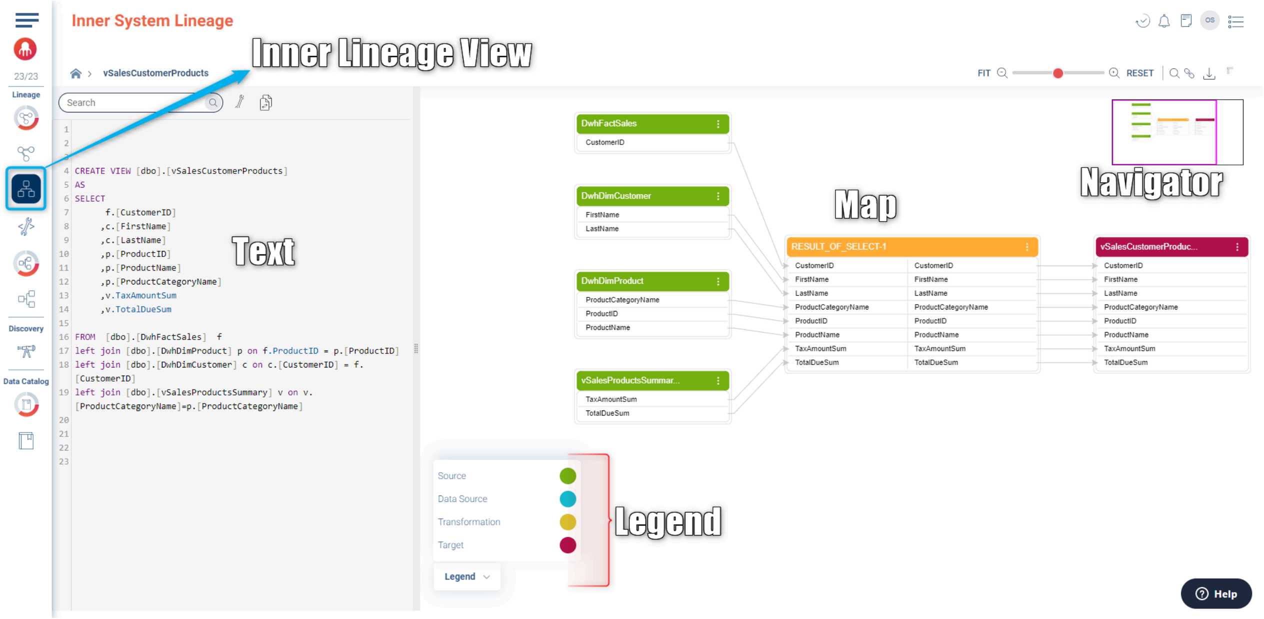
By leveraging Inner System Lineage, users can achieve several benefits. They can improve data governance by gaining a deeper understanding of data flows and relationships, facilitating compliance with regulatory requirements. It also enhances data quality management by identifying potential data lineage issues or bottlenecks that may impact data accuracy or timeliness.

Furthermore, Inner System Lineage in Cloudera Octopai simplifies the process of impact analysis. Users can easily assess the potential effects of making changes to a specific data source or transformation logic by tracing the downstream impact on other data elements and associated reports or processes.

In summary, Inner System Lineage in Cloudera Octopai empowers users with a clear and visual representation of data lineage, enabling them to understand the data flow, identify dependencies, and improve data governance and quality management. By leveraging this feature, organizations can gain valuable insights into their data landscape and make informed decisions regarding data management and analytics processes.

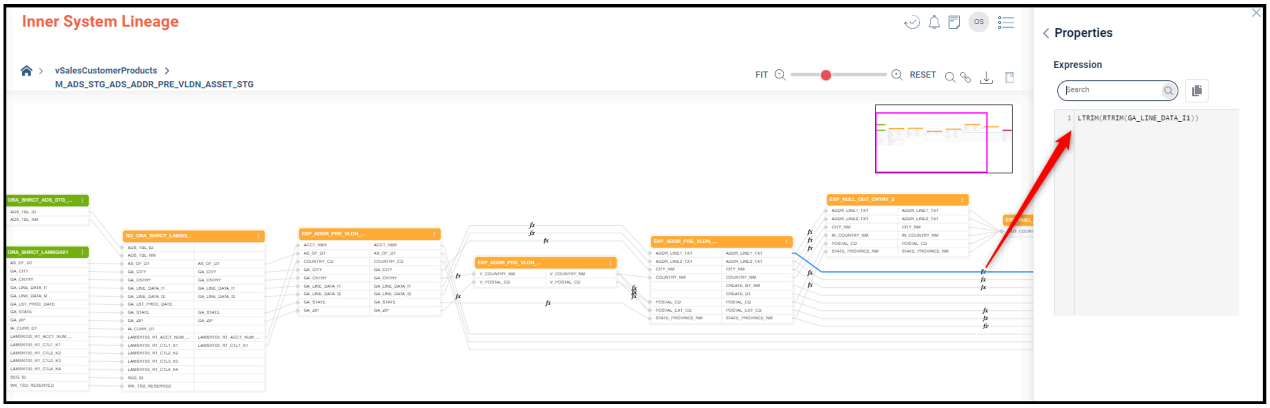
Top use cases include:

- Visualizing the logic of a report, ETL, or database object data flow
- Locating dependencies within a report



```
CREATE VIEW [dbo].[vSalesCustomerProducts]
AS
SELECT
      f.[CustomerID]
     ,c.[FirstName]
     ,c.[LastName]
     ,p.[ProductID]
     ,p.[ProductName]
     ,p.[ProductCategoryName]
     ,v.TaxAmountSum
     ,v.TotalDueSum

FROM  [dbo].[DwhFactSales]  f
left join [dbo].[DwhDimProduct] p on f.ProductID = p.[ProductID]
left join [dbo].[DwhDimCustomer] c on c.[CustomerID] = f.
[CustomerID]
left join [dbo].[vSalesProductsSummary] v on v.
[ProductCategoryName]=p.[ProductCategoryName]
```

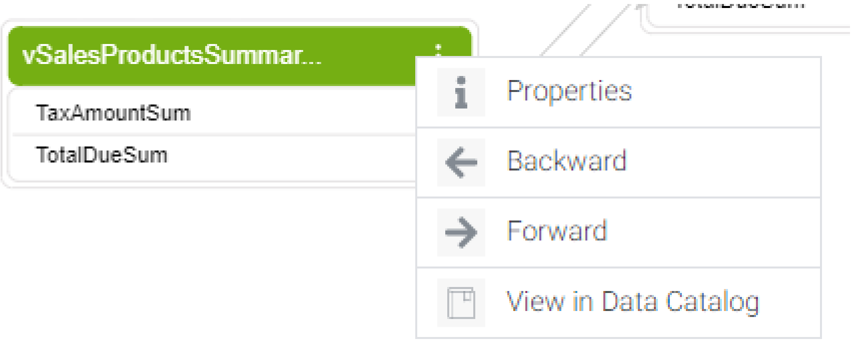
Double click on the "fx" sign or hexagonal icon in the upper left corner of the table to see the properties of the expression / function.



```
LTRIM(RTRIM(GA_LINE_DATA_I1))
```

Expand your lineage

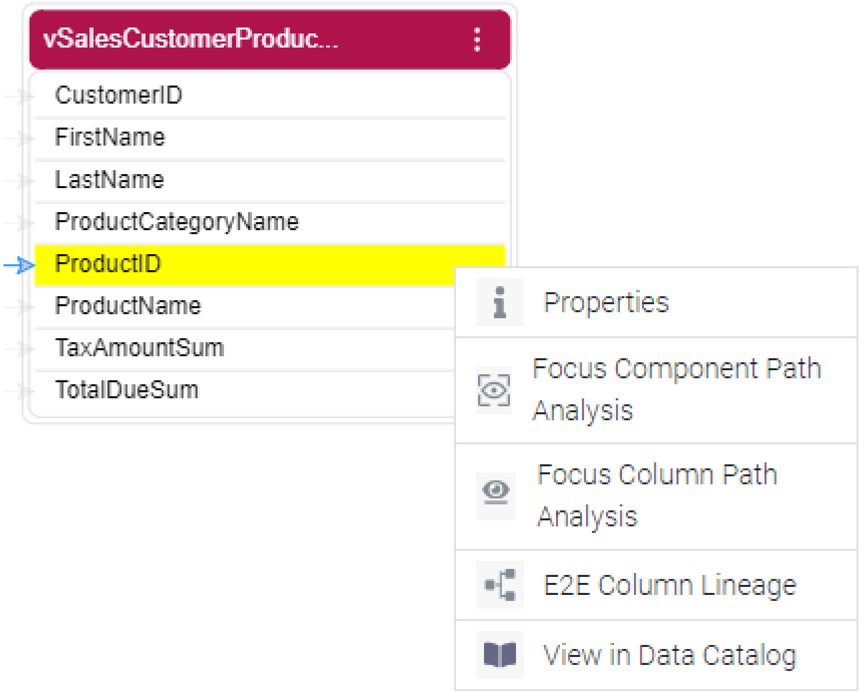
From a source/Target Table, click on the three dots on the top-right corner to access more options



- Get the table properties
- Hop Backwards to the table source (only from source tables - green)
- Hop Forward to the table target (only from target tables - red)

- Hop to Catalog Module

Figure 6: Inner System Lineage Feature Overview



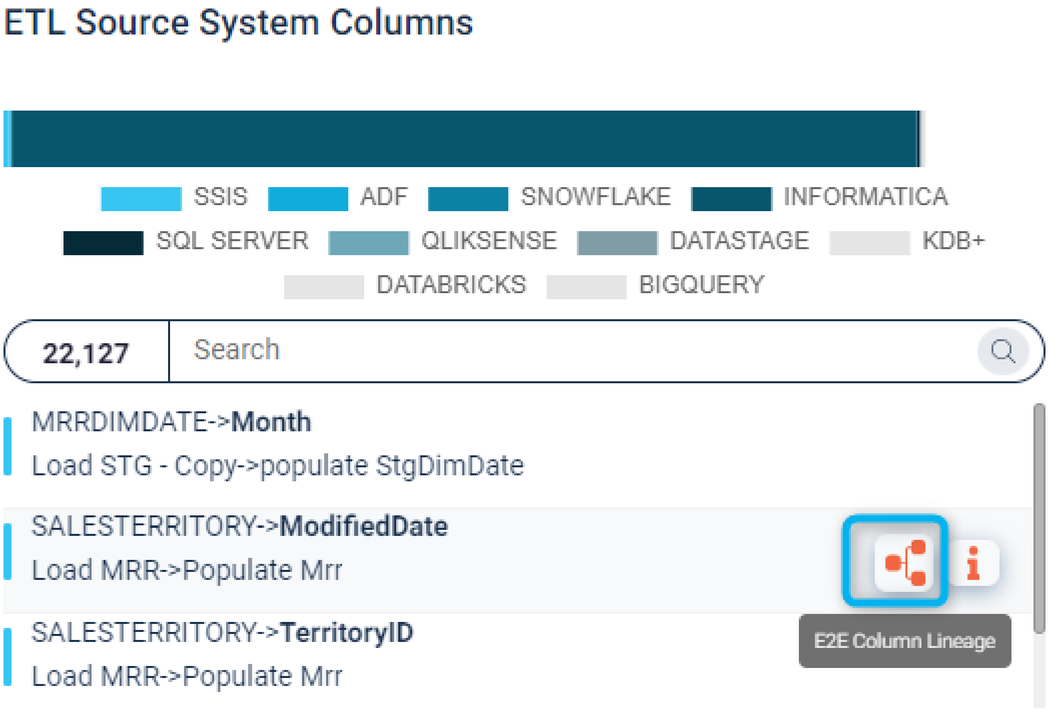
Click on the three dots next to the column name to open the drop-down menu with the following options:

- **Column Properties:** The column properties will appear on the left side of the screen
- **Focus Component Path Analysis:** Focused Component Path Analysis
- **Focus Column Path Analysis:** Focused Column Path Analysis
- **E2E Column Lineage:** End-to-End Column Lineage
- **View in Data Catalog:** Data Catalog Module

End-to-End Column Lineage

E2E Column Lineage

Hop to End to End Column Lineage



ETL Source System Columns

SSIS ADF SNOWFLAKE INFORMATICA
SQL SERVER QLIKSENSE DATASTAGE KDB+
DATABRICKS BIGQUERY

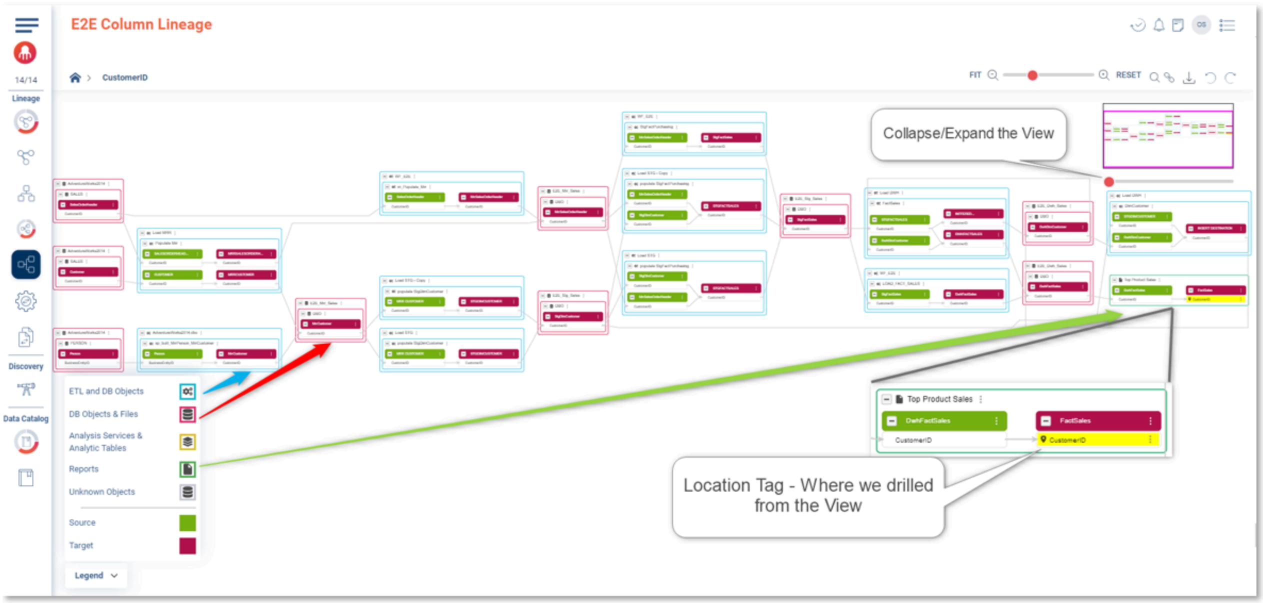
22,127 Search

MRRDIMDATE->**Month**
Load STG - Copy->populate StgDimDate

SALESTERRITORY->**ModifiedDate**
Load MRR->Populate Mrr

SALESTERRITORY->**TerritoryID**
Load MRR->Populate Mrr

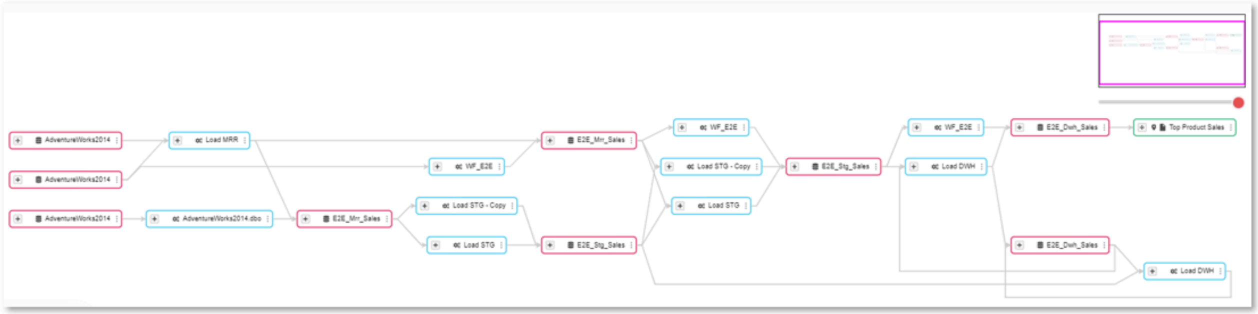
E2E Column Lineage



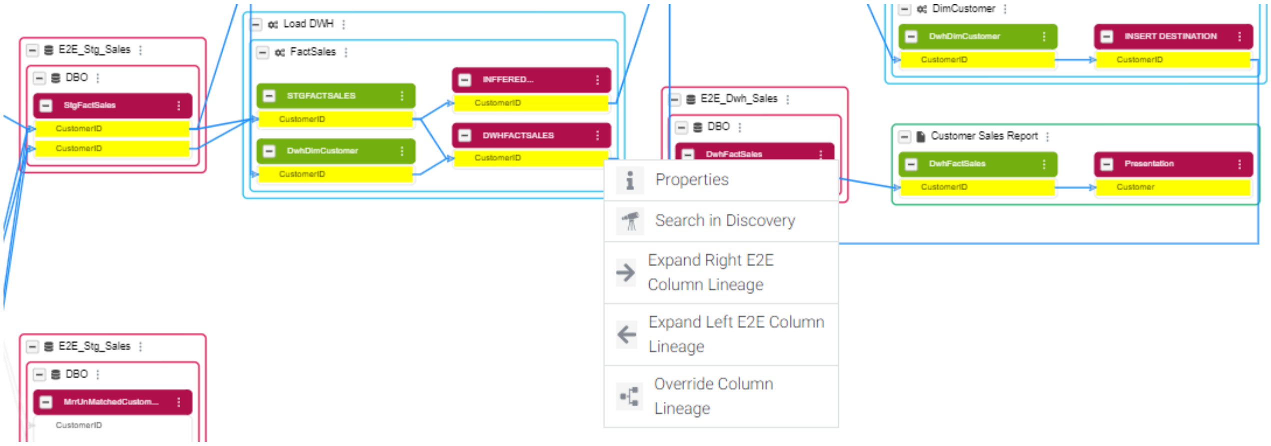
Collapse/Expand the View

Location Tag - Where we drilled from the View

Collapse/Expand the View from the bar under the Navigation panel



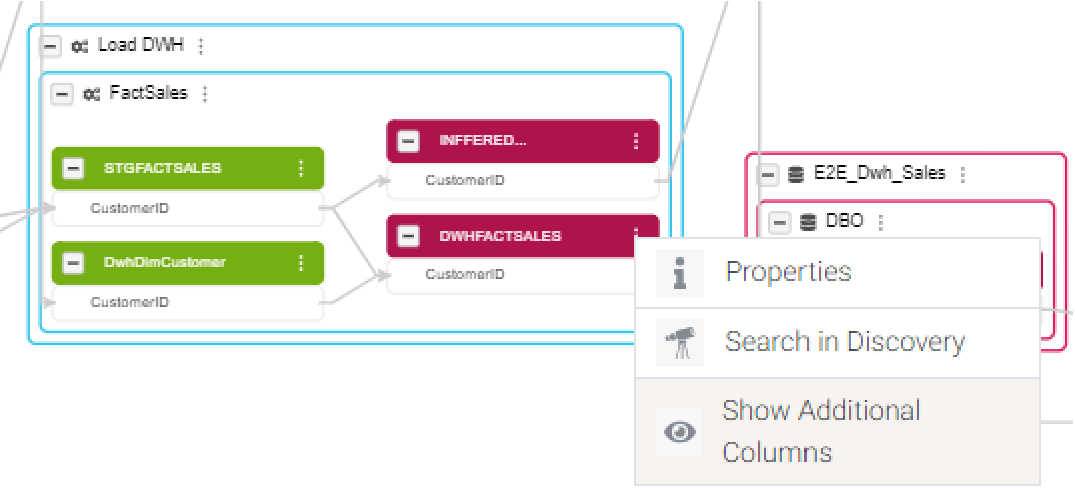
E2E Column Lineage Functionalities over each Column



Click on the three dots next to the column to open the drop menu with the following options:

- Properties - Column Properties will pop up on the left side of the screen
- Search in Discovery - Hop to Discovery Module
- Expand Right E2E Column Lineage - Impact Analysis
- Expand Left E2E Column Lineage - Root Cause Analysis
- Override Column Lineage - Start the E2E Lineage from this Column

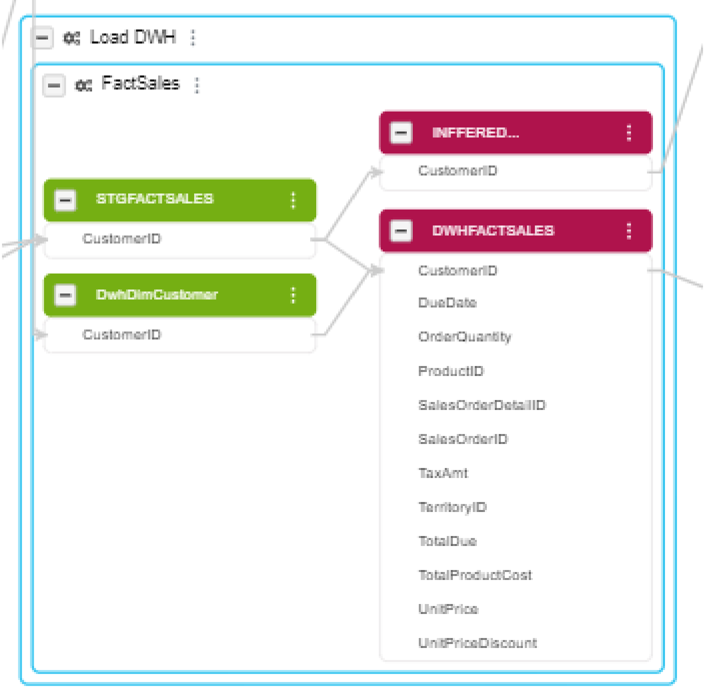
E2E Column Lineage Functionalities over each Table



Click on the three dots next to the table to open the drop menu with the following options:

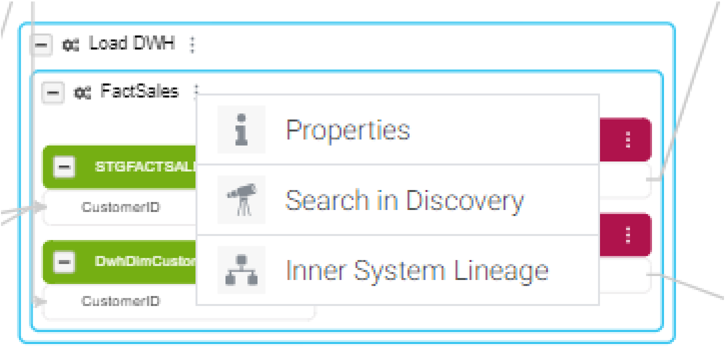
- Properties - Table Properties will pop up on the left side of the screen
- Search in Discovery - Hop to Discovery Module

- Show Additional Columns - Will show all the columns related to the table



Same options can be selected like the E2E functionalities over each column.

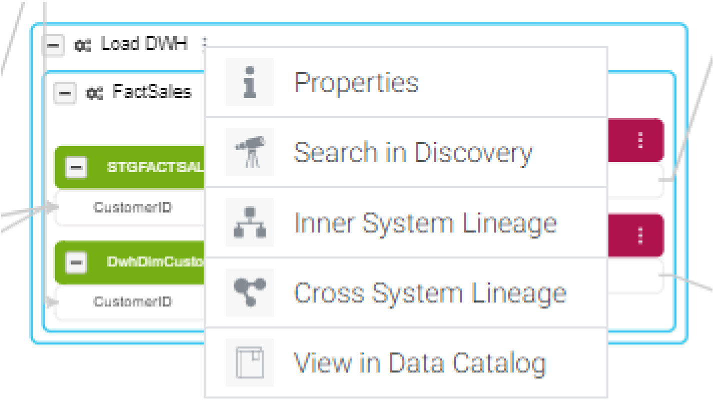
E2E Column Lineage Functionalities over each Inner Square



Click on the three dots next to the inner square name to open the drop menu with the following options:

- Properties - Table Properties will pop up on the left side of the screen
- Search in Discovery - Hop to Discovery Module
- Inner System Lineage - Hop to the Inner System Lineage

E2E Column Lineage Functionalities over each Outer Square



Click on the three dots next to the table to open the drop menu with the following options:

- Properties - Table Properties will pop up on the left side of the screen
- Search in Discovery - Hop to Discovery Module
- Inner System Lineage - Hop to the Inner System Lineage
- Cross System Lineage - Hop to the Cross System Lineage
- View in Data Catalog - Hop to Catalog Module

Live Lineage

Welcome to the comprehensive guide on Live Lineage, a robust tool designed to streamline data management tasks. This guide aims to provide a detailed walkthrough of Live Lineage's functionalities, highlighting practical use cases, and providing useful tips and tricks.

Introduction

Welcome to the comprehensive guide on Live Lineage, a robust tool designed to streamline data management tasks. This guide aims to provide a detailed walkthrough of Live Lineage's functionalities, highlighting practical use cases, and providing useful tips and tricks.

Live Lineage can be applied in various scenarios:

1. **Fixing Broken Data Lineage Due to Script Errors** : Utilize Live Lineage to detect and correct errors causing disruptions to data flow.
2. **Simulating Script Changes** : Test a script change that could affect a data lineage flow before deploying it to production.
3. **Script Migration** : Live Lineage enables confident migration of scripts from one database system to another.

The Live Lineage Visualizer supports a wide range of technologies, including SQL Server, Oracle, Teradata, Netezza, Vertica, Snowflake, MySQL, Hive, PostgreSQL, DB2, Redshift, Google BigQuery, and SAP Hana.

Key Features of Live Lineage

1. Streamlined Migration Projects

What it does: Live Lineage enables you to conduct seamless transitions between different systems, such as Oracle and Snowflake. It allows you to identify potential issues and resolve them before initiating the migration process, ensuring a smooth transition.

How to use it: Access the Live Lineage module, select the script you want to migrate, and use the Visualizer tool to review potential migration issues and resolve them.

Pro tip: Start with a few scripts before migrating all your scripts to avoid a total system halt if issues arise.

2. Syntax Error Detection and Resolution

What it does: Leveraging real-time SQL script visualization, Live Lineage identifies and corrects syntax errors proactively, maintaining a smooth data pipeline.

How to use it: Input your script into the Live Lineage Visualizer. The module will scan and highlight syntax errors, offering potential solutions.

Pro tip: Regularly review the syntax error alerts and address errors promptly to avoid delays and complications.

3. Script Updates

What it does: Live Lineage empowers users to simulate changes before deploying them to production, guaranteeing accurate and reliable script modifications.

How to use it: Choose the script you want to update, make the necessary changes, then use the simulation feature to assess the impact of these changes.

Pro tip: Frequently test script changes using the simulation feature to preemptively identify potential issues and disruptions.

Working with Live Lineage

1. Navigating Live Lineage

To use Live Lineage, enter your Cloudera Octopai platform and navigate to the Live Lineage module. This module enables you to input or edit scripts and visualize their data lineage and potential impact of any changes. You can search for key words or expressions within the script or correlate specific data lineage tables with the script sections. Use actions like Play, Delete or Copy the script for ongoing activities.

Whilst you edit a script or simulating the script behavior upon migration, utilize the Error space to understand the errors you need to fix.

2. Accessing Scripts via Inner System Lineage Path or Discovery

Live Lineage also allows users to access and edit existing scripts through the Inner System Lineage path or Discovery modules. To do this, double-click on components containing the script and press “Edit”. The changes and activities will be recorded in the Recent Activities.

Pro tip: Utilize the Inner System Lineage path to quickly access and modify existing scripts without having to navigate through your entire database especially in context of Change Impact Analysis and Migration.

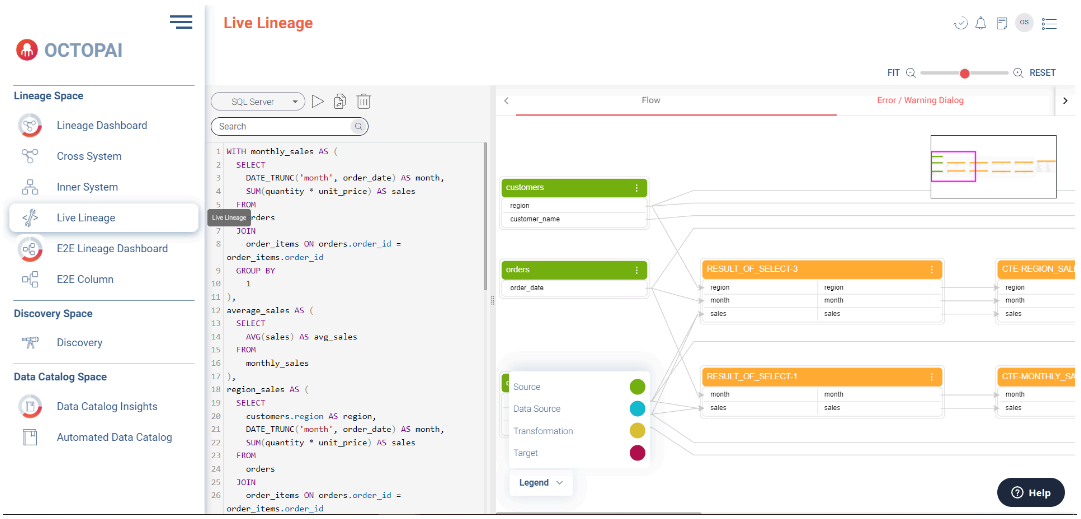
3. Recent Activities

Whether you're editing an existing script via your Inner System Lineage path or creating a new one, all activities are recorded under Recent Activities. Navigate to your Recent Activities to access this data. Activities executed using Live Lineage will be labeled as "Show Live Lineage."

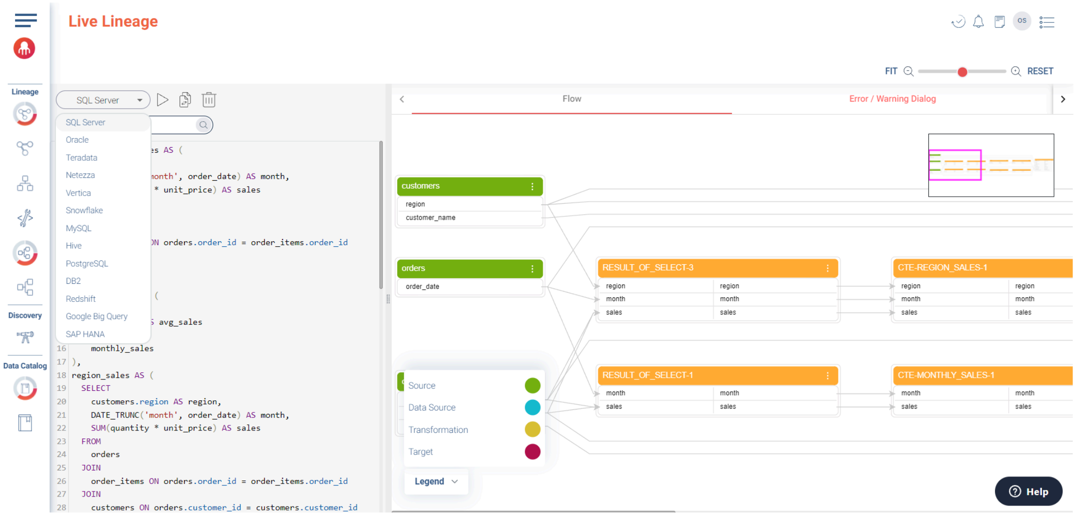
4. Working with Scripts

If the script you've worked on is derived from an existing script, the Recent Activities will display the original script's name. If you're editing the script or adding a new one, the activity will be marked as "Custom SQL Script."

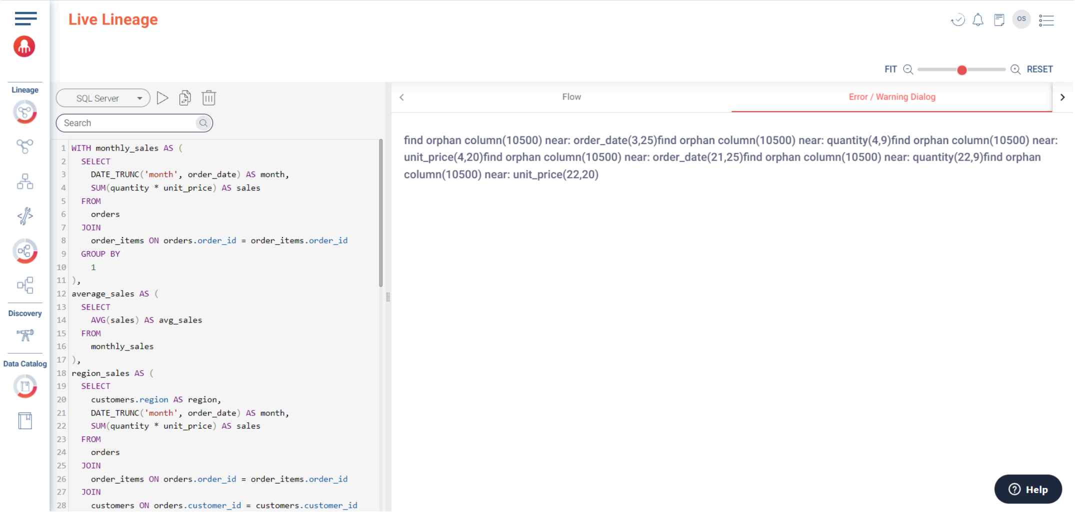
Pro tip: Keeping track of your Recent Activities allows you to monitor your script changes and workflow, and can be a lifesaver when troubleshooting.



```
WITH monthly_sales AS (
  SELECT
    DATE_TRUNC('month', order_date) AS month,
    SUM(quantity * unit_price) AS sales
  FROM
    orders
  JOIN
    order_items ON orders.order_id =
order_items.order_id
  GROUP BY
    1
),
average_sales AS (
  SELECT
    AVG(sales) AS avg_sales
  FROM
    monthly_sales
),
region_sales AS (
  SELECT
    customers.region AS region,
    DATE_TRUNC('month', order_date) AS month,
    SUM(quantity * unit_price) AS sales
  FROM
    orders
  JOIN
    order_items ON orders.order_id =
order_items.order_id
```

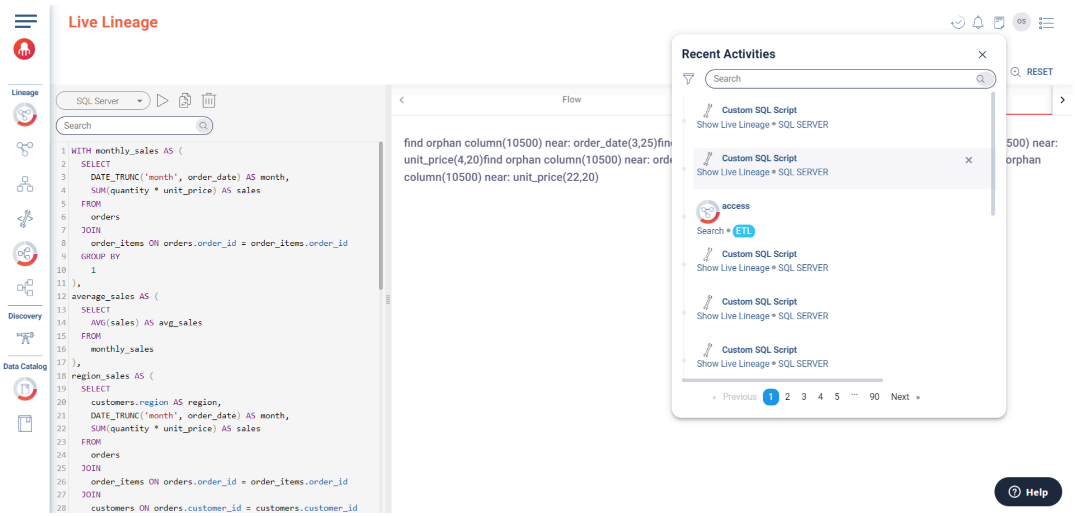


```
es AS (

onth', order_date) AS month,
 * unit_price) AS sales



N orders.order_id = order_items.order_id



 (

 avg_sales
    monthly_sales
),
region_sales AS (
  SELECT
    customers.region AS region,
    DATE_TRUNC('month', order_date) AS month,
    SUM(quantity * unit_price) AS sales
  FROM
    orders
  JOIN
    order_items ON orders.order_id = order_items.order_id
  JOIN
    customers ON orders.customer_id = customers.customer_id
```



```
WITH monthly_sales AS (
  SELECT
    DATE_TRUNC('month', order_date) AS month,
    SUM(quantity * unit_price) AS sales
  FROM
    orders
  JOIN
    order_items ON orders.order_id = order_items.order_id
  GROUP BY
    1
),
average_sales AS (
  SELECT
    AVG(sales) AS avg_sales
  FROM
    monthly_sales
),
region_sales AS (
  SELECT
    customers.region AS region,
    DATE_TRUNC('month', order_date) AS month,
    SUM(quantity * unit_price) AS sales
  FROM
    orders
  JOIN
    order_items ON orders.order_id = order_items.order_id
  JOIN
    customers ON orders.customer_id = customers.customer_id
```

find orphan column(10500) near: order_date(3,25)find orphan column(10500) near: quantity(4,9)find orphan column(10500) near: unit_price(4,20)find orphan column(10500) near: order_date(21,25)find orphan column(10500) near: quantity(22,9)find orphan column(10500) near: unit_price(22,20)



```
WITH monthly_sales AS (
  SELECT
    DATE_TRUNC('month', order_date) AS month,
    SUM(quantity * unit_price) AS sales
  FROM
    orders
  JOIN
    order_items ON orders.order_id = order_items.order_id
  GROUP BY
    1
),
average_sales AS (
  SELECT
    AVG(sales) AS avg_sales
  FROM
    monthly_sales
),
region_sales AS (
  SELECT
    customers.region AS region,
    DATE_TRUNC('month', order_date) AS month,
    SUM(quantity * unit_price) AS sales
  FROM
    orders
  JOIN
    order_items ON orders.order_id = order_items.order_id
  JOIN
    customers ON orders.customer_id = customers.customer_id
```

find orphan column(10500) near: order_date(3,25)fin unit_price(4,20)find orphan column(10500) near: orde column(10500) near: unit_price(22,20)

500) near: orphan

Recent Activities

Search

Custom SQL Script
Show Live Lineage • SQL SERVER

Custom SQL Script
Show Live Lineage • SQL SERVER

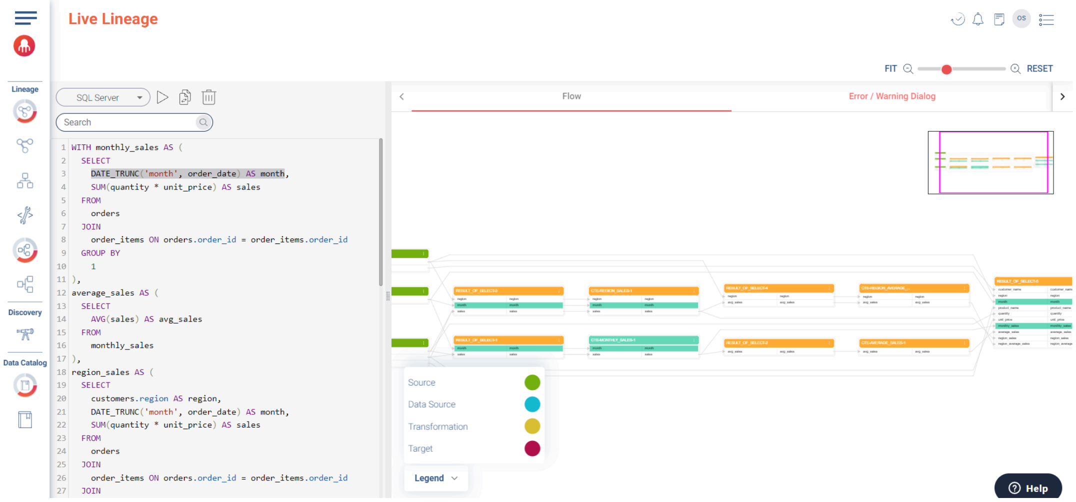
access
Search • ETL

Custom SQL Script
Show Live Lineage • SQL SERVER

Custom SQL Script
Show Live Lineage • SQL SERVER

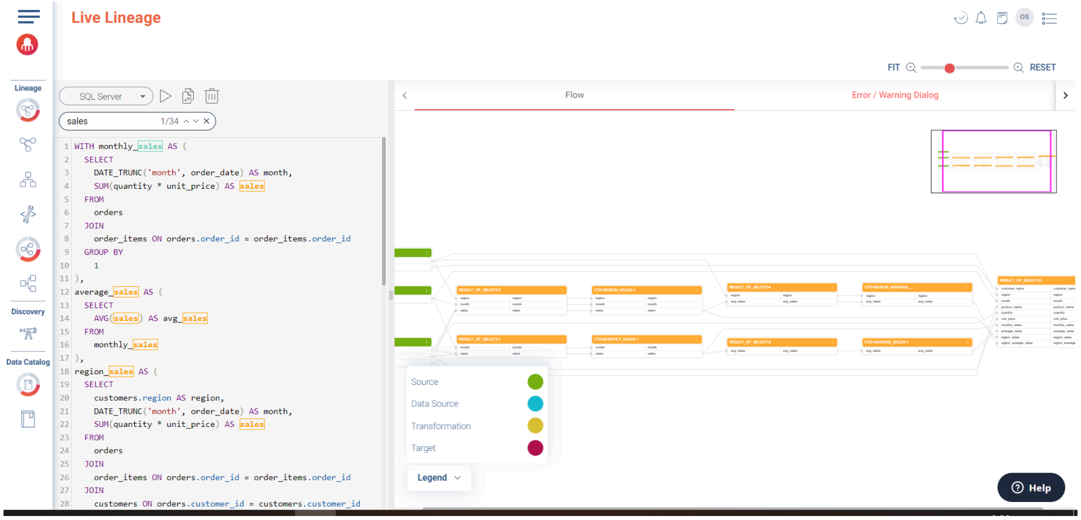
Custom SQL Script
Show Live Lineage • SQL SERVER

« Previous 1 2 3 4 5 ··· 90 Next »



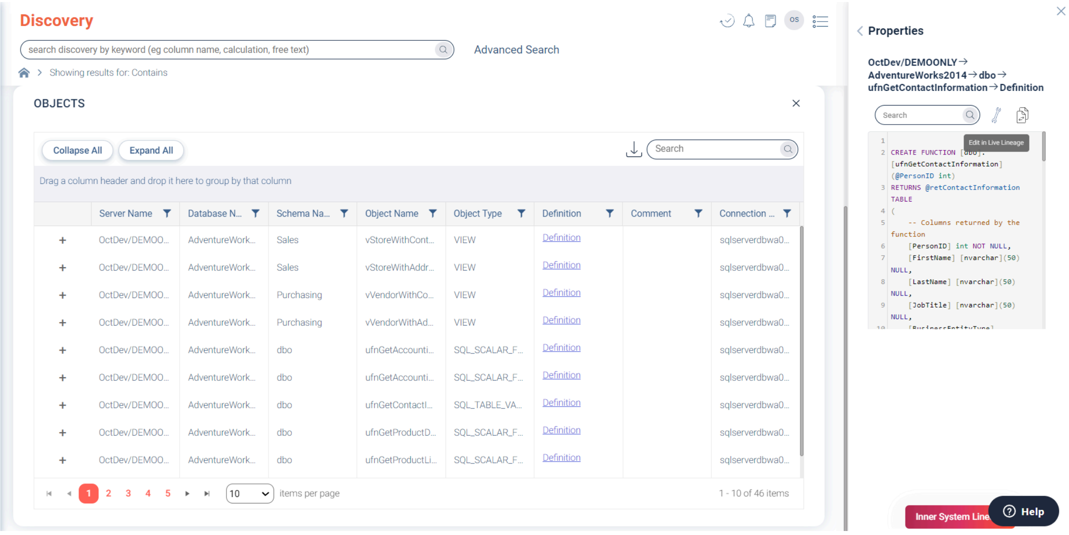
```
WITH monthly_sales AS (
  SELECT
    DATE_TRUNC('month', order_date) AS month,
    SUM(quantity * unit_price) AS sales
  FROM
    orders
  JOIN
    order_items ON orders.order_id = order_items.order_id
  GROUP BY
    1
),
average_sales AS (
  SELECT
    AVG(sales) AS avg_sales
  FROM
    monthly_sales
),
region_sales AS (
  SELECT
    customers.region AS region,
    DATE_TRUNC('month', order_date) AS month,
    SUM(quantity * unit_price) AS sales
  FROM
    orders
  JOIN
    order_items ON orders.order_id = order_items.order_id
  JOIN
```

Source	🟢
Data Source	🔵
Transformation	🟡
Target	🔴



```
WITH monthly_sales AS (
  SELECT
    DATE_TRUNC('month', order_date) AS month,
    SUM(quantity * unit_price) AS sales
  FROM
    orders
  JOIN
    order_items ON orders.order_id = order_items.order_id
  GROUP BY
    1
),
average_sales AS (
  SELECT
    AVG(sales) AS avg_sales
  FROM
    monthly_sales
),
region_sales AS (
  SELECT
    customers.region AS region,
    DATE_TRUNC('month', order_date) AS month,
    SUM(quantity * unit_price) AS sales
  FROM
    orders
  JOIN
    order_items ON orders.order_id = order_items.order_id
  JOIN
    customers ON orders.customer_id = customers.customer_id
```

Source	🟢
Data Source	🔵
Transformation	🟡
Target	🔴



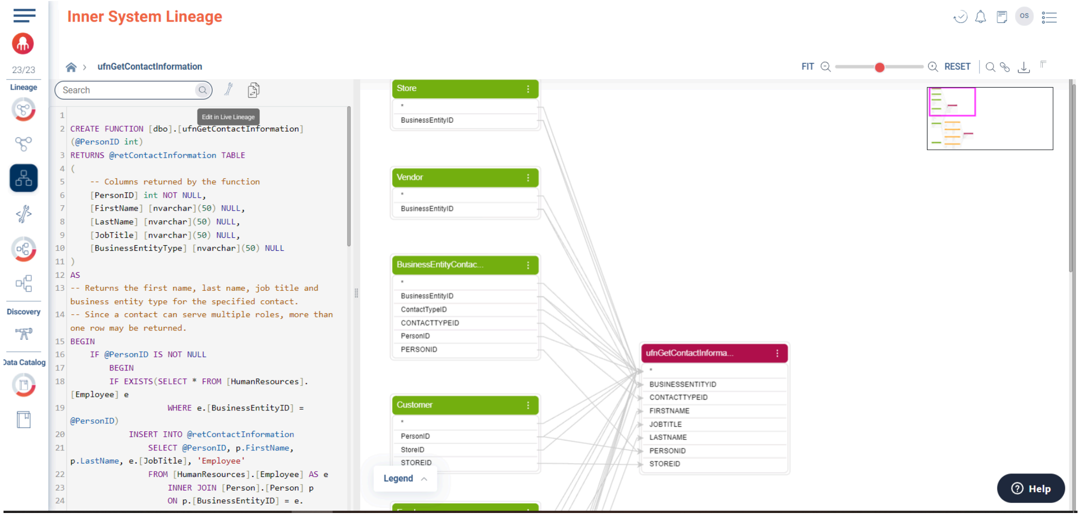
search discovery by keyword (eg column name, calculation, free text)

	Server Name	Database N...	Schema Na...	Object Name	Object Type	Definition	Comment	Connection ...
+	OctDev/DEMOO...	AdventureWork...	Sales	vStoreWithCont...	VIEW	Definition		sqlserverdbwa0...
+	OctDev/DEMOO...	AdventureWork...	Sales	vStoreWithAddr...	VIEW	Definition		sqlserverdbwa0...
+	OctDev/DEMOO...	AdventureWork...	Purchasing	vVendorWithCo...	VIEW	Definition		sqlserverdbwa0...
+	OctDev/DEMOO...	AdventureWork...	Purchasing	vVendorWithAd...	VIEW	Definition		sqlserverdbwa0...
+	OctDev/DEMOO...	AdventureWork...	dbo	ufnGetAccounti...	SQL_SCALAR_F...	Definition		sqlserverdbwa0...
+	OctDev/DEMOO...	AdventureWork...	dbo	ufnGetAccounti...	SQL_SCALAR_F...	Definition		sqlserverdbwa0...
+	OctDev/DEMOO...	AdventureWork...	dbo	ufnGetContactI...	SQL_TABLE_VA...	Definition		sqlserverdbwa0...
+	OctDev/DEMOO...	AdventureWork...	dbo	ufnGetProductD...	SQL_SCALAR_F...	Definition		sqlserverdbwa0...
+	OctDev/DEMOO...	AdventureWork...	dbo	ufnGetProductLi...	SQL_SCALAR_F...	Definition		sqlserverdbwa0...

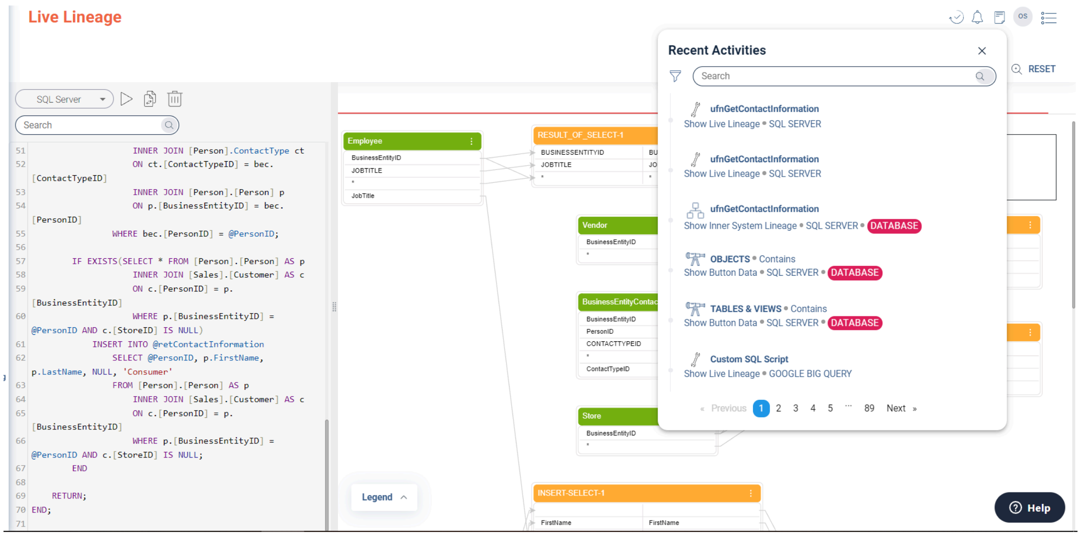
1 2 3 4 5 · 10

OctDev/DEMOONLY → AdventureWorks2014 → dbo → ufnGetContactInformation → Definition

```
CREATE FUNCTION [dbo].[ufnGetContactInformation](@PersonID int)
RETURNS @retContactInformation TABLE
(
    -- Columns returned by the function
    [PersonID] int NOT NULL,
    [FirstName] [nvarchar](50) NULL,
    [LastName] [nvarchar](50) NULL,
    [JobTitle] [nvarchar](50) NULL,
    [BusinessEntityType]
```



```
CREATE FUNCTION [dbo].[ufnGetContactInformation](@PersonID int)
RETURNS @retContactInformation TABLE
(
    -- Columns returned by the function
    [PersonID] int NOT NULL,
    [FirstName] [nvarchar](50) NULL,
    [LastName] [nvarchar](50) NULL,
    [JobTitle] [nvarchar](50) NULL,
    [BusinessEntityType] [nvarchar](50) NULL
)
AS
-- Returns the first name, last name, job title and business entity type for the specified contact.
-- Since a contact can serve multiple roles, more than one row may be returned.
BEGIN
    IF @PersonID IS NOT NULL
        BEGIN
        IF EXISTS(SELECT * FROM [HumanResources].[Employee] e
                    WHERE e.[BusinessEntityID] = @PersonID)
            INSERT INTO @retContactInformation
                SELECT @PersonID, p.FirstName, p.LastName, e.[JobTitle], 'Employee'
                FROM [HumanResources].[Employee] AS e
                    INNER JOIN [Person].[Person] p
                    ON p.[BusinessEntityID] = e.
```



```
                    INNER JOIN [Person].ContactType ct
                    ON ct.[ContactTypeID] = bec.[ContactTypeID]
                    INNER JOIN [Person].[Person] p
                    ON p.[BusinessEntityID] = bec.[PersonID]
                WHERE bec.[PersonID] = @PersonID;

        IF EXISTS(SELECT * FROM [Person].[Person] AS p
                    INNER JOIN [Sales].[Customer] AS c
                    ON c.[PersonID] = p.[BusinessEntityID]
                    WHERE p.[BusinessEntityID] = @PersonID AND c.[StoreID] IS NULL)
            INSERT INTO @retContactInformation
                SELECT @PersonID, p.FirstName, p.LastName, NULL, 'Consumer'
                FROM [Person].[Person] AS p
                    INNER JOIN [Sales].[Customer] AS c
                    ON c.[PersonID] = p.[BusinessEntityID]
                    WHERE p.[BusinessEntityID] = @PersonID AND c.[StoreID] IS NULL;
        END

    RETURN;
END;
```

Recent Activities

Search

ufnGetContactInformation
Show Live Lineage • SQL SERVER

ufnGetContactInformation
Show Live Lineage • SQL SERVER

ufnGetContactInformation
Show Inner System Lineage • SQL SERVER • DATABASE

OBJECTS • Contains
Show Button Data • SQL SERVER • DATABASE

TABLES & VIEWS • Contains
Show Button Data • SQL SERVER • DATABASE

Custom SQL Script
Show Live Lineage • GOOGLE BIG QUERY

« Previous 1 2 3 4 5 ··· 89 Next »

Octomize AI - Lineage Studio empowered by GenAI Copilot

Holistic Data Management: The integration of Live Lineage and Octomize offers a comprehensive solution for automating, optimizing, and interpreting SQL queries, with real-time data lineage visualization.

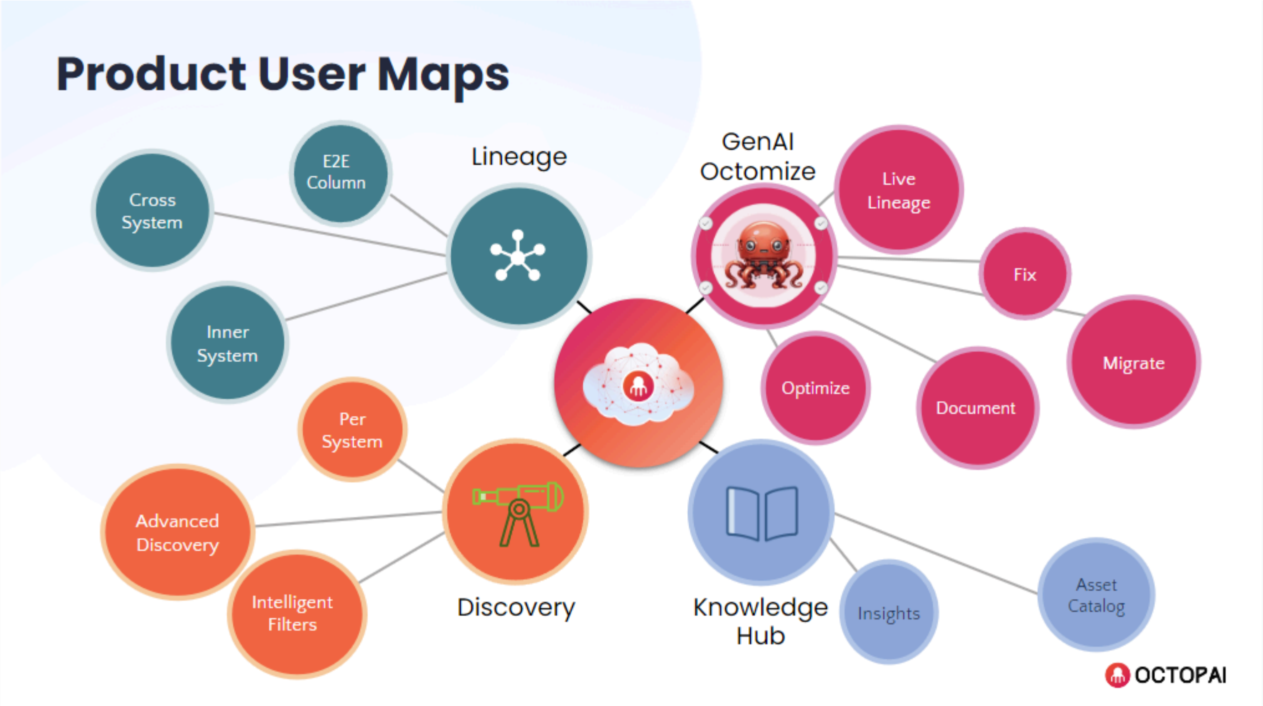
Key Benefits

- **Holistic Data Management**: The integration of **Live Lineage** and **Octomize** offers a comprehensive solution for automating, optimizing, and interpreting SQL queries, with real-time data lineage visualization.
- **Domain-Specific AI**: **GenAI** technology is purpose-built for tackling data domain challenges, delivering AI-driven solutions that are relevant and actionable.
- **Immediate Impact Assessment**: **Live Lineage** enables real-time visualization of data sources, transformations, and targets, effectively mitigating operational risks.
- **Cost Efficiency & Risk Mitigation**: Save on labor and financial resources by avoiding costly errors and compliance issues.
- **Data Democratization**: **Octomize** simplifies complex SQL into understandable language, enabling non-technical stakeholders to participate in data-driven decision-making.
- **Security**: Integrated with Azure OpenAI to ensure a secure workspace for your data operations.
- **Tailored Experience**: Pre-engineered prompts and domain-specific optimizations ensure direct applicability to the challenges faced by data teams.

Feature Capabilities

- **Query Fixing**: Corrects and enhances SQL syntax, raising the quality of your code. When integrated with **Live Lineage's** real-time visualization, the impact is magnified—not only fixing queries but also making data-driven decisions on how corrections affect the entire data ecosystem. This synergy ensures both correctness and optimal data utility, improving efficiency and compliance.
- **Query Optimization**: Optimizes SQL queries, reducing execution times significantly. Combined with **Live Lineage's** dynamic lineage mapping, it offers real-time visibility into the downstream effects of optimizations. This elevates query optimization from a technical tweak to a strategic enhancement.
- **System Migration**: Simplifies script migration by adjusting for compatibility. Supported systems include **SQL Server, Oracle, Teradata, Netezza, Vertica, Snowflake, MySQL, Hive, PostgreSQL, DB2, Redshift, Google Big Query, SAP HANA, Spark, Java, Python**. Migration is supported from and to any of these systems, enabling seamless script conversion across various environments. With **Live Lineage**, you gain insight into how migrations impact the existing data landscape, enabling a proactive, frictionless transition.
- **Query Interpretation**: Translates complex SQL into plain language, democratizing data access across the organization. Coupled with **Live Lineage**, non-technical stakeholders gain context, empowering them to understand and utilize data effectively within the broader data ecosystem.
- **Documentation Generation**: Produces business, technical documentation as well as compliance and security risk assessments, ensuring that all stakeholders have the necessary insights into data operations. This capability drives productivity gains and supports simulations for applying changes—whether it's a script already harvested by Cloudera Octopai or a new one.

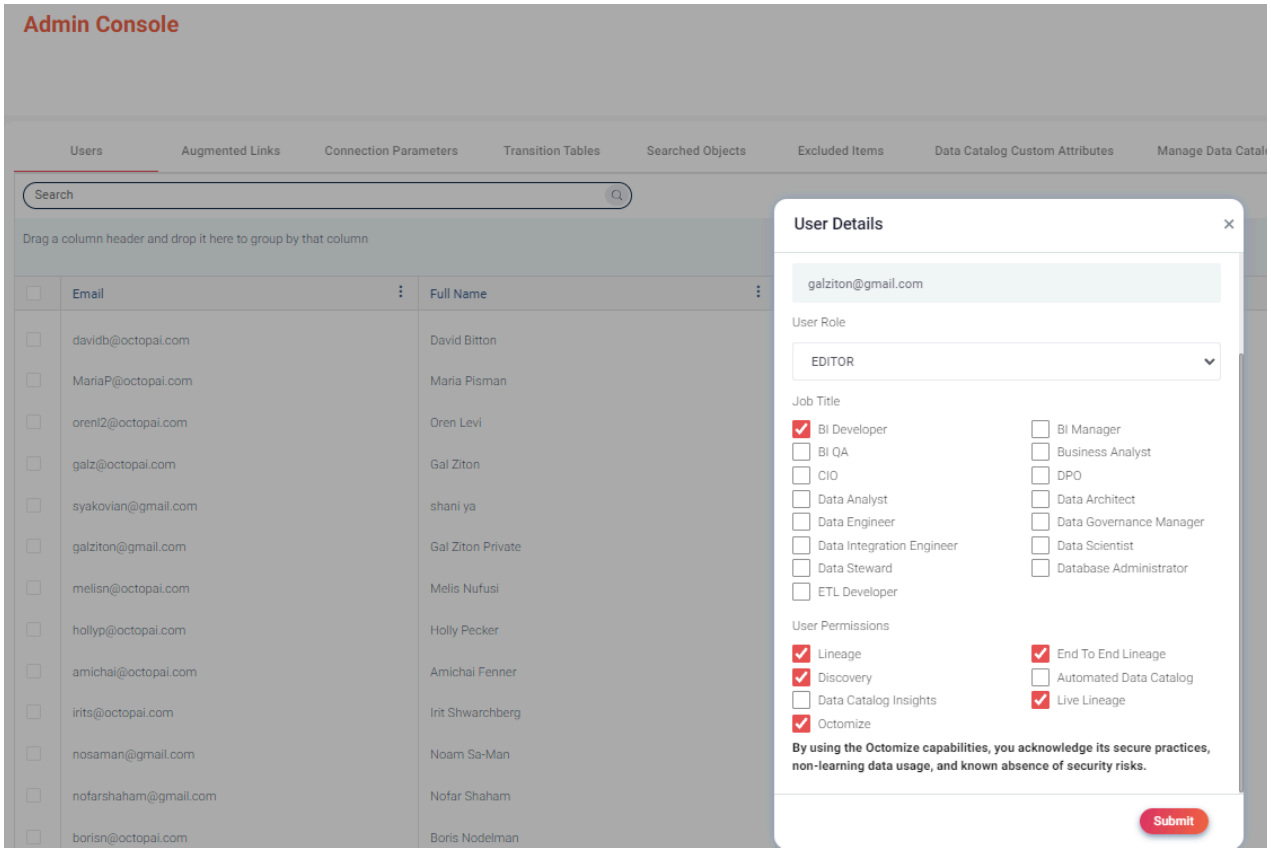
The integration of **Octomize GenAI** Copilot with **Live Lineage** creates a powerful synergy, enhancing efficiency, foresight, and risk mitigation. This is a true "1 + 1 = 3" scenario, where the combined solution exceeds the sum of its parts.



Product User Maps

How to Enable

Contact Support or your Customer Success Manager for activation. Note that **Octomize** requires the **Live Lineage** module to be activated first. Once activated, your Admin can enable it for specific users.



Admin Console

Drag a column header and drop it here to group by that column

□	Email	Full Name
□	davidb@octopai.com	David Bitton
□	MariaP@octopai.com	Maria Pisman
□	orenl2@octopai.com	Oren Levi
□	galz@octopai.com	Gal Ziton
□	syakovian@gmail.com	shani ya
□	galziton@gmail.com	Gal Ziton Private
□	melisn@octopai.com	Melis Nufusi
□	hollyp@octopai.com	Holly Pecker
□	amichai@octopai.com	Amichai Fenner
□	irits@octopai.com	Irit Shwarchberg
□	nosaman@gmail.com	Noam Sa-Man
□	nofarshaham@gmail.com	Nofar Shaham
□	borisn@octopai.com	Boris Nodelman

User Details ×

galziton@gmail.com

User Role

EDITOR

Job Title

☑ BI Developer
□ BI QA
□ CIO
□ Data Analyst
□ Data Engineer
□ Data Integration Engineer
□ Data Steward
□ ETL Developer
□ BI Manager
□ Business Analyst
□ DPO
□ Data Architect
□ Data Governance Manager
□ Data Scientist
□ Database Administrator

User Permissions

☑ Lineage
☑ Discovery
□ Data Catalog Insights
☑ Octomize
☑ End To End Lineage
□ Automated Data Catalog
☑ Live Lineage

By using the Octomize capabilities, you acknowledge its secure practices, non-learning data usage, and known absence of security risks.

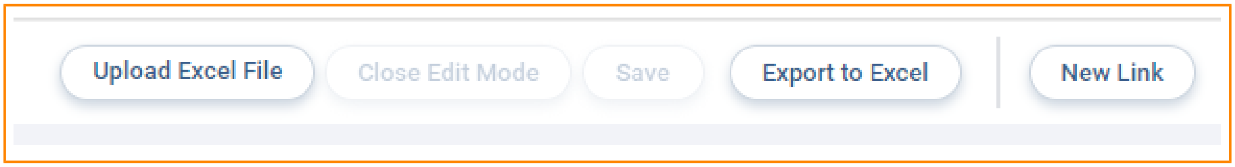
Submit

Augmented Links

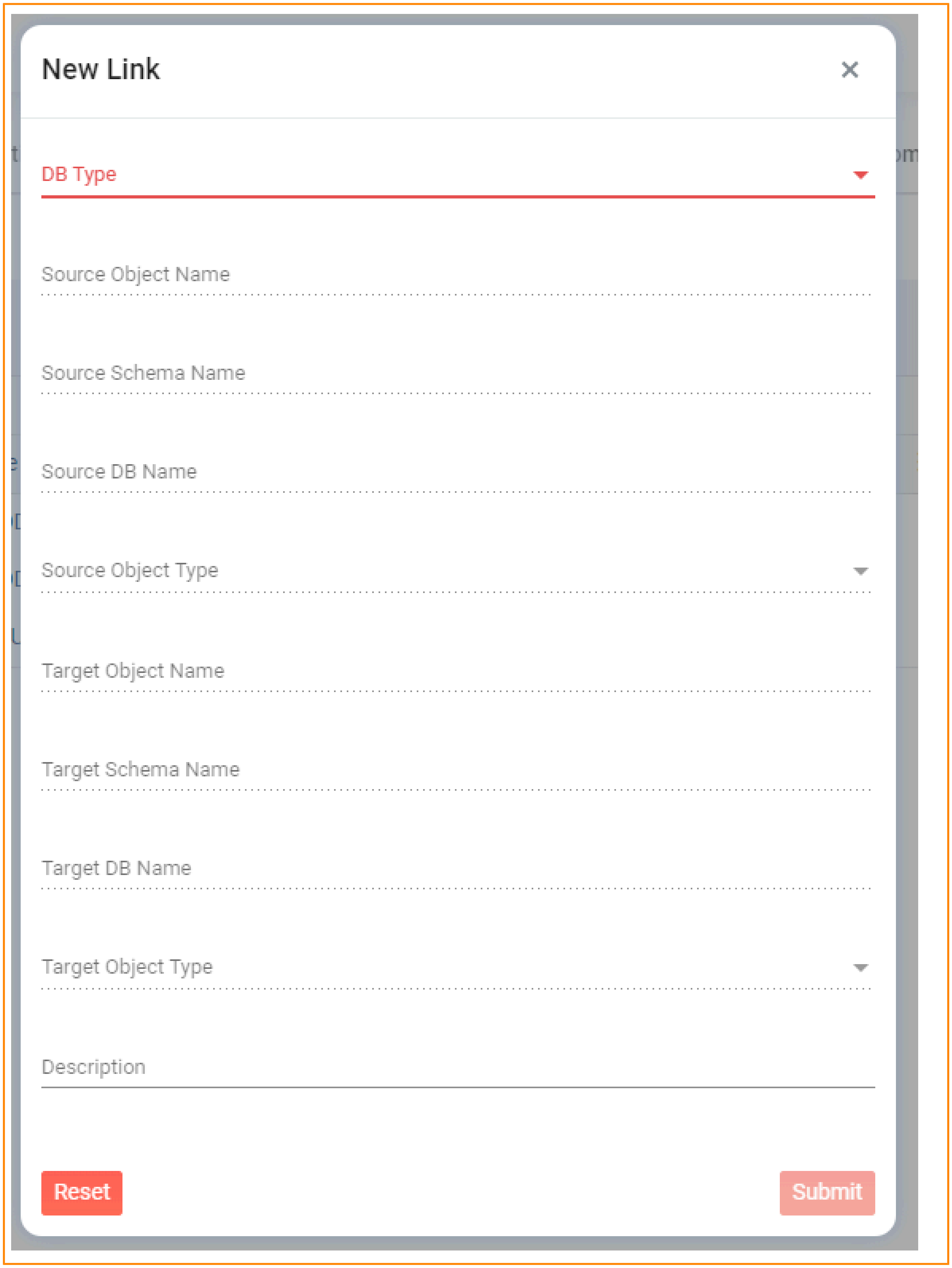
Create manual or bulk augmented lineage links in the Admin Console to connect database objects in Cross System Lineage.

How to Create an Augmented Link

- Go to Admin Console --> Augmented Links



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



New Link

DB Type

Source Object Name

Source Schema Name

Source DB Name

Source Object Type

Target Object Name

Target Schema Name

Target DB Name

Target Object Type

Description

  - 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

Automated Discovery Space - Your Search Engine for Cross-System Data Asset Discovery

Discovery Space in Cloudera Octopai Data Lineage is a high-performance search engine designed for data experts who need instant access to metadata, scripts, and stored procedures across multiple systems. Think of it as Goo...

Find the Data You Need—Instantly and Accurately

Discovery Space in Cloudera Octopai is a high-performance search engine designed for data experts who need instant access to metadata, scripts, and stored procedures across multiple systems. Think of it as Google Search for your data ecosystem, enabling fast, context-aware queries that preserve the native structure, language, and formats of the harvested systems.

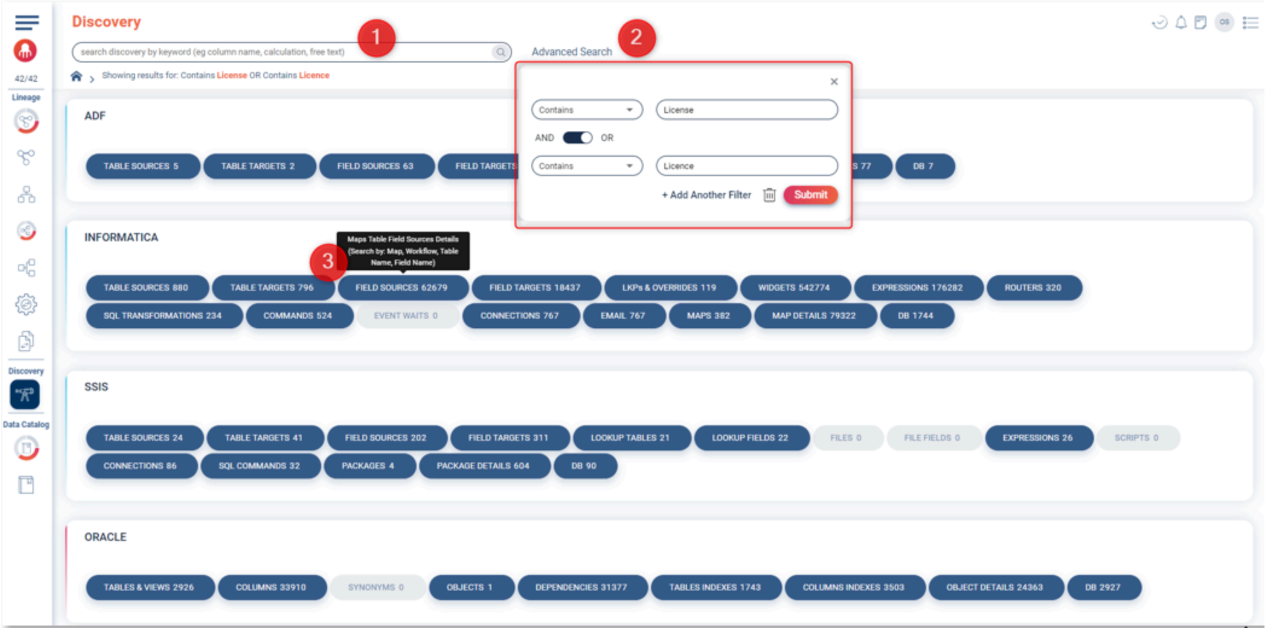
Key Capabilities of Discovery Space:

1. **Comprehensive Metadata Indexing:**
    - Discovery Space represents the full scope of harvested metadata, including:
        - Tables, columns, reports, and dashboards
        - ETL processes, scripts, stored procedures
        - Business logic and dependencies
    - The search engine operates across databases, BI tools, ETL platforms, and cloud environments, ensuring full visibility into your data landscape.
2. **Intelligent Search for Precision:**
    - Supports fuzzy and exact search parameters to narrow results efficiently, allowing users to locate relevant objects without manual filtering.
    - Queries are performed using an intelligent search engine that understands system-specific syntax and maintains the original context of the metadata.
3. **Native Format & Terminology Preservation:**
    - Unlike generic search tools, Discovery Space preserves the structure and terminology of each source system, ensuring that results align with the way data is defined and stored in its native environment.
    - This is critical for impact analysis, root cause investigations, and governance workflows, where context and accuracy are essential.
4. **Downloadable Insights for Impact Analysis:**
    - Discovery Space does not alter or annotate data, but instead enables users to download search results for further scoping and impact assessment.
    - These insights can be cross-referenced in Knowledge Hub, where additional documentation, collaboration, and governance processes can take place.

Optimized for Data Engineers, Analysts, and Governance Teams

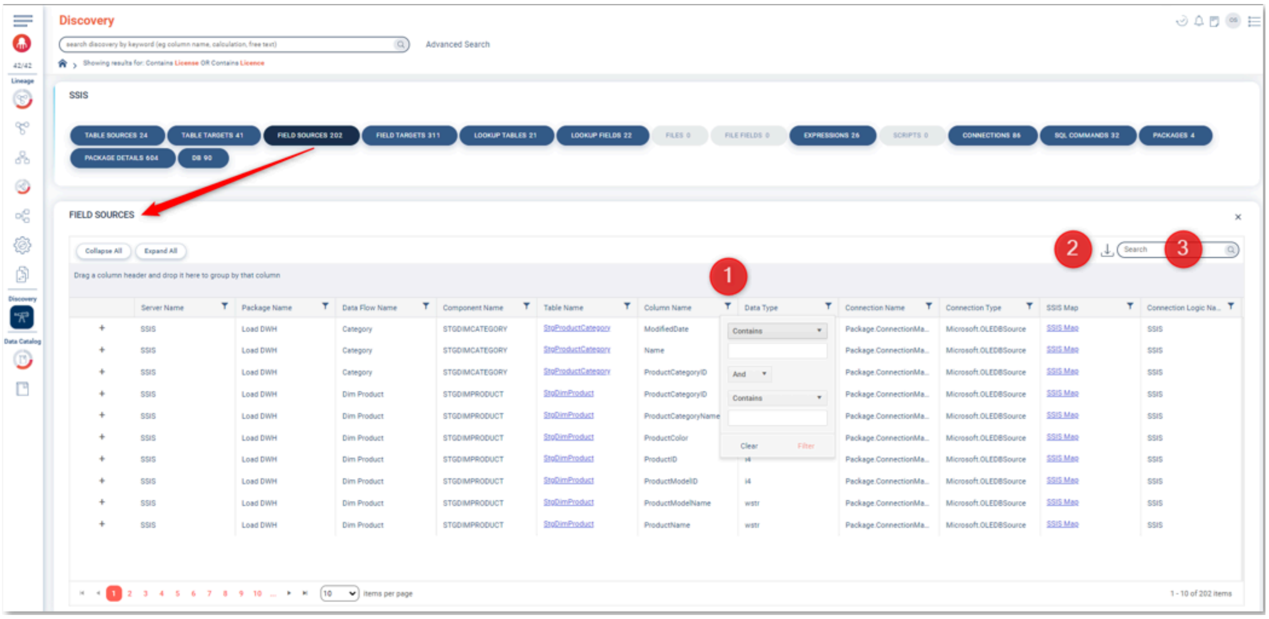
With **Discovery Space** , data professionals can: **Locate critical data assets instantly** across complex, hybrid environments **Analyze metadata dependencies** to assess risks and change impacts **Streamline investigations** with precise, system-specific searches

Systems Section:



1. **Search Box** Searches for any value across the BI landscape
2. **Advance Search** Search values using AND / OR when it comes to a non-conclusive term or searching for more than one value
3. **How Cloudera Octopai Searches the Value** Hover over the tag to know how Cloudera Octopai searches the value

Detailed Level:



	Server Name	Package Name	Data Flow Name	Component Name	Table Name	Column Name	Data Type	Connection Name	Connection Type	SSIS Map	Connection Logic Na...
+	SSIS	Load DWH	Category	STGDIMCATEGORY	StgProductCategory	ModifiedDate		Package.ConnectionMa...	Microsoft.OLEDBSource	SSIS Map	SSIS
+	SSIS	Load DWH	Category	STGDIMCATEGORY	StgProductCategory	Name		Package.ConnectionMa...	Microsoft.OLEDBSource	SSIS Map	SSIS
+	SSIS	Load DWH	Category	STGDIMCATEGORY	StgProductCategory	ProductCategoryID		Package.ConnectionMa...	Microsoft.OLEDBSource	SSIS Map	SSIS
+	SSIS	Load DWH	Dim Product	STGDIMPRODUCT	StgDimProduct	ProductCategoryID		Package.ConnectionMa...	Microsoft.OLEDBSource	SSIS Map	SSIS
+	SSIS	Load DWH	Dim Product	STGDIMPRODUCT	StgDimProduct	ProductCategoryName		Package.ConnectionMa...	Microsoft.OLEDBSource	SSIS Map	SSIS
+	SSIS	Load DWH	Dim Product	STGDIMPRODUCT	StgDimProduct	ProductColor		Package.ConnectionMa...	Microsoft.OLEDBSource	SSIS Map	SSIS
+	SSIS	Load DWH	Dim Product	STGDIMPRODUCT	StgDimProduct	ProductID	i4	Package.ConnectionMa...	Microsoft.OLEDBSource	SSIS Map	SSIS
+	SSIS	Load DWH	Dim Product	STGDIMPRODUCT	StgDimProduct	ProductModelID	i4	Package.ConnectionMa...	Microsoft.OLEDBSource	SSIS Map	SSIS
+	SSIS	Load DWH	Dim Product	STGDIMPRODUCT	StgDimProduct	ProductModelName	wstr	Package.ConnectionMa...	Microsoft.OLEDBSource	SSIS Map	SSIS
+	SSIS	Load DWH	Dim Product	STGDIMPRODUCT	StgDimProduct	ProductName	wstr	Package.ConnectionMa...	Microsoft.OLEDBSource	SSIS Map	SSIS

1. **Column Filter Funnel** Narrow down the search within a column

2. **Export the list to Excel** Use the information for your own purposes like pivot tables, workload & effort planning, etc.
3. **Specific tool search** Helps to search for a specific term within the open tool

Apply Filters on which Metadata Sources you wish to apply the discovery function:

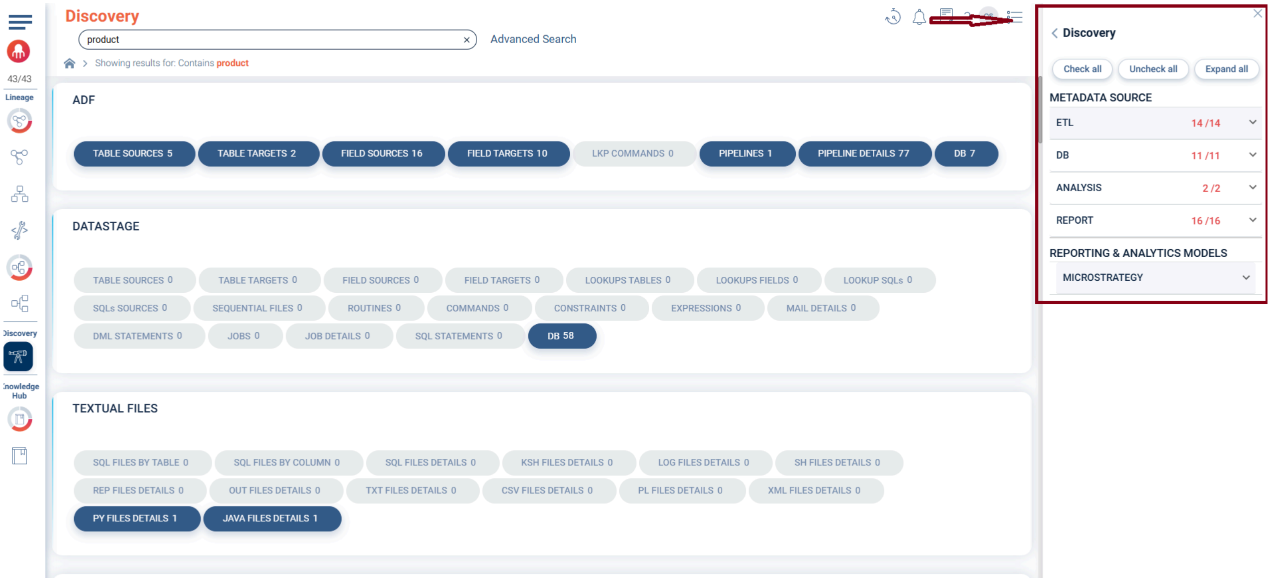


Tableau Intelligent Graph Connector for operational metadata intelligence

Learn about setting up Tableau Intelligent Graph Connector to extract operational metadata for usage, performance and auditing data.

Permissions prerequisites

The following permissions and prerequisites are valid for setting up Tableau Intelligent Graph Connector:

- **API License** – Ensure that the API of Tableau Usage is enabled through your Tableau license. Refer to your Tableau administrator or support if you need help enabling the API.
- **Login Permissions** – Ensure you have login permissions to the Tableau Usage PostgreSQL server.
- **Database Access** – You will need PostgreSQL server and port information, along with a username and password. Ensure that the SELECT permissions are granted for the following public data tables:
  - _sites
  - users
  - views
  - hist_comments
  - hist_views
  - _http_requests
  - historical_events
  - projects
  - hist_users
  - workbooks
  - system_users
  - _sessions

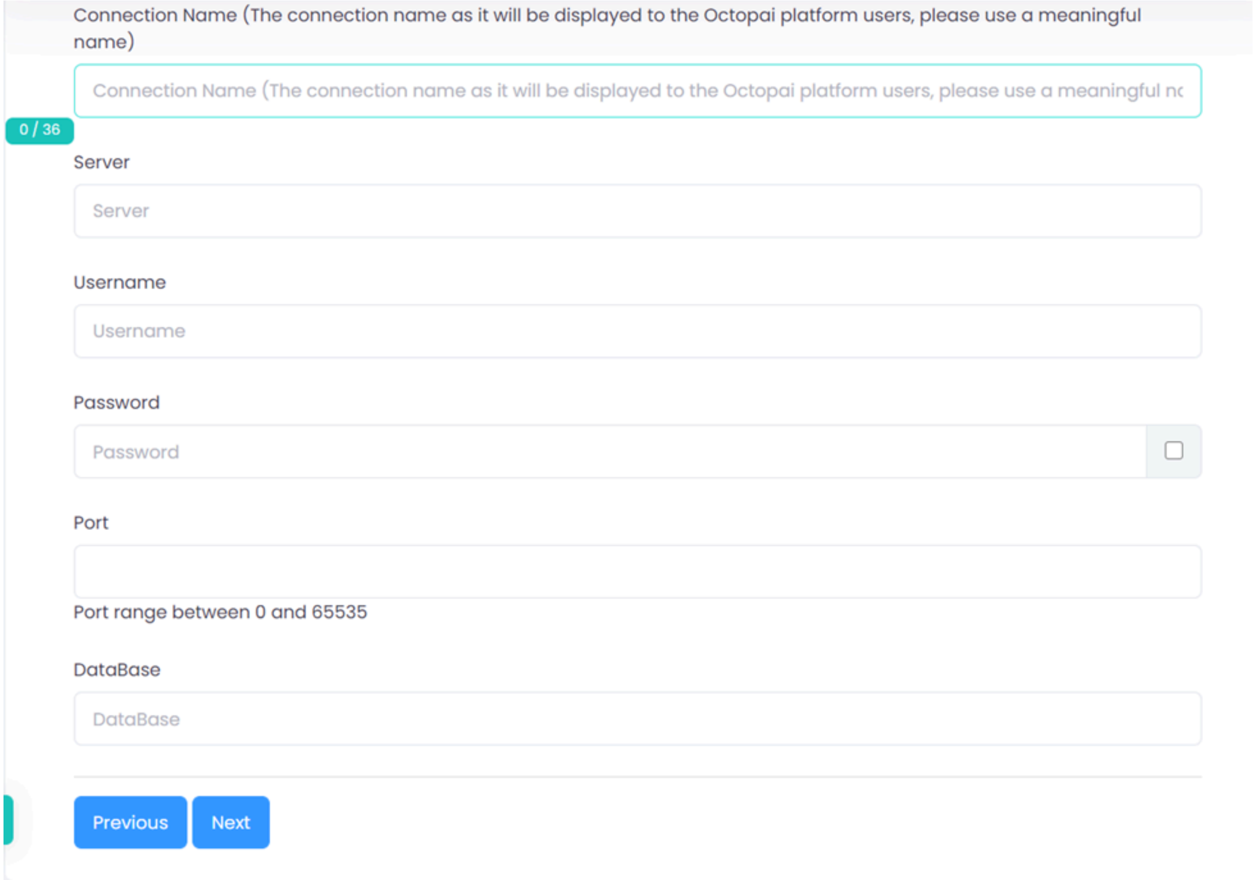
Setting up Tableau Metadata source

To configure Tableau Metadata in Cloudera Octopai:

1. **Open Cloudera Octopai Client.**

   Launch the **Octopai Client** on your system.
2. **Navigate to the metadata source setup.**
   - In the Cloudera Octopai Client interface, go to the **Metadata Sources** section.
   - Locate and select **Tableau** as the metadata source.
3. **Input the server credentials.**
   - Enter the relevant PostgreSQL server details, including **Server** , **Port** , **User** , and **Password** . This information must correspond to the Tableau PostgreSQL database.
   - Ensure you provide the correct login credentials and that the required permissions are in place (as listed in the prerequisites).
4. **Select data tables.**
   - Select the relevant tables for which you have permissions, such as those mentioned in the prerequisites.
5. **Save the configuration.**
   - Once you have input the necessary details, save your configuration. The Cloudera Octopai Client will now have access to Tableau's Operational Metadata.

Figure 7: Tableau metadata in Cloudera Octopai configuration



Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful nc

0 / 36

Server

Server

Username

Username

Password

Password

Port

Port range between 0 and 65535

DataBase

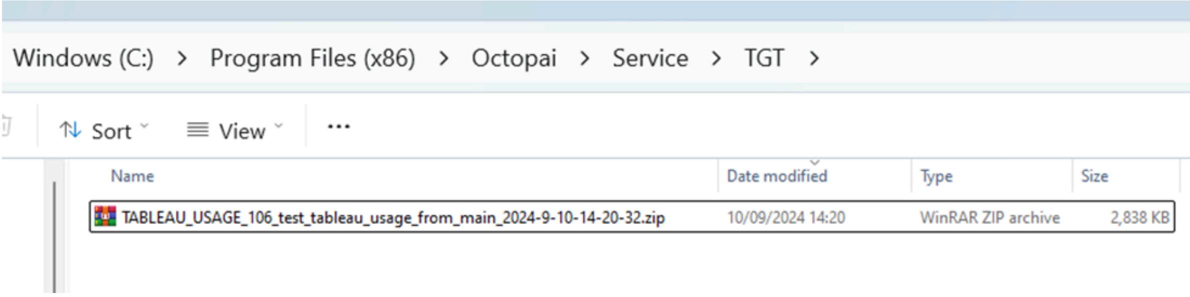
DataBase

Verifying the extracted metadata file

Once the metadata extraction is complete, follow these steps to verify the extracted files:

1. **Access the target folder.**
   - On the server where the **Octopai Client** is installed, navigate to the **Target (TGT) Folder** .
   - **Default Path** : C:\Program Files (x86)\Octopai\Service\TGT
2. **Locate the connector file.**
   - Inside the TGT folder, you will find a **.zip** file named after the **Tableau Connector** .
   - Example: Tableau_Metadata.zip
3. **Open the zip file.**
   - Extract the contents of the **.zip** file.

   Figure 8: Open the ZIP file


Windows (C:) > Program Files (x86) > Octopai > Service > TGT >
Sort
View
...

Name	Date modified	Type	Size
TABLEAU_USAGE_106_test_tableau_usage_from_main_2024-9-10-14-20-32.zip	10/09/2024 14:20	WinRAR ZIP archive	2,838 KB

4. **Verify file contents.**
   - Ensure the quantity of inner files matches what you expect based on your metadata extraction process.
   - Review the **quality** of the files by checking the format and ensuring that the data is consistent with what was pulled from the Tableau server.

Expected files and quality check

- The extracted .zip must contain several files representing different metadata components, for example views, projects, and historical events.
- Ensure that all necessary metadata files are present. Any missing or corrupted files could indicate an issue with the extraction process.

Cloudera Octopai Connector for Apache NiFi

Learn how the Cloudera Octopai Data Lineage connector for Apache NiFi enables visibility into data movement across systems by capturing and visualizing lineage derived from NiFi flows.

Overview

Apache NiFi is a core orchestration platform in modern data architectures, responsible for ingesting, routing, transforming, and delivering data across heterogeneous environments. The Cloudera Octopai Data Lineage connector for Apache NiFi extracts metadata from NiFi flows and constructs lineage that exposes how data moves between systems, technologies, and platforms.

The connector enables the following capabilities:

- Building cross-system, inner system and end-to-end column lineage for NiFi flows.
- Populating the Knowledge Hub assets automatically and enabling users to discover NiFi assets.
- Enabling governance, impact analysis, and operational visibility across enterprise data pipelines.

Why Apache NiFi data lineage matters

NiFi often serves as the integration layer that connects files, databases, object stores, streaming platforms, and cloud services. Understanding these flows is critical for several reasons.

Visibility into data movement

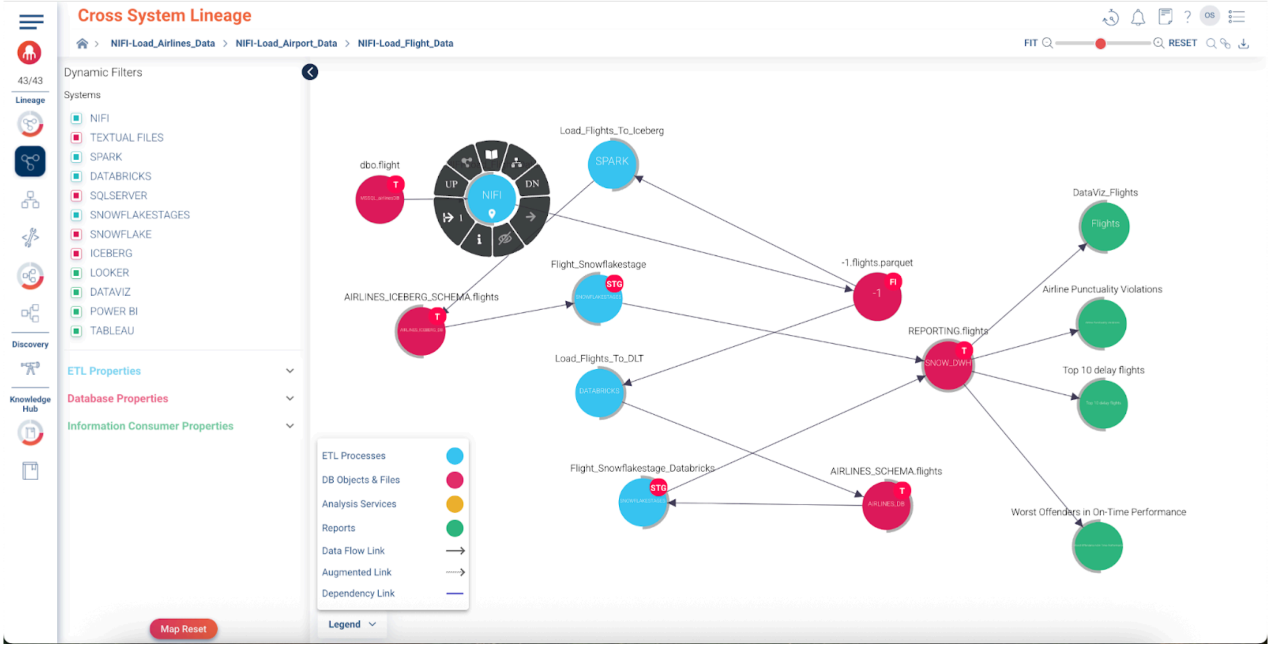
NiFi connects diverse sources and targets. Lineage reveals how data enters, moves through, and exits the platform.

Cross-system complexity

NiFi commonly bridges legacy, hybrid, and cloud environments. Cross-system lineage enables teams to track data as it moves across technologies and organizational boundaries.

Operational insight

Understanding dependencies between systems helps teams troubleshoot failures, assess the impact of change, and reduce risk during migrations or platform modernization initiatives.



Supported NiFi versions

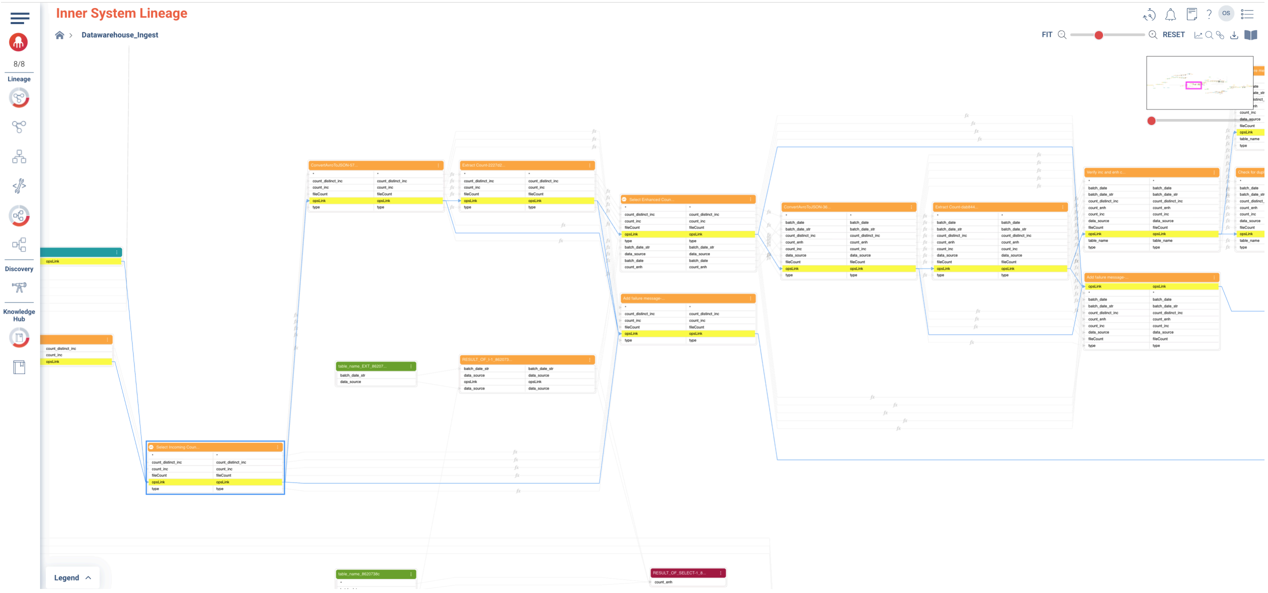
The connector is compatible with Apache NiFi versions 1.2.8 through 2.7.2.

Using supported versions ensures consistent metadata extraction, stable API behavior, and reliable lineage generation.

Lineage model overview

The connector builds lineage in layers, starting with operational flow relationships and extending to system-level source and target context.

Processor-level operational lineage (opsLink)



The foundation of NiFi lineage is the processor-level operational link, referred to as an *opsLink*.

An opsLink represents a direct execution relationship between two NiFi components:

- A source processor and a target processor connected by a NiFi connection.
- InputPorts and OutputPorts are treated as processors for lineage purposes.

opsLinks are derived by parsing the ProcessorGroup configuration JSON and capturing:

- Processors, InputPorts, and OutputPorts.
- Connections between components.
- Associated source or target context when available, such as database, table, topic, bucket, or file location.

This processor-level lineage forms the operational flow graph of a NiFi ProcessGroup and serves as the backbone for cross-system lineage views.

Nested ProcessGroups behavior

When a ProcessorGroup contains other nested ProcessorGroups, inner lineage is scoped strictly to the selected ProcessorGroup:

- Only processors, InputPorts, and OutputPorts that belong directly to the current ProcessorGroup are displayed in the inner lineage view.
- Nested ProcessorGroups are not expanded or traversed as part of inner lineage.

When one ProcessorGroup is connected to another ProcessorGroup:

- The relationship between ProcessorGroups is not shown in inner lineage.
- These relationships are visualized only in end-to-end lineage or cross-system lineage views.

This separation ensures that inner lineage remains focused on execution flow within a single ProcessGroup, while cross-group dependencies are handled at higher-level lineage views.

Supported NiFi components

The connector supports a broad set of NiFi processors for lineage extraction.

For each supported processor, the documentation lists:

- Component Type (FQCN)
- NiFiProcessorType

Table 1: Supported components

Component Type (FQCN)	NiFiProcessorType
org.apache.nifi.processors.standard.QueryDatabaseTable	QueryDatabaseTable
org.apache.nifi.processors.kafka.pubsub.ConsumeKafka_1_0	ConsumeKafka
org.apache.nifi.kafka.processors.ConsumeKafka	ConsumeKafka
org.apache.nifi.processors.kafka.pubsub.ConsumeKafkaRecord_2_6	ConsumeKafkaRecord
org.apache.nifi.processors.kafka.pubsub.PublishKafka_1_0	PublishKafka
org.apache.nifi.processors.kafka.pubsub.PublishKafka_2_0	PublishKafka
org.apache.nifi.processors.kafka.pubsub.PublishKafka_2_6	PublishKafka
org.apache.nifi.processors.kafka.pubsub.PublishKafkaRecord_2_0	PublishKafkaRecord
org.apache.nifi.processors.kafka.pubsub.PublishKafkaRecord_2_6	PublishKafkaRecord
org.apache.nifi.processors.kudu.PutKudu	PutKudu
org.apache.nifi.processors.standard.FlattenJson	FlattenJson
org.apache.nifi.processors.attributes.UpdateAttribute	UpdateAttribute
org.apache.nifi.processors.aws.s3.PutS3Object	PutS3Object
org.apache.nifi.processors.standard.PutSQL	PutSQL
org.apache.nifi.processors.standard.RouteOnAttribute	RouteOnAttribute
org.apache.nifi.processors.standard.RouteOnContent	RouteOnContent
org.apache.nifi.processors.aws.s3.ListS3	ListS3
org.apache.nifi.processors.standard.ExecuteStreamCommand	ExecuteStreamCommand
org.apache.nifi.processors.standard.ReplaceText	ReplaceText
org.apache.nifi.processors.hadoop.DeleteHDFS	DeleteHDFS
org.apache.nifi.processors.standard.GenerateFlowFile	GenerateFlowFile
org.apache.nifi.processors.aws.s3.FetchS3Object	FetchS3Object
org.apache.nifi.processors.parquet.PutParquet	PutParquet
org.apache.nifi.processors.standard.InvokeHTTP	InvokeHTTP
org.apache.nifi.processors.standard.GenerateTableFetch	GenerateTableFetch
org.apache.nifi.processors.standard.ConvertRecord	ConvertRecord
org.apache.nifi.processors.hadoop.FetchHDFS	FetchHDFS
org.apache.nifi.processors.standard.EvaluateJsonPath	EvaluateJsonPath
org.apache.nifi.csv.CSVReader	CSVReader
org.apache.nifi.processors.standard.AttributesToJSON	AttributesToJSON
org.apache.nifi.processors.standard.ExecuteSQL	ExecuteSQL
org.apache.nifi.processors.standard.ExecuteScript	ExecuteScript
org.apache.nifi.processors.script.ExecuteScript	ExecuteScript
org.apache.nifi.processors.standard.LogMessage	LogMessage
org.apache.nifi.processors.standard.ExecuteSQLRecord	ExecuteSQLRecord
org.apache.nifi.processors.hive.PutHive3QL	PutHive3QL
org.apache.nifi.processors.office.ConvertExcelToCSVProcessor	ConvertExcelToCSVProcessor
org.apache.nifi.processors.cdp.objectstore.PutCDPObjectStore	PutCDPObjectStore

Component Type (FQCN)	NiFiProcessorType
org.apache.nifi.processors.cdp.objectstore.DeleteCDPObjectStore	DeleteCDPObjectStore
org.apache.nifi.csv.CSVRecordSetWriter	CSVRecordSetWriter
org.apache.nifi.processors.standard.MergeContent	MergeContent
org.apache.nifi.processors.standard.QueryRecord	QueryRecord
org.apache.nifi.processors.standard.ExtractText	ExtractText
org.apache.nifi.dbcp.DBCPConnectionPool	DBCPConnectionPool
org.apache.nifi.processors.standard.DistributeLoad	DistributeLoad
org.apache.nifi.processors.standard.SplitJson	SplitJson
org.apache.nifi.processors.standard.JoltTransformJSON	JoltTransformJSON
org.apache.nifi.processors.standard.JoltTransformRecord	JoltTransformRecord
org.apache.nifi.processors.mongodb.GetMongo	GetMongo
org.apache.nifi.processors.mongodb.PutMongo	PutMongo
org.apache.nifi.processors.hive.SelectHive3QL	SelectHive3QL
org.apache.nifi.processors.standard.SplitText	SplitText
org.apache.nifi.processors.standard.FetchSFTP	FetchSFTP
org.apache.nifi.processors.avro.ConvertAvroToJSON	ConvertAvroToJSON
org.apache.nifi.processors.standard.UpdateRecord	UpdateRecord
org.apache.nifi.processors.parquet.ConvertAvroToParquet	ConvertAvroToParquet
org.apache.nifi.processors.enrich.JoinEnrichment	JoinEnrichment
org.apache.nifi.processors.enrich.ForkEnrichment	ForkEnrichment
org.apache.nifi.processors.standard.MergeRecord	MergeRecord
org.apache.nifi.processors.standard.InferAvroSchema	InferAvroSchema
org.apache.nifi.processors.kite.InferAvroSchema	InferAvroSchema
org.apache.nifi.processors.aws.s3.DeleteS3Object	DeleteS3Object
org.apache.nifi.processors.standard.SplitRecord	SplitRecord
org.apache.nifi.processors.standard.ConvertJSONToAvro	ConvertJSONToAvro
org.apache.nifi.processors.standard.PutSFTP	PutSFTP
org.apache.nifi.processors.hadoop.PutHDFS	PutHDFS
org.apache.nifi.processors.standard.PutDatabaseRecord	PutDatabaseRecord
org.apache.nifi.processors.standard.ConvertJSONToSQL	ConvertJSONToSQL
org.apache.nifi.processors.iceberg.PutIceberg	PutIceberg
org.apache.nifi.processors.standard.PutFile	PutFile
com.demoulas.nifi.processors.GetSFTPFileInfo	GetSFTPFileInfo
com.demoulas.nifi.processors.MoveSFTP	MoveSFTP
org.apache.nifi.processors.standard.CompressContent	CompressContent
org.apache.nifi.processors.standard.GetSFTP	GetSFTP
org.apache.nifi.processors.standard.Notify	Notify
org.apache.nifi.processors.standard.RetryFlowFile	RetryFlowFile
org.apache.nifi.processors.standard.Wait	Wait

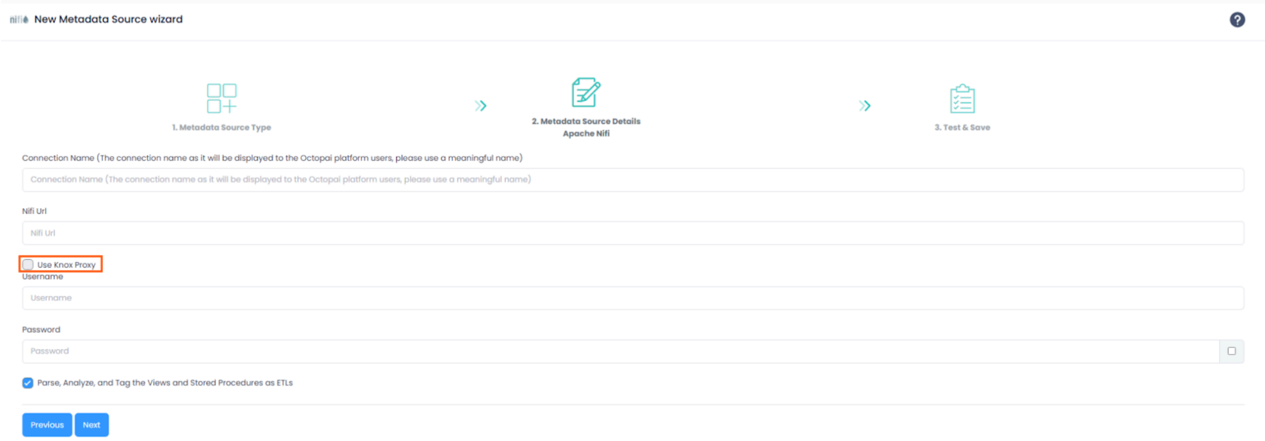
Note: Processors that are not listed in the table are handled using basic pass-through logic.

Knox configuration

The connector supports environments where NiFi is deployed behind an Apache Knox Gateway.

Use Knox Proxy disabled (default)

Figure 9: Use Knox Proxy option in the New Metadata Source wizard



Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Nifi Url

Nifi Url

□ Use Knox Proxy

Username

Username

Password

Password

☑ Parse, Analyze, and Tag the Views and Stored Procedures as ETLs

If Use Knox Proxy is unchecked in the New Metadata Source wizard:

- Authentication uses NiFi native token-based authentication.
- The extractor sends credentials to the NiFi API endpoint: POST /nifi-api/access/token
- NiFi returns a JWT bearer token.
- Subsequent API requests use the token in the Authorization header.

Use Knox Proxy enabled

If Use Knox Proxy is checked in the New Metadata Source wizard:

- Authentication uses HTTP Basic Authentication through Knox.
- Credentials are sent with each request in the Authorization header.
- Knox validates credentials and forwards authenticated requests to NiFi.
- No token exchange is performed.

When to use Knox Proxy

Use Knox Proxy when:

- NiFi is accessed through an Apache Knox Gateway.
- Authentication is centrally managed by Knox.
- The NiFi API is exposed through a Knox proxy URL.

Do not use Knox Proxy when:

- Connecting directly to NiFi without a gateway.
- Using NiFi native authentication.
- The NiFi URL points directly to the NiFi server.

Limitations

The following limitations apply:

- Dynamic parameters embedded inside table names or query identifiers, such as ${db.table.fullname}, are not resolved.
- Site-to-site connections are not currently supported.

Installation and setup

Installation and enablement are performed as part of Cloudera environment configuration.

For assistance with configuration or enablement, contact your Cloudera representative or Cloudera Octopai Data Lineage Support.

Roadmap direction

The connector will continue to evolve with enhancements including:

- Expanded processor coverage.

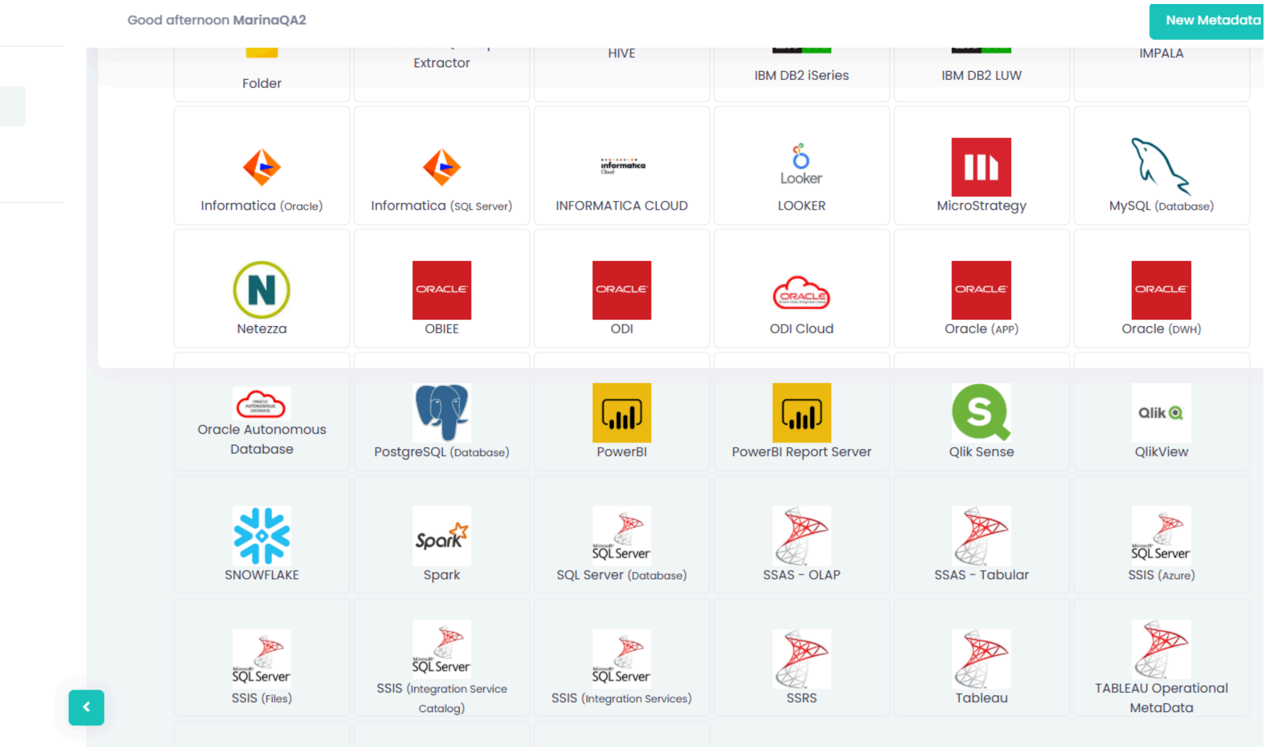
Configuring Cloudera Octopai Connector for Apache Spark

Learn about installing and configuring the Spline-based Cloudera Octopai Data Lineage Connector for Apache Spark to capture automated metadata lineage.

About this task

License requirement: Ensure Spark is included in your Cloudera Octopai subscription before proceeding.

Figure 10: Cloudera Octopai Connector architecture overview



Supported capabilities

The connector captures lineage in the following scenarios:

- **Spline agent lineage** – Lineage capture is limited to what the Spline agent can parse from Spark SQL execution plans.
- **Active jobs** – Only running or newly executed jobs are collected.
- **Persistent actions** – Read and write operations that touch persistent storage (tables or files) are recorded.
- **Cluster configuration** – Spark must be configured with the Spline properties in spark-defaults.conf.
- **Explicit application name** – For job names to appear, jobs must define the application name explicitly:

  ```
  spark = SparkSession.builder \
      .appName("Spark UDF Example") \
      .getOrCreate()
  ```
- **Customer-managed environments** – Spark clusters are deployed and managed within the customer environment.

Limitations

The following constraints apply to the connector:

- **Successful jobs only** – Lineage is generated for jobs that finish without errors.
- **Persistent storage focus** – Operations that remain in-memory are excluded from lineage capture.
- **Named jobs required** – Jobs without an explicit name produce lineage records without a meaningful identifier.
- **Kerberos support** – Kerberos and delegation tokens are not yet supported; use basic authentication when sending lineage to the Spline server.
- **Spline parsing scope** – Only Spark operations that Spline supports will appear in lineage.
- **Streaming jobs** – Spark Structured Streaming workloads (for example, Kafka flows) are not captured.
- **Partial execution** – Only code paths that are executed (for example, a conditional branch that runs) appear in lineage.
- **User-defined functions** – UDF logic is not parsed, although their invocation appears in the execution plan.

Before you begin

Before starting the installation, ensure the following:

- A running Spark Cluster (Spark 2.x or 3.x)
- Access to HDFS for storing lineage files
- Cloudera Manager or similar access to configure Spark cluster properties
- Access permissions to upload JAR files to HDFS and edit Spark configurations

Note: Lineage is captured only for data operations that persist results to storage. DataFrames kept solely in memory do not generate lineage. When intermediate DataFrames are not written to a persistent target, lineage can appear incomplete.

Procedure

1. Clone the Cloudera Octopai customized Spline Agent.

   Clone the repository from the Cloudera Octopai customized branch:

   ```
   git clone https://github.com/OCTOPAILTD/spline-spark-agent.git
   cd spline-spark-agent
   git checkout OCT-27187_Enable_writing_to_files
   ```

2. Build the Spline Agent bundle.

   Choose the relevant folder according to your Spark version and navigate to it. The Jar is shipped with the Cloudera Octopai Agent and the user needs to upload the correct jar according to its spark version.

   For example, for Spark 3.5:

   ```
   PS C:\GIT\spline-spark-agent\bundle-3.5> mvn clean package
   ```

   After the build, you will find the Spline Agent JAR file under:

   ```
   bundle-3.5/target/spark-3.5-spline-agent-bundle_2.12-2.2.1.jar
   ```

3. Upload the JAR to HDFS.

   Upload the built Spline Agent JAR file to your HDFS /tmp folder:

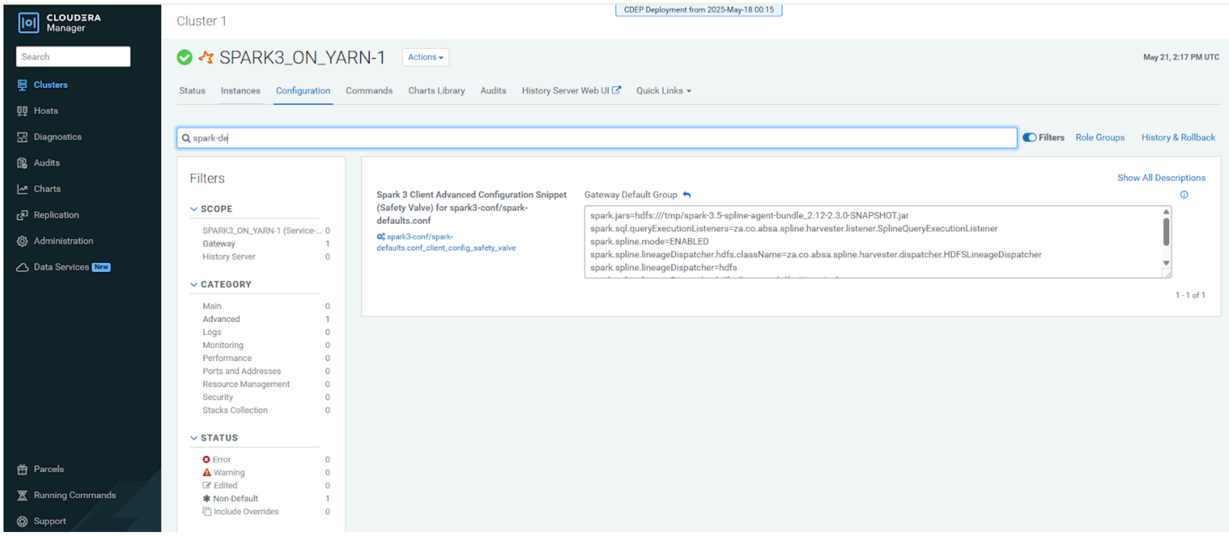
   ```
   hdfs dfs -put spark-3.5-spline-agent-bundle_2.12-2.2.1.jar /tmp/
   ```

4. Configure Spark defaults.

   Add the following properties to your Spark cluster configuration (**spark-defaults.conf**) through Cloudera Manager or equivalent:

   ```
   spark.jars=hdfs:///tmp/spark-3.5-spline-agent-bundle_2.12-2.2.1.jar
   spark.sql.queryExecutionListeners=za.co.absa.spline.harvester.listener.S
   plineQueryExecutionListener
   spark.spline.mode=ENABLED
   spark.spline.lineageDispatcher=hdfs
   spark.spline.lineageDispatcher.hdfs.className=za.co.absa.spline.harvester.
   dispatcher.HDFSLineageDispatcher
   spark.spline.lineageDispatcher.hdfs.directory=hdfs:///tmp/spline
   spark.driver.memory=4g
   ```

   Figure 11: Sample Spark defaults configuration



Spark 3 Client Advanced Configuration Snippet (Safety Valve) for spark3-conf/spark-defaults.conf
spark3-conf/spark-defaults.conf_client_config_safety_valve

Gateway Default Group

```
spark.jars=hdfs:///tmp/spark-3.5-spline-agent-bundle_2.12-2.3.0-SNAPSHOT.jar
spark.sql.queryExecutionListeners=za.co.absa.spline.harvester.listener.SplineQueryExecutionListener
spark.spline.mode=ENABLED
spark.spline.lineageDispatcher.hdfs.className=za.co.absa.spline.harvester.dispatcher.HDFSLineageDispatcher
spark.spline.lineageDispatcher=hdfs
```

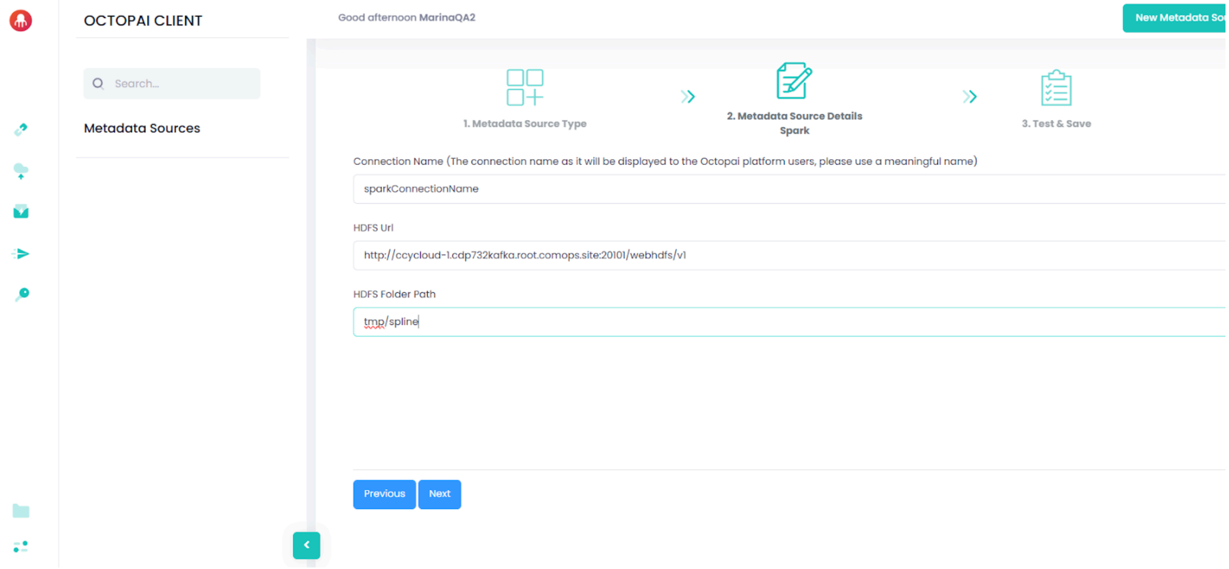
5. Create the HDFS lineage directory.

   Create the directory where lineage files will be written and set permissions:

   ```
   hdfs dfs -mkdir /tmp/spline
   ```

```
hdfs dfs -chown hive /tmp/spline
```

Figure 12: HDFS directory creation example



Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

sparkConnectionName

HDFS Url

http://ccycloud-1.cdp732kafka.root.comops.site:20101/webhdfs/v1

HDFS Folder Path

tmp/spline

6. Set permissions.

   Ensure the user running the Spark jobs has permission to write lineage files to /tmp/spline.

   This can typically be done by ensuring the Spark job runs under a user who has write access to /tmp/spline in HDFS.

What to do next

After completing the installation, verify the following:

- Spline Agent JAR is built and uploaded to HDFS.
- Spark cluster configuration is updated with Spline properties.
- /tmp/spline folder is created and write-access is configured.
- Spark cluster is restarted or configuration is refreshed.
- Test Spark jobs are producing lineage files in /tmp/spline.

⚠ **Important:**

- Ensure the Spline Agent bundle matches your **Spark** and **Scala** versions.
- **Only successful jobs** with **persistent outputs** will generate lineage.
- If no lineage appears, verify:
  - The job reads/writes from/to persistent sources.
  - The Spark job includes the correct configuration parameters. Check the Spark logs.
  - The JAR file was correctly uploaded to HDFS and accessible.

Cloudera Octopai Connector for Apache Spark — Installation and Setup Guide

Learn how to install and configure the Cloudera Octopai Connector for Apache Spark, based on Spline technology, to enable automated metadata extraction and lineage tracking.

About this task

The Cloudera Octopai Connector for Apache Spark supports both non-secured Spark clusters and Spark clusters secured with Kerberos authentication.

In Kerberos-secured environments, lineage capture relies on WebHDFS delegation tokens acquired by the Cloudera Octopai Client running on Windows. Kerberos authentication and delegation token acquisition occur within a Linux environment running on Windows Subsystem for Linux (WSL). This Linux layer is required because the Hadoop and WebHDFS security tools necessary for delegation tokens are available only in Linux.

The Cloudera Octopai Client, running on Windows, orchestrates the process but relies on Linux (through WSL) to authenticate with the Kerberos KDC and securely access HDFS.

Before you begin

Prerequisites

Before starting the installation, ensure the following:

- Spark must be included in your Cloudera Octopai license.
- A running Spark Cluster (Spark 2.x or 3.x).
- Access to HDFS for storing lineage files generated by the Spline agent.
- Cloudera Manager or similar access to configure Spark cluster properties.
- Access permissions to upload jar files to HDFS and edit Spark configuration files, such as spark-defaults.conf.
- Cloudera Manager or similar administrative access to configure Spark cluster properties.

Additional prerequisites for Kerberos-secured environments only

Kerberos-secured Spark clusters require additional client-side setup to enable secure HDFS access.

Windows Requirements (Cloudera Octopai Client)

- A Windows host for running the Cloudera Octopai Client service. This host acts as the control plane for lineage ingestion.
- MIT Kerberos for Windows installed, version 4.1 or later. This provides Windows-side Kerberos tooling and ticket validation.
- A Kerberos configuration file (krb5.ini) provided by the customer. This file defines realms, KDCs, and domain mappings.
- A Spark service principal keytab stored securely on the Windows host. This keytab is used to authenticate non-interactively against the Kerberos KDC.
- Network access from the Windows host to the following:
  - Port 88 (TCP and UDP) for Kerberos KDC access.
  - Port 749 (TCP) for Kerberos Admin Server access.
  - The WebHDFS endpoint on the configured port (for example, port 20101).

Linux Requirements (WSL Ubuntu)

- A Linux environment is mandatory for Kerberos-secured Spark clusters.
- WSL enabled, with Ubuntu installed. This Linux environment runs alongside Windows and is used exclusively for authentication and token handling.

- Kerberos utilities installed inside Ubuntu, including:
    - krb5-user
    - curl
    - jq
- A Linux Kerberos configuration file located at /etc/krb5.conf. This file is copied from the krb5.ini file on Windows to ensure consistent realm configuration.
- Ability to run Kerberos commands such as kinit and klist inside the Linux environment. These commands are used to authenticate, validate tickets, and troubleshoot authentication issues.
- Linux is required because Hadoop WebHDFS delegation tokens cannot be generated using Windows-native tooling. The Linux Kerberos and Hadoop security stack is required to securely acquire and manage these tokens.

About this task

What is Supported

1. **Lineage based on the Spline agent:** Lineage capture is based on what Spline is capable of parsing from Spark SQL execution plans.
2. Running jobs only: Only currently running or newly executed jobs are captured for lineage.
3. Persistent actions only: Actions that involve reading from or writing to persistent storage (such as tables or files) are captured.
4. Cluster configuration: Spark cluster must be configured to include Spline-specific properties in spark-defaults.conf.
5. Explicit Application Name: For job names to appear, jobs must define the application name explicitly:

```
spark = SparkSession.builder \
                        .appName("Spark UDF Example") \
                        .getOrCreate()
```

6. Deployment inside Customer Environment: Spark clusters must be deployed and managed by the customer.
7. Authentication modes: Non-secured Spark clusters, using standard HDFS authentication, are supported without additional setup. Kerberos-secured Spark clusters are supported using WebHDFS delegation tokens acquired by the Cloudera Octopai Client through Linux running on WSL.

About this task

Limitations

1. Lineage for Successfully Completed Jobs Only: Lineage is captured only for Spark jobs that complete successfully.
2. Persistent Storage Only: Only persistent read/write actions to tables or file systems are captured. In-memory DataFrame operations that are not written to storage are not captured.
3. Explicit Job Naming Required: If no job name is explicitly set in the Spark code, the lineage record will not include a meaningful job name.
4. Kerberos Authentication Scope:
    - Kerberos authentication is supported through delegation tokens.
    - Direct Kerberos authentication inside Spark executors is not supported.
    - Keytabs and Kerberos credentials are not deployed into Spark containers.
5. Limited Parsing Scope: Only operations that Spline supports and can parse from the Spark SQL execution plan are included in lineage.
6. Streaming Jobs Not Supported: Spark Structured Streaming (e.g., Kafka read/write) is not captured.
7. Partial Code Execution: Only the parts of the code that are executed (e.g., code within true condition branches) will produce lineage.
8. UDFs: User-Defined Functions (UDFs) are not parsed. Their usage appears in the plan, but the internal logic is not captured.

About this task

Important Usage Note

Important: Lineage is captured only for data operations that are persisted to storage. DataFrames that are not saved to persistent storage will not have lineage captured. As a result, there may be cases where the lineage appears incomplete or broken if intermediate DataFrames are used in-memory without being written to a persistent target.

In Kerberos-secured environments, successful lineage generation depends on the Cloudera Octopai Client successfully acquiring a valid WebHDFS delegation token. Failure to acquire a token or token expiration will prevent the Spark job from writing lineage files.

Procedure

1. **1.** Clone the Cloudera Octopai Customized Spline Agent

   Clone the repository from the Cloudera Octopai customized branch:

   ```
   git clone https://github.com/OCTOPAILTD/spline-spark-agent.git
                           cd spline-spark-agent
                           git checkout OCT-27187_Enable_writing_to_files
   ```

2. **2.** Build the Spline Agent Bundle

   Choose the relevant folder according to your Spark version and navigate to it. The jar is shipped with the Cloudera Octopai Agent and the user needs to upload the correct jar according to its spark version.

   For example, for Spark 3.5:

   ```
   PS C:\GIT\spline-spark-agent\bundle-3.5> mvn clean package
   ```

   After the build, you will find the Spline Agent jar file under:

   ```
   bundle-3.5/target/spark-3.5-spline-agent-bundle_2.12-2.2.1.jar
   ```

3. **3.** Upload the jar to HDFS

   Upload the built Spline Agent jar file to your HDFS /tmp folder:

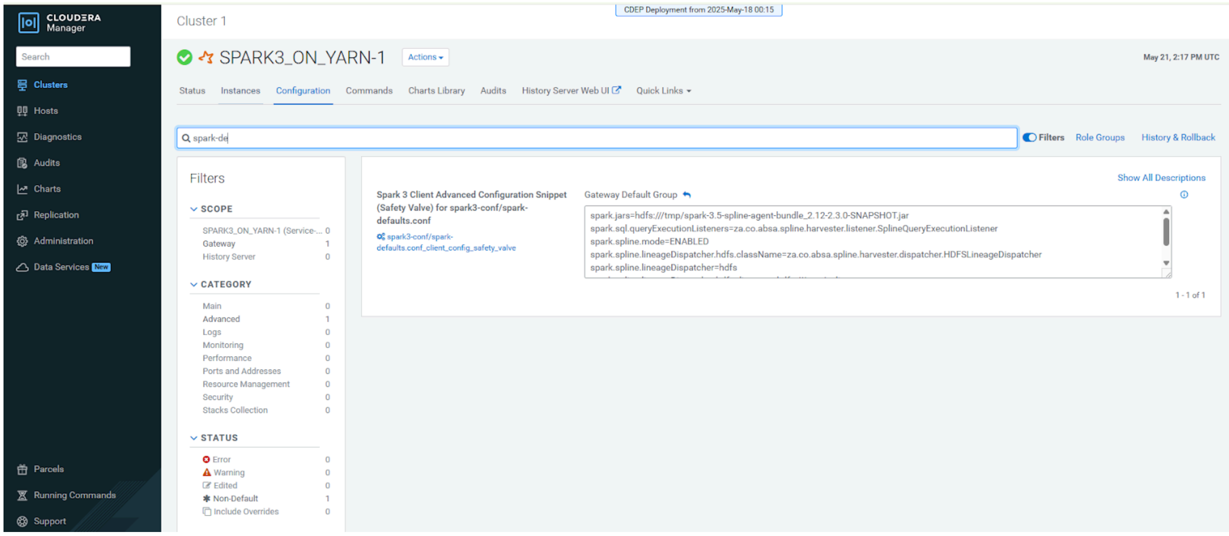
   ```
   hdfs dfs -put spark-3.5-spline-agent-bundle_2.12-2.2.1.jar /tmp/
   ```

4. **4.** Configure Spark Defaults

   Add the following properties to your Spark cluster configuration (spark-defaults.conf) through Cloudera Manager or equivalent:

   ```
   spark.jars=hdfs:///tmp/spark-3.5-spline-agent-bundle_2.12-2.2.1.jar
                           spark.sql.queryExecutionListeners=za.co.absa.spl
   ine.harvester.listener.SplineQueryExecutionListener
                           spark.spline.mode=ENABLED
                           spark.spline.lineageDispatcher=hdfs
                           spark.spline.lineageDispatcher.hdfs.className=za.c
   o.absa.spline.harvester.dispatcher.HDFSLineageDispatcher
                           spark.spline.lineageDispatcher.hdfs.directory=h
   dfs:///tmp/spline
   ```

```
spark.driver.memory=4g
```



Spark 3 Client Advanced Configuration Snippet (Safety Valve) for spark3-conf/spark-defaults.conf

spark3-conf/spark-defaults.conf_client_config_safety_valve

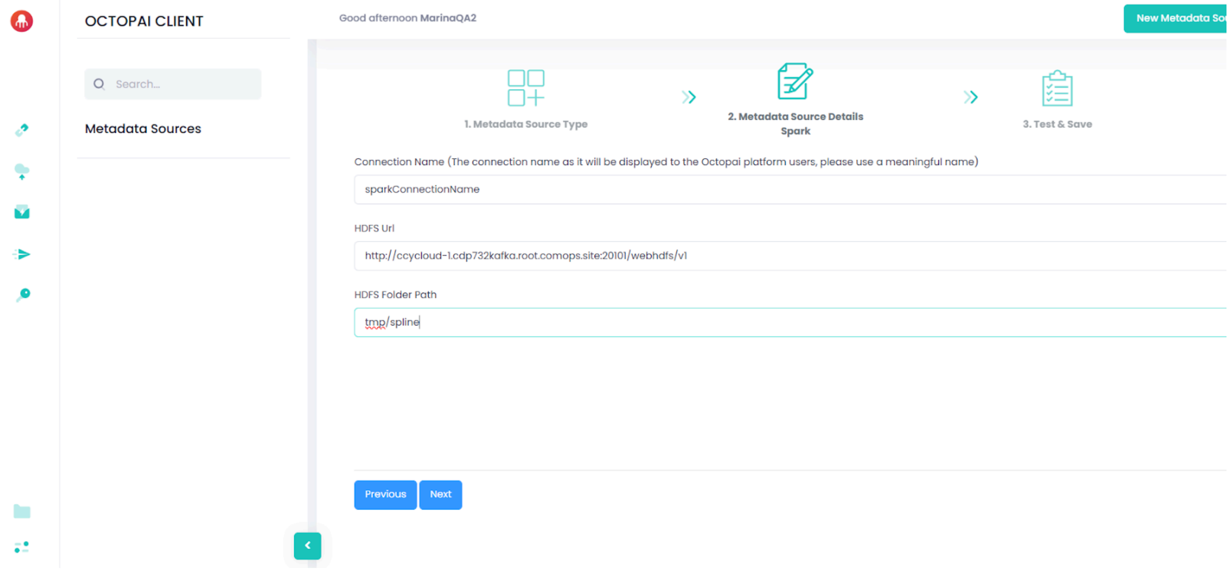
Gateway Default Group

```
spark.jars=hdfs:///tmp/spark-3.5-spline-agent-bundle_2.12-2.3.0-SNAPSHOT.jar
spark.sql.queryExecutionListeners=za.co.absa.spline.harvester.listener.SplineQueryExecutionListener
spark.spline.mode=ENABLED
spark.spline.lineageDispatcher.hdfs.className=za.co.absa.spline.harvester.dispatcher.HDFSLineageDispatcher
spark.spline.lineageDispatcher=hdfs
```

5. Create the HDFS Lineage Directory

Create the directory where lineage files will be written and set permissions:

```
hdfs dfs -mkdir /tmp/spline
                                              hdfs dfs -chown hive /tmp/spline
```



Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

sparkConnectionName

HDFS Url

http://ccycloud-1.cdp732kafka.root.comops.site:20101/webhdfs/v1

HDFS Folder Path

tmp/spline

6. Set Permissions

Ensure the user running the Spark jobs has permission to write lineage files to /tmp/spline.

This can typically be done by ensuring the Spark job runs under a user who has write access to /tmp/spline in HDFS.

What to do next

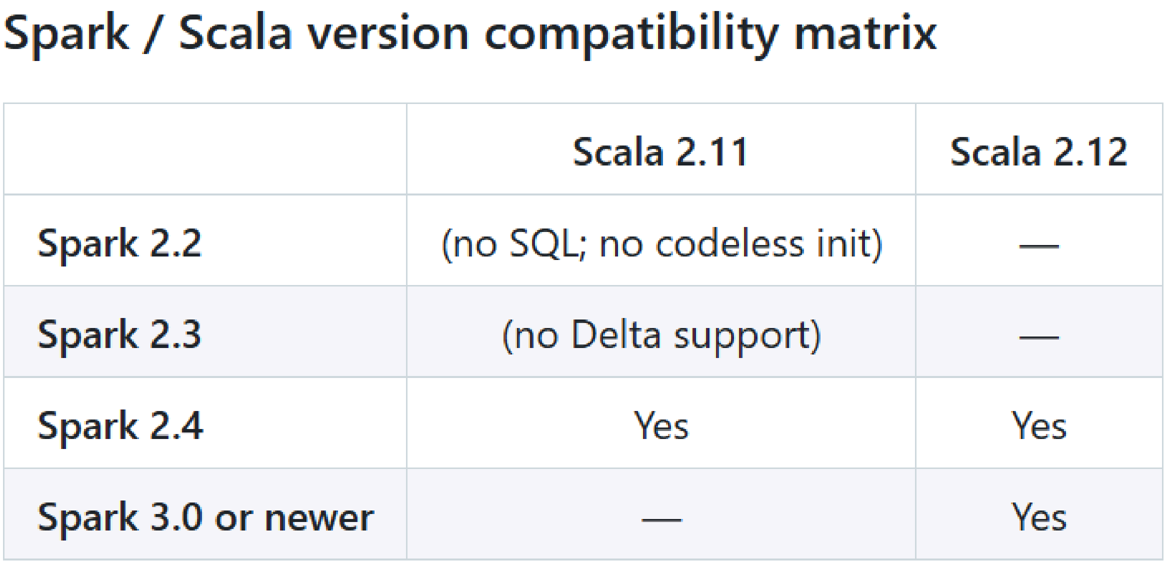
Post-Installation Checklist

- Spline Agent jar is built and uploaded to HDFS
- Spark cluster configuration is updated with Spline properties
- /tmp/spline folder is created and write-access is configured

- Spark cluster is restarted or the configuration is refreshed
- Test Spark jobs are producing lineage files in /tmp/spline
- For Kerberos-secured environments:
  - Kerberos authentication is validated on Windows.
  - Kerberos authentication is validated inside Linux (WSL).
  - WebHDFS delegation tokens are successfully generated.
  - The Cloudera Octopai Client service is running.

Important Notes

- Provide the SE with the Spark, Scala, and Java version details used in your environment. Reach out to Cloudera support to generate the appropriate Spline connector jars.
- Ensure the Spline Agent bundle matches your Spark and Scala versions.



Spark / Scala version compatibility matrix

	Scala 2.11	Scala 2.12
Spark 2.2	(no SQL; no codeless init)	—
Spark 2.3	(no Delta support)	—
Spark 2.4	Yes	Yes
Spark 3.0 or newer	—	Yes

- Only successful jobs with persistent outputs will generate lineage.
- If no lineage appears, verify:
  - The job reads and writes to and from persistent sources
  - The Spark job includes the correct configuration parameters. Check the Spark logs
  - The jar file was correctly uploaded to HDFS and accessible

Appendix A. Sample krb5.ini Configuration:

This appendix provides a reference Kerberos configuration file example. Use this example to validate realm, KDC, and domain mappings. The actual values must be provided by your organization.

```
[libdefaults]
default_realm = ROOT.COMOPS.SITE
dns_lookup_realm = false
dns_lookup_kdc = false
ticket_lifetime = 24h
renew_lifetime = 7d
forwardable = true
[realms]
ROOT.COMOPS.SITE = {
  kdc = ccycloud-1.cdp.root.comops.site
  admin_server = ccycloud-1.cdp.root.comops.site
}

[domain_realm]
```

```
.root.comops.site = ROOT.COMOPS.SITE
root.comops.site = ROOT.COMOPS.SITE
```

On Windows, the file must be placed under C:\ProgramData\MIT\Kerberos5\krb5.ini

On Linux, the file must be placed under /etc/krb5.conf

Correct realm and domain mappings are required for successful Kerberos authentication and delegation token acquisition.

Appendix B. Kerberos Verification and Expected Results:

This appendix describes how to verify that the Kerberos authentication chain is functioning correctly on both Windows and Linux. These checks must be completed before the lineage can be written to HDFS in Kerberos-secured environments.

Windows Verification

Run the following command from a PowerShell window:

```
"C:\Program Files\MIT\Kerberos\bin\kinit.exe" -kt C:\Octopai\spark.keytab spark@ROOT.COMOPS.SITE
```

Expected result: No output indicates successful authentication.

If an error occurs, verify the keytab, principal name, realm configuration, and network connectivity to the KDC.

Linux Verification (WSL Ubuntu)

Run the following commands from PowerShell:

```
wsl kinit -kt /mnt/c/Octopai/spark.keytab spark@ROOT.COMOPS.SITE
wsl klist
```

Expected result: The klist output should display a valid Kerberos ticket, including fields such as Valid starting, Expires, and Service principal.

If no ticket is shown, verify that the krb5.conf file exists under /etc, the keytab path is correct, and the Linux environment can access the KDC.

After all Windows and Linux verifications are complete, start the Cloudera Octopai Client service:

```
Start-Service OctopaiClient
```

Completing these steps successfully verifies that Kerberos authentication and WebHDFS delegation token acquisition are configured correctly.

Configuring Kafka and Kafka Connect Connector in Cloudera Octopai

Learn how to configure Kafka and Kafka Connect connector in Cloudera Octopai Client using Kerberos authentication (SASL/GSSAPI).

Before you begin

Before configuring Kafka and Kafka Connect connectors in Cloudera Octopai Client, ensure the following components are available and properly configured:

- **Kerberos infrastructure:** Active Kerberos Key Distribution Center (KDC), valid Kerberos realm configuration, and network connectivity from the Octopai Client host to the KDC.

- **MIT Kerberos for Windows:** Install MIT Kerberos for Windows on the machine running Cloudera Octopai Client. The default installation path is C:\Program Files\MIT\Kerberos\bin\kinit.exe. Cloudera Octopai Client uses kinit to acquire Kerberos tickets.
- **Kerberos configuration file:** The Kerberos configuration file must exist at C:\ProgramData\MIT \Kerberos5\krb5.ini.

  Example configuration:

  ```
  [libdefaults]
                              default_realm = ROOT.COMOPS.SITE
                              dns_lookup_realm = false
                              dns_lookup_kdc = false
                              ticket_lifetime = 24h
                              renew_lifetime = 7d
                              forwardable = true

                              [realms]
                              ROOT.COMOPS.SITE = {
                              kdc = ccycloud-1.cdp.root.comops.site
                              admin_server = ccycloud-1.cdp.root.comops.site
                              }

                              [domain_realm]
                              .root.comops.site = ROOT.COMOPS.SITE
                              root.comops.site = ROOT.COMOPS.SITE
  ```
- **Kerberos credentials:** Obtain a Kerberos principal (for example, kafka-user@REALM) and its associated keytab file (for example, C:\octopai\kafka-user.keytab). Ensure the keytab file is securely stored and accessible.
- **Kafka cluster configuration:** Ensure the Kafka cluster is configured with SASL/GSSAPI enabled, the Kafka service principal is configured on the brokers (for example, kafka/hostname@REALM), and the broker hostnames are resolvable using fully qualified domain names.

Procedure

1. Create a keytab file.

   On a Kerberos administration server, create a keytab file for the Kerberos principal used by Cloudera Octopai Client using this command:

   ```
   ktutil
                           addent -password -p kafka-user@REALM -k 1 -e aes2
   56-cts
                           wkt /path/to/kafka-user.keytab
                           quit
   ```

   Or request the keytab file from your Kerberos administrator using this command:

   ```
   kadmin -q "ktadd -k /path/to/kafka-user.keytab kafka-user@REALM"
   ```

   Copy the keytab file securely to the Windows server running the Cloudera Octopai Client.

2. Verify Kerberos authentication before configuring Kafka or Kafka Connect.
   a) Open PowerShell.
   b) Navigate to the Kerberos binaries directory:

   ```
   cd "C:\Program Files\MIT\Kerberos\bin"
   ```

   c) Obtain a Kerberos ticket:

   ```
   .\kinit.exe -kt "C:\octopai\kafka-user.keytab" kafka-user@REALM
   ```

   d) Confirm the validity of the ticket:

   ```
   .\klist.exe
   ```

   Note: A valid ticket confirms that Kerberos is configured correctly.

3. Configure Kafka metadata source in Cloudera Octopai Client.
   a) Start a new connection:

      1. Open Cloudera Octopai Client.
      2. Select New Connection.
      3. Choose Kafka from the vendor list.

   b) Provide basic connection information using the following values:

      - Bootstrap Servers:

        Kafka broker hostnames and ports

        Example: kafka1.example.com:9092
      - Schema Registry URL (optional):

        Example: http://schema-registry.example.com:8081

   c) Select Kerberos as the authentication method.
   d) Configure Kerberos settings:

      - Kerberos principal:

        Example: kafka-user@REALM
      - Keytab path:

        Absolute path to the keytab file

        Example: C:\octopai\kafka-user.keytab

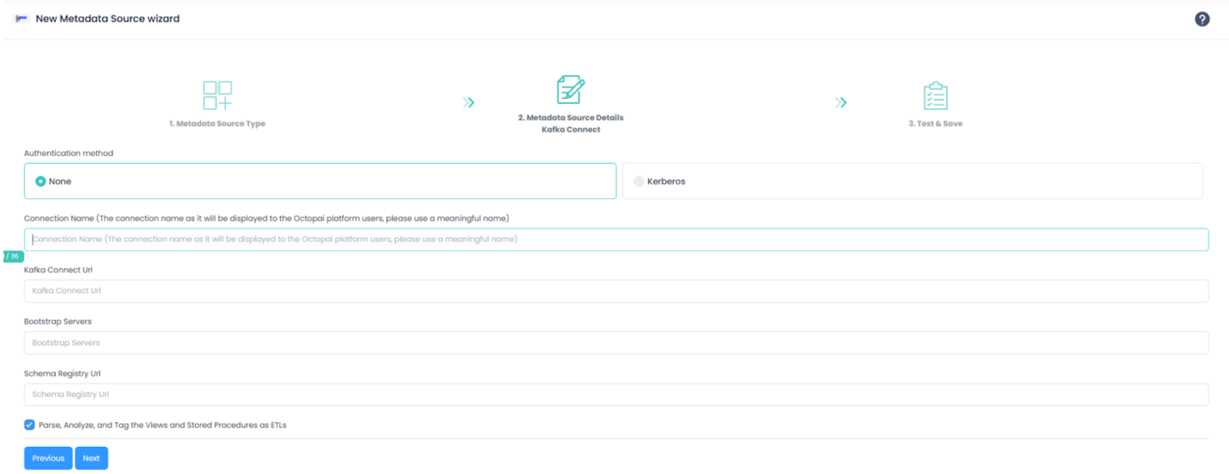
      Important:

      - The realm names are case sensitive.
      - The keytab file must be readable by the Cloudera Octopai Client service account.
      - Broker addresses must use fully qualified domain names.

   e) To test the connection, click Test Connection.
   f) Click Save to store the connection.

4. Set up the Kafka Connect metadata source.

   Note: Kafka Connect metadata sources are configured in Cloudera Octopai Client using a dedicated connector.

   New Metadata Source wizard

   1. Metadata Source Type » 2. Metadata Source Details Kafka Connect » 3. Test & Save

   Authentication method

   ● None ○ Kerberos

   Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

   Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

   / 36

   Kafka Connect Url

   Kafka Connect Url

   Bootstrap Servers

   Bootstrap Servers

   Schema Registry Url

   Schema Registry Url

   ☑ Parse, Analyze, and Tag the Views and Stored Procedures as ETLs

   Previous Next

   a) Configure connection parameters without authentication.

      If the authentication method is **None**, provide the following values:

      - Kafka Connect URL:
        - Required
        - Kafka Connect REST endpoint
      - Bootstrap Servers:
        - Required
        - Kafka broker hostnames
      - Schema Registry URL:
        - Optional

   b) Configure connection parameters with Kerberos authentication.

      If the authentication method is Kerberos, follow the steps described earlier to configure Kerberos.

      Required additional fields:

      - Kerberos principal
      - Keytab path

5. Verify the extracted metadata files.

   a) After the extraction completes, navigate to this folder:

      ```
      C:\Program Files (x86)\Octopai\Service\TGT
      ```

   b) Open the ZIP file matching the connector name.

   c) Verify the presence and structure of the extracted files.

Kerberos authentication errors:

- Verify the principal format: username@REALM
- Confirm the keytab path and permissions.
- Validate the realm configuration in krb5.ini

Kafka connectivity issues:

- Verify the network connectivity.

- Use fully qualified domain names.
- Confirm the SASL listener configuration.

Ticket expiration:

- Verify the ticket_lifetime and renew_lifetime values.
- Adjust the renewal configuration, if required.

Clock skew:

Synchronize the system time:

```
w32tm /resync /force
```

Error during the extraction:

- Collect the logs from C:\Program Files (x86)\Octopai\Service\log

```
POWER_BI_103_21520221025220220215    15/02/2022 10:02    LOG File    3 KB
```

- Send the logs with the connector number and name to Cloudera Support.

AWS Glue

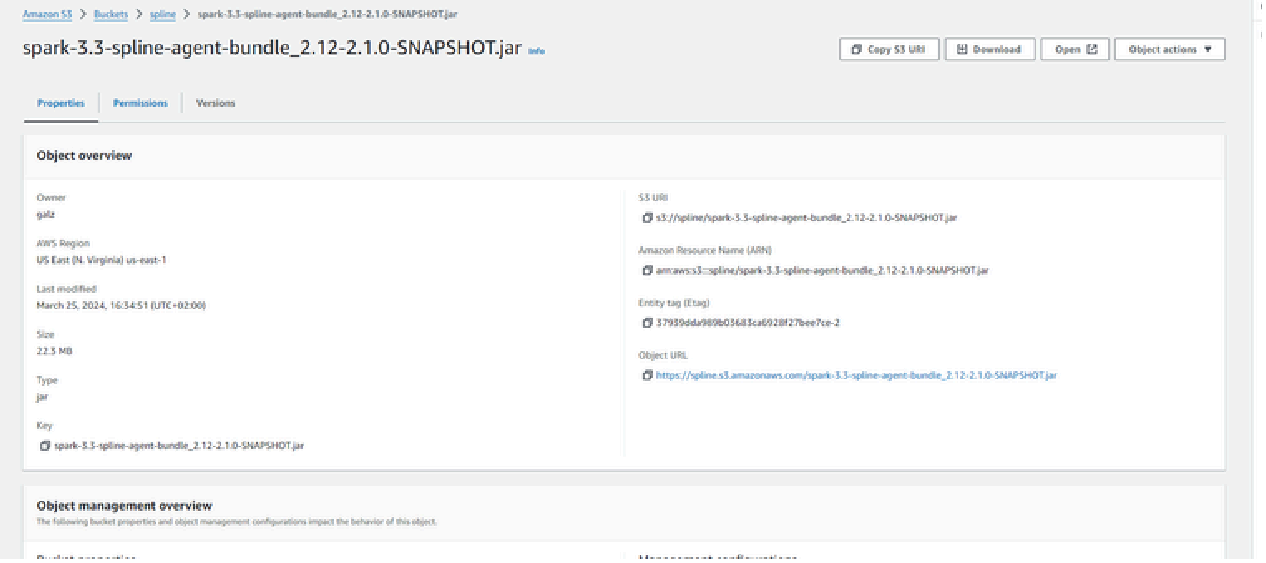
Learn how to configure AWS Glue jobs with Spline integration.

Note: Supported versions: Aws Glue 3- Spark 3.1, Scala 2

How to set up the permissions

Configure Spark Jar:

On S3, create a folder named lib and copy the Spline jar S3 URL related to Spark. This action needs to be done once.


Amazon S3 > Buckets > spline > spark-3.3-spline-agent-bundle_2.12-2.1.0-SNAPSHOT.jar
Info
Copy S3 URI
Download
Open
Object actions ▼
Properties
Permissions
Versions

spark-3.3-spline-agent-bundle_2.12-2.1.0-SNAPSHOT.jar

Object overview

Owner
galz

AWS Region
US East (N. Virginia) us-east-1

Last modified
March 25, 2024, 16:34:51 (UTC+02:00)

Size
22.3 MB

Type
jar

Key
spark-3.3-spline-agent-bundle_2.12-2.1.0-SNAPSHOT.jar

S3 URI
s3://spline/spark-3.3-spline-agent-bundle_2.12-2.1.0-SNAPSHOT.jar

Amazon Resource Name (ARN)
arn:aws:s3:::spline/spark-3.3-spline-agent-bundle_2.12-2.1.0-SNAPSHOT.jar

Entity tag (Etag)
37939dda989b03683ca6928f27bee7ce-2

Object URL
https://spline.s3.amazonaws.com/spark-3.3-spline-agent-bundle_2.12-2.1.0-SNAPSHOT.jar

Object management overview

The following bucket properties and object management configurations impact the behavior of this object.

How to Configure parameters for each Job:

For each job, include in the Job Parameters a parameter named conf with the following values:

```

```

```
        spark.spline.producer.url=https://databricks.spline.octopai.com/pro
ducer
        --conf spark.sql.queryExecutionListeners=za.co.absa.spline.harvest
er.listener.SplineQueryExecutionListener
```

The value of spark.spline.producer.url should be set to the URL of the producer created for the customer in the Jenkins job found at https://jenkins.octopai.com/job/Deploy-Spline-Cluster/.

For example, use the URL https://databricks.spline.octopai.com/producer.

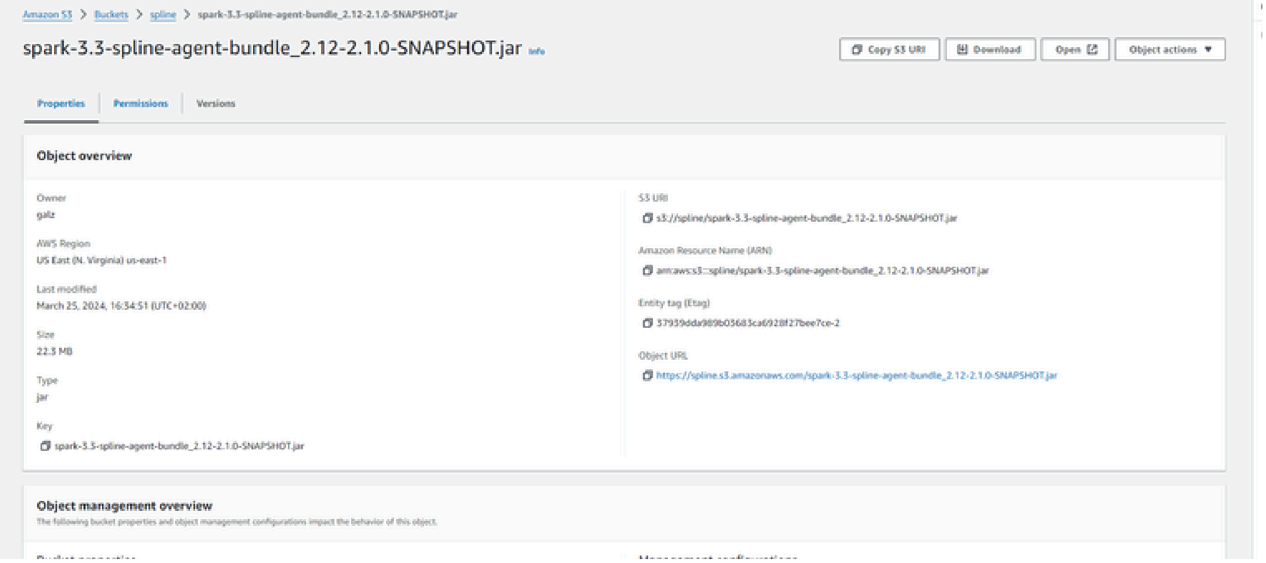
How to Configure parameters for each Job

For each job, include in the "Job Parameters" a parameter named conf with the following values:

```

        spark.spline.producer.url=
        https://databricks.spline.octopai.com/producer
         --conf spark.sql.queryExecutionListeners=za.co.absa.spline.harvest
er.listener.SplineQueryExecutionListener
```

The value of spark.spline.producer.url should be set to the URL of the producer created for the customer in the Jenkins job found at https://jenkins.octopai.com/job/Deploy-Spline-Cluster/ .

For example, use the URL https://databricks.spline.octopai.com/producer .


Amazon S3 > Buckets > spline > spark-3.3-spline-agent-bundle_2.12-2.1.0-SNAPSHOT.jar
Info
Copy S3 URI
Download
Open
Object actions
Properties
Permissions
Versions

spark-3.3-spline-agent-bundle_2.12-2.1.0-SNAPSHOT.jar

Object overview

Owner
galz

AWS Region
US East (N. Virginia) us-east-1

Last modified
March 25, 2024, 16:34:51 (UTC+02:00)

Size
22.5 MB

Type
jar

Key
spark-3.3-spline-agent-bundle_2.12-2.1.0-SNAPSHOT.jar

S3 URI
s3://spline/spark-3.3-spline-agent-bundle_2.12-2.1.0-SNAPSHOT.jar

Amazon Resource Name (ARN)
arn:aws:s3:::spline/spark-3.3-spline-agent-bundle_2.12-2.1.0-SNAPSHOT.jar

Entity tag (Etag)
57939dda989b05683ca6928f27bee7ce-2

Object URL
https://spline.s3.amazonaws.com/spark-3.3-spline-agent-bundle_2.12-2.1.0-SNAPSHOT.jar

Object management overview
The following bucket properties and object management configurations impact the behavior of this object.

Configure Databricks in Cloudera Octopai

Learn how to integrate Databricks with Cloudera Octopai Data Lineage based on your catalog type, including Unity Catalog, Hive Metastore, or hybrid (Unity Catalog and Hive Metastore) deployments.

About this task

Before configuring Databricks in Cloudera Octopai, review the prerequisites that apply to your catalog environment. The configuration requirements vary depending on whether you use Unity Catalog, Hive Metastore, or a hybrid deployment combining both.

Requirements for Unity Catalog

Unity Catalog environments require system table access and Databricks SQL connectivity. To extract lineage, Cloudera Octopai must authenticate with a service principal and query Unity Catalog lineage system tables.

Note: Databricks admin permissions are required to view and manage Unity Catalog system tables and configure the access controls needed for lineage extraction.

Requirements for Hive Metastore

For environments using only Hive Metastore, ensure that the user or machine identity meets the following requirements:

- Permission to view and access the workspace folders containing the notebooks.
- Read access to the projects or directories selected for metadata extraction.
- Can view the relevant Hive Metastore objects referenced by the notebooks.

Requirements for Unity Catalog and Hive Metastore Hybrid Deployments

In hybrid environments, the following requirements must be met:

- Unity Catalog prerequisites, including SQL Warehouse access and permissions.
- Hive Metastore assets must also be included.
- Cloudera Octopai combines metadata sources to provide extended lineage coverage.

Perform the following steps to configure Databricks in Cloudera Octopai:

Procedure

1. Create a service principal (required)

   Unity Catalog lineage extraction requires a machine identity with access to governed metadata.

   You must ensure the following:

   - Create a Databricks-managed service principal.
   - Enable Workspace access and Databricks SQL access.

2. Enable or create an SQL Warehouse (required)

Cloudera Octopai relies on querying Databricks system tables, which requires a running SQL Warehouse.

You must ensure the following:

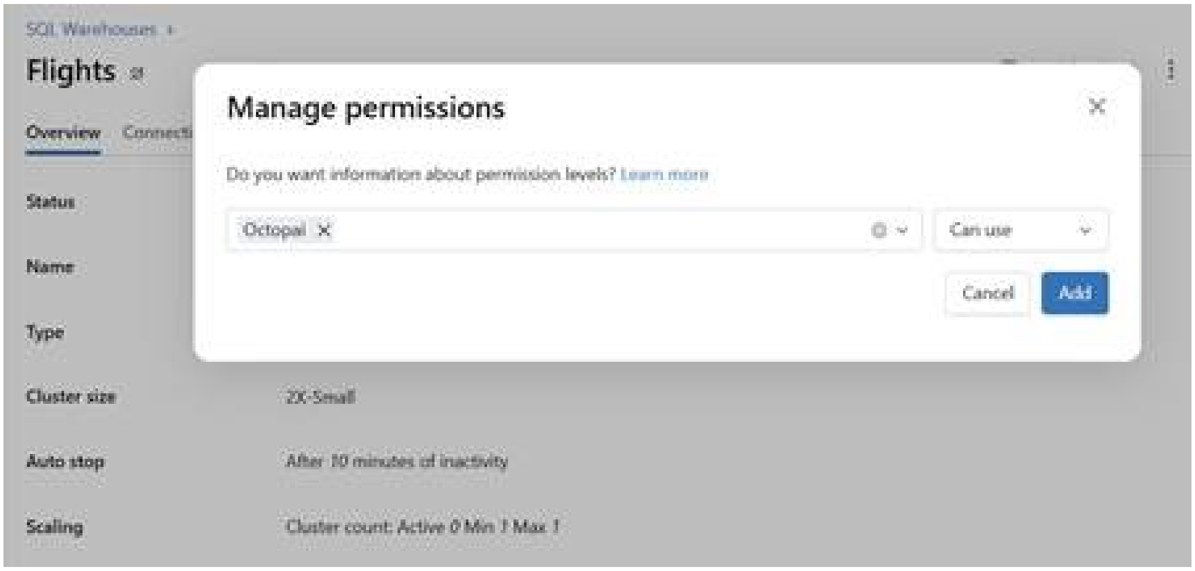
- Create or enable a Databricks SQL Warehouse.
- Allow access to required system schemas.

Perform the following steps:

a. In Databricks, go to the SQL Warehouses tab.

b. If no SQL Warehouse exists,, click Create SQL Warehouse and configure it as required.

c. Assign the service principal Manager permissions to the warehouse by selecting Can use.



Manage permissions

Do you want information about permission levels? Learn more

Octopai × — Can use

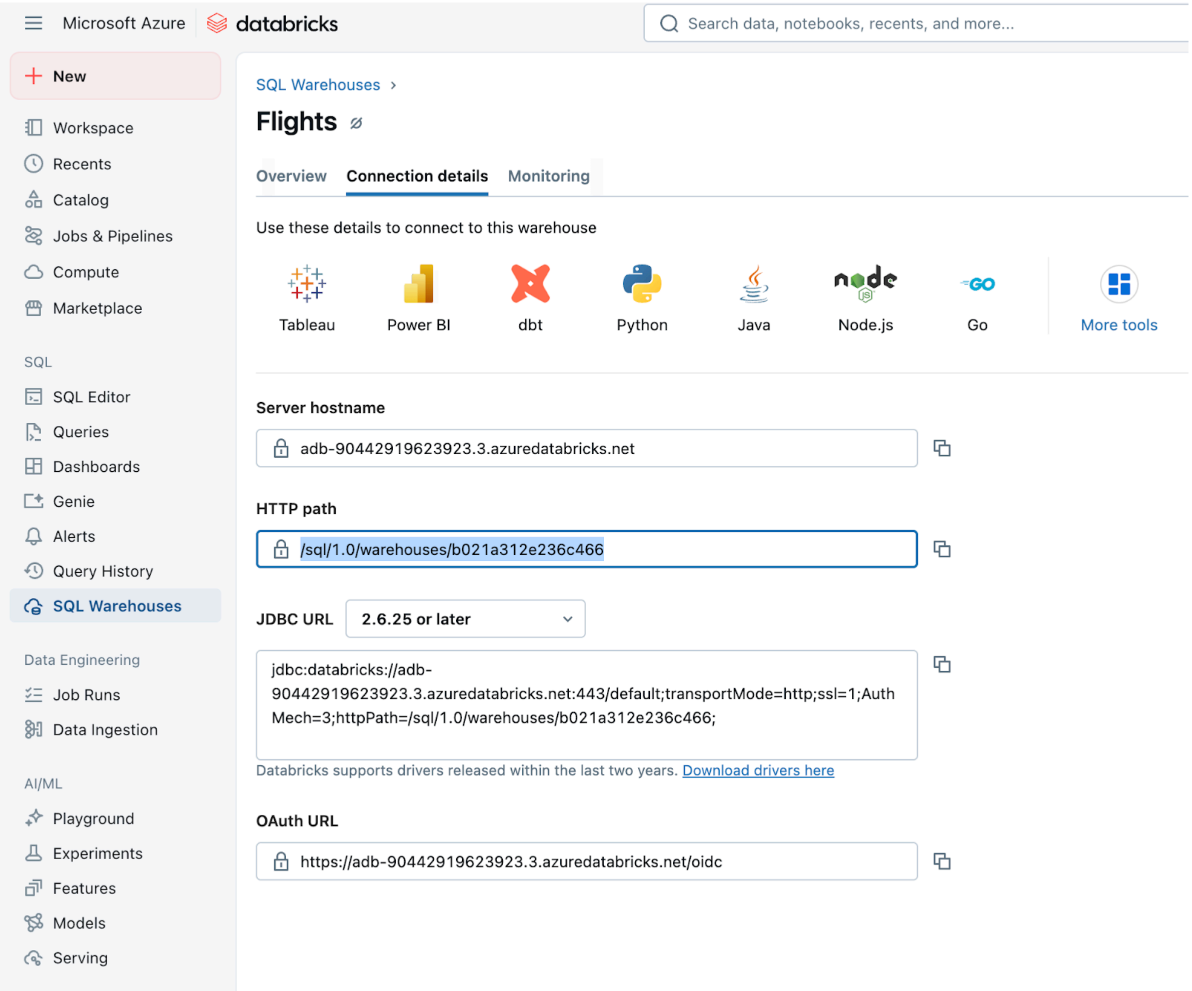
Cluster size 2X-Small

Auto stop After 10 minutes of inactivity

Scaling Cluster count: Active 0 Min 1 Max 1

d. Open the SQL Warehouse and navigate to Connection Details.

e. Copy the HTTP path. You will need this path for the integration process.



Use these details to connect to this warehouse

Server hostname

adb-90442919623923.3.azuredatabricks.net

HTTP path

/sql/1.0/warehouses/b021a312e236c466

JDBC URL 2.6.25 or later

jdbc:databricks://adb-90442919623923.3.azuredatabricks.net:443/default;transportMode=http;ssl=1;AuthMech=3;httpPath=/sql/1.0/warehouses/b021a312e236c466;

Databricks supports drivers released within the last two years. Download drivers here

OAuth URL

https://adb-90442919623923.3.azuredatabricks.net/oidc

3. Ensure Unity Catalog–Enabled Compute

- Unity Catalog must be enabled at the workspace/account level.
- A cluster that supports Unity Catalog access must be available.

4. Grant Unity Catalog Lineage Permissions (required)

The service principal must have SELECT access on the system lineage tables (system.access.table_lineage and system.access.column_lineage) and read access on relevant catalogs and schemas.

a. Open the Catalog in Databricks.

b. Search for:

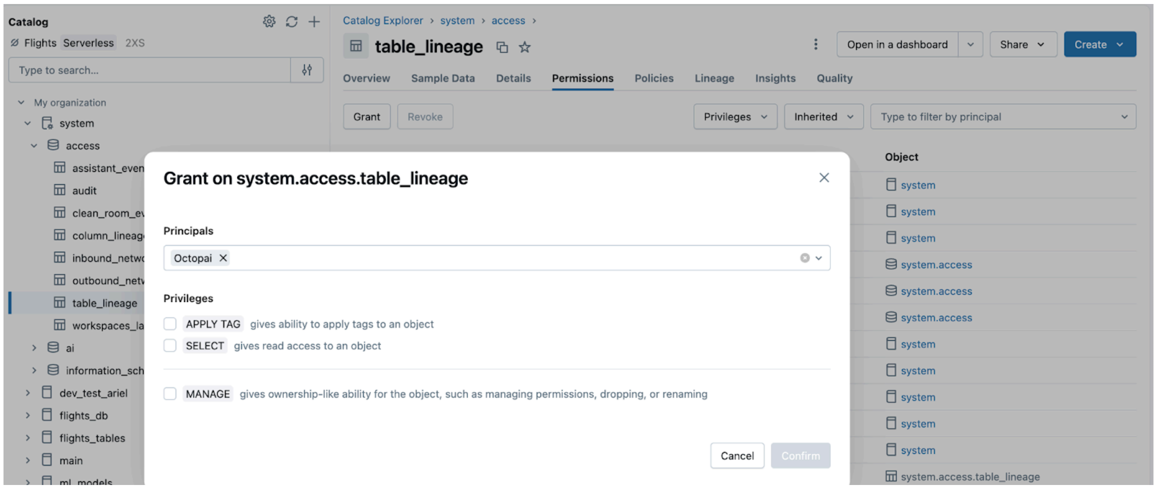
- Catalog: system
- Schema: access
- Tables: table_lineage and column_lineage

c. The tables are automatically created by Databricks.

Note: You must have admin permissions to view and manage the tables.

a. For each table, perform the following steps:

- Open the Permissions tab.
- Click Grant.
- Select the service principal created earlier.
- Enable Select Permission.



Grant on system.access.table_lineage

Principals

Octopai

Privileges

□ APPLY TAG gives ability to apply tags to an object

□ SELECT gives read access to an object

□ MANAGE gives ownership-like ability for the object, such as managing permissions, dropping, or renaming

5. Download the ODBC Driver

- Download and install the Simba ODBC Driver for Databricks from the official Databricks download page: https://www.databricks.com/spark/odbc-drivers-download
- Select the appropriate version for your operating system (Windows or Linux).

6. Collect the required workspace information:

    - Workspace URL
    - Workspace ID (required)
    - Account ID (optional)

    a) Find the Databricks Account ID

    The account ID is available in the Databricks Account Console.

    1. Open the account console: https://accounts.cloud.databricks.com/
    2. Log in using your organization's credentials (SSO may be required).
    3. In the top-right corner, select your username/email to open the dropdown menu.
    4. Databricks displays the Account ID as a UUID value, for example: 55eb1a01-48d5-4008-8dbd-03dd8447 a595
    5. Copy this value.

    b) Find the Databricks Workspace ID

    The workspace ID is embedded directly in your Databricks workspace URL.

    1. Open your Databricks workspace in the browser, for example:

        ```
        https://adb-90442919623923.3.azuredatabricks.net/
        ```

        or

        ```
        https://adb-90442919623923.3.azuredatabricks.net/?o=90442919623923
        ```

    2. Locate the parameter ?o= in the URL, for example:

        https://mycompany.cloud.databricks.com/?o=90442919623923 # Workspace ID = 90442919623923
    3. If you do not find the ?o= parameter, navigate to Sidebar Data Science & Engineering .

        The URL will update to include the workspace ID:

        ```
        https://mycompany.cloud.databricks.com/?o=90442919623923#workspace/
        ```

    This verifies the workspace ID value.

Related Information

Databricks Lineage in Cloudera Octopai Data Lineage

Configure Databricks authentication in Cloudera Octopai

Learn about the authentication methods available for connecting Cloudera Octopai Data Lineage to Databricks, including machine-to-machine authentication using service principals and user authentication using Personal Access Tokens.

Cloudera Octopai Data Lineage supports two authentication approaches for Databricks integration, applicable to Unity Catalog, Hive Metastore, or hybrid (Unity Catalog and Hive Metastore) deployment.

Option 1: Machine-to-machine authentication (service principal)

This is the recommended approach for production deployments.

For environments using Unity Catalog, you must configure a service principal. This is because Cloudera Octopai must authenticate using an identity with permission to query Unity Catalog system lineage tables.

To configure Databricks authentication, ensure the following:

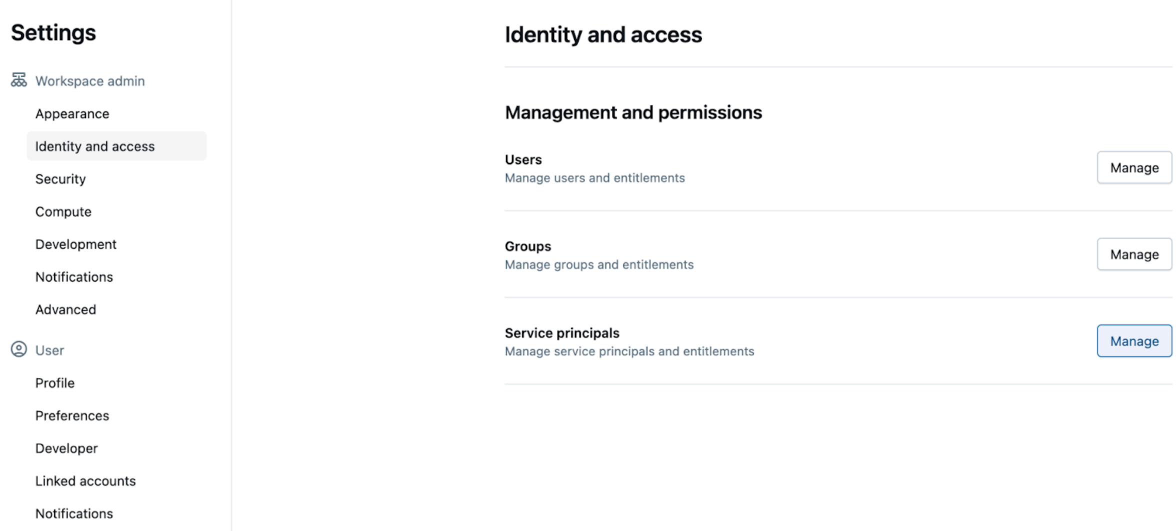
- Create a Databricks service principal.
- Assign it to the workspace.
- Grant the required Unity Catalog and system table permissions.
- Generate OAuth credentials for secure access.

This method enables automated extraction without relying on a personal user account.

Create a dedicated service principal

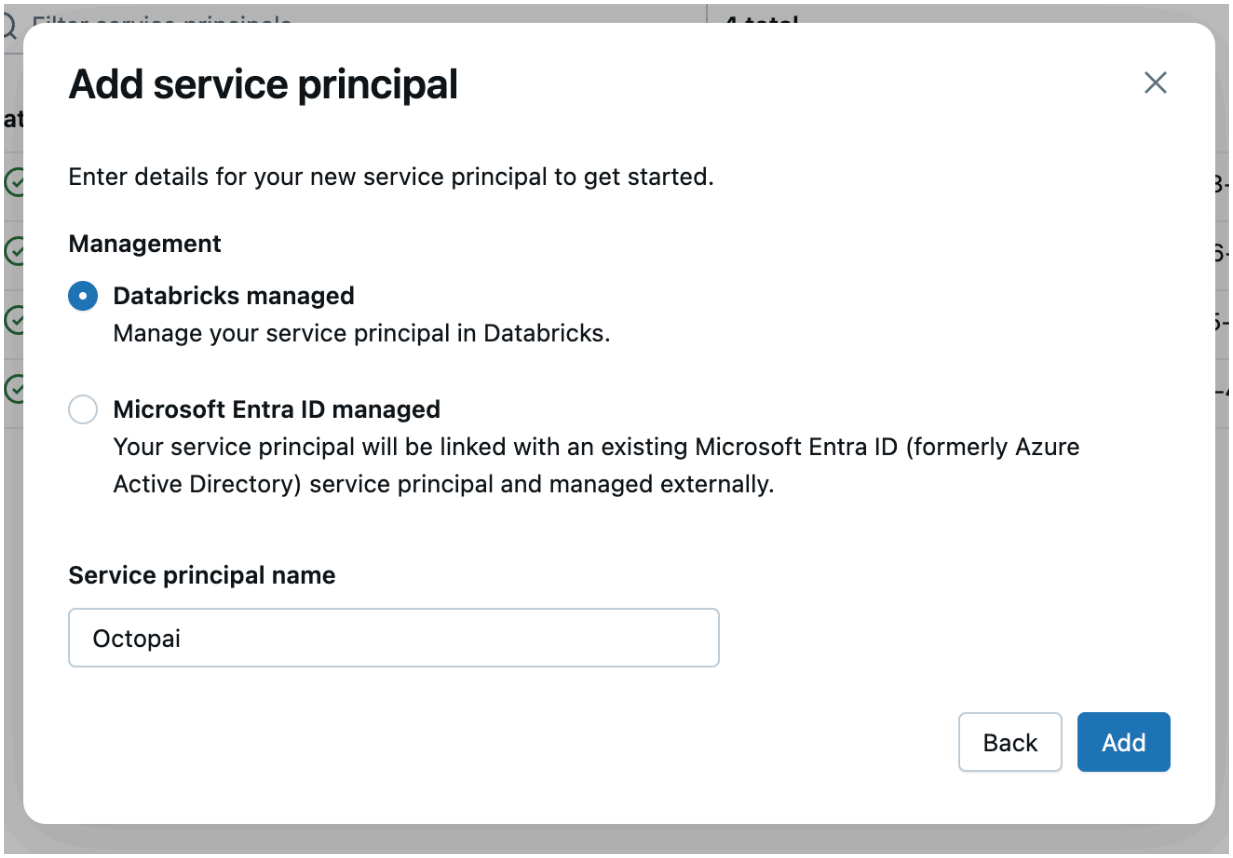
To create a dedicated service principal, perform the following steps:

1. In the Databricks workspace, navigate to Settings.
2. Go to Identity and Access Service Principals .



3. Click Manage, then select Add Service Principal.

4. Choose Databricks Managed and assign a descriptive name (for example, octopai).



Add service principal

Enter details for your new service principal to get started.

Management

● Databricks managed
Manage your service principal in Databricks.

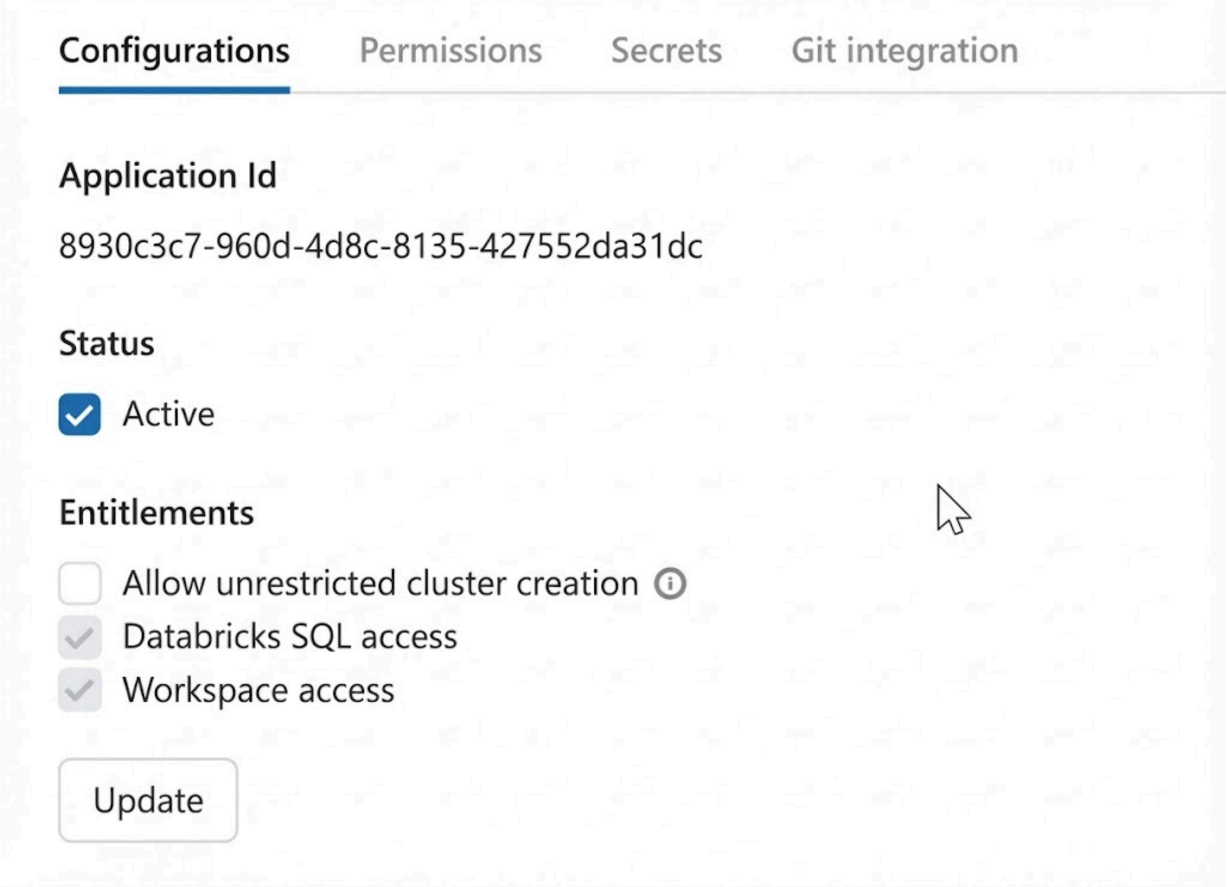
○ Microsoft Entra ID managed
Your service principal will be linked with an existing Microsoft Entra ID (formerly Azure Active Directory) service principal and managed externally.

Service principal name

Octopai

5. Open the created service principal and navigate to the Configurations tab.

6. Select Databricks SQL Access and Workspace Access.



Application Id

8930c3c7-960d-4d8c-8135-427552da31dc

Status

☑ Active

Entitlements

□ Allow unrestricted cluster creation

☑ Databricks SQL access

☑ Workspace access

Option 2: User authentication token (Personal Access Token)

Alternatively, you can authenticate using a Databricks user token (applicable only for HMS).

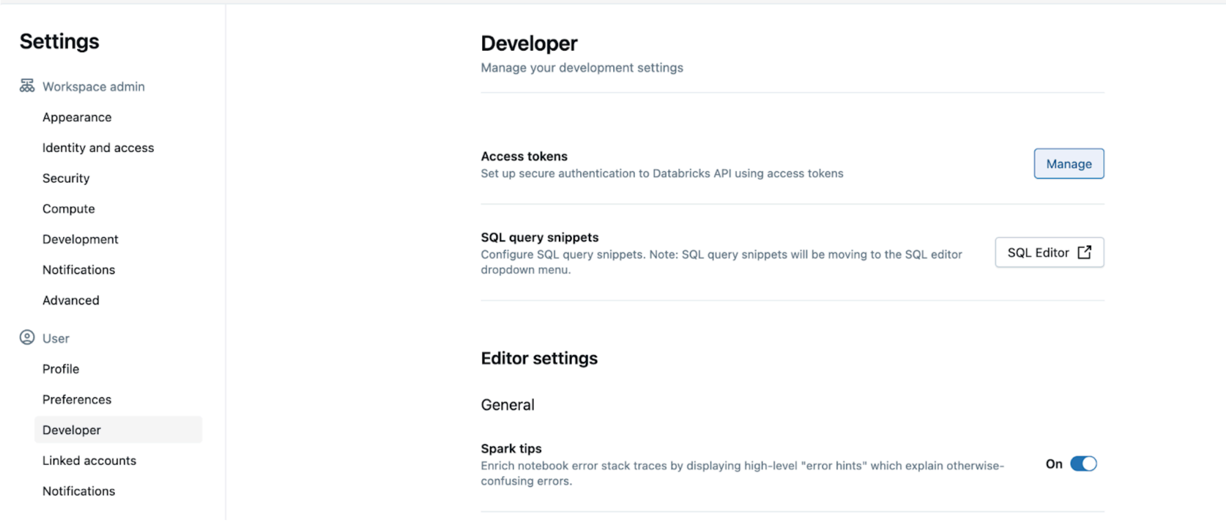
You must ensure the following:

- Generate a Personal Access Token (PAT).
- Provide the token during the Cloudera Octopai setup.

Generate a Personal Access Token

Perform the following steps to generate a Personal Access Token:

1. In Databricks, navigate to Settings Developer Access Tokens (Manage).



Developer

Manage your development settings

Access tokens

Set up secure authentication to Databricks API using access tokens

SQL query snippets

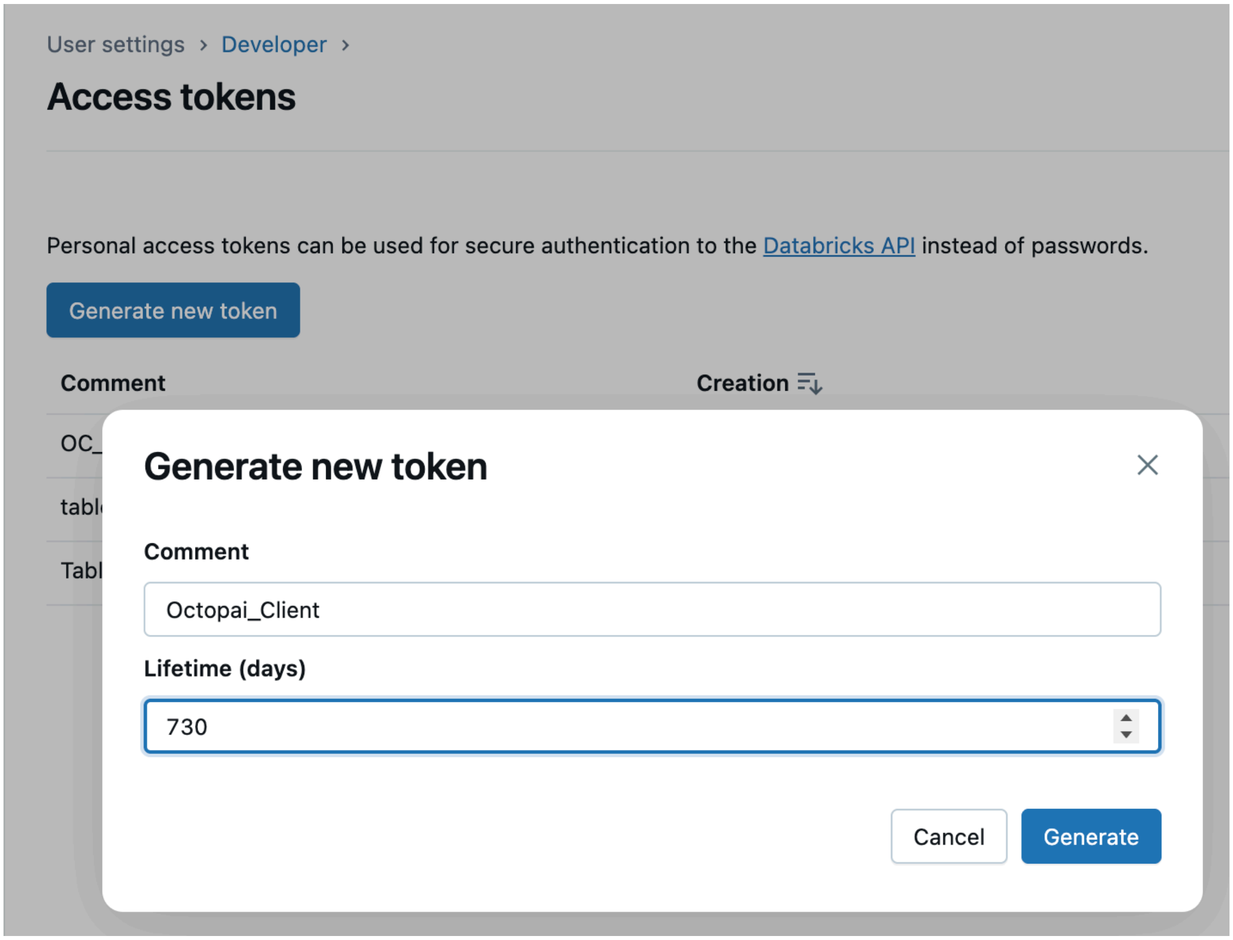
Configure SQL query snippets. Note: SQL query snippets will be moving to the SQL editor dropdown menu.

Spark tips

Enrich notebook error stack traces by displaying high-level "error hints" which explain otherwise-confusing errors.

2. Click Generate New Token.
3. Set the maximum lifespan for the token.

Note: The token must be periodically regenerated.



Access tokens

Personal access tokens can be used for secure authentication to the Databricks API instead of passwords.

Generate new token

Comment: Octopai_Client

Lifetime (days): 730

Related Information

Databricks Lineage in Cloudera Octopai Data Lineage

Configure Databricks Metadata Source in Cloudera Octopai

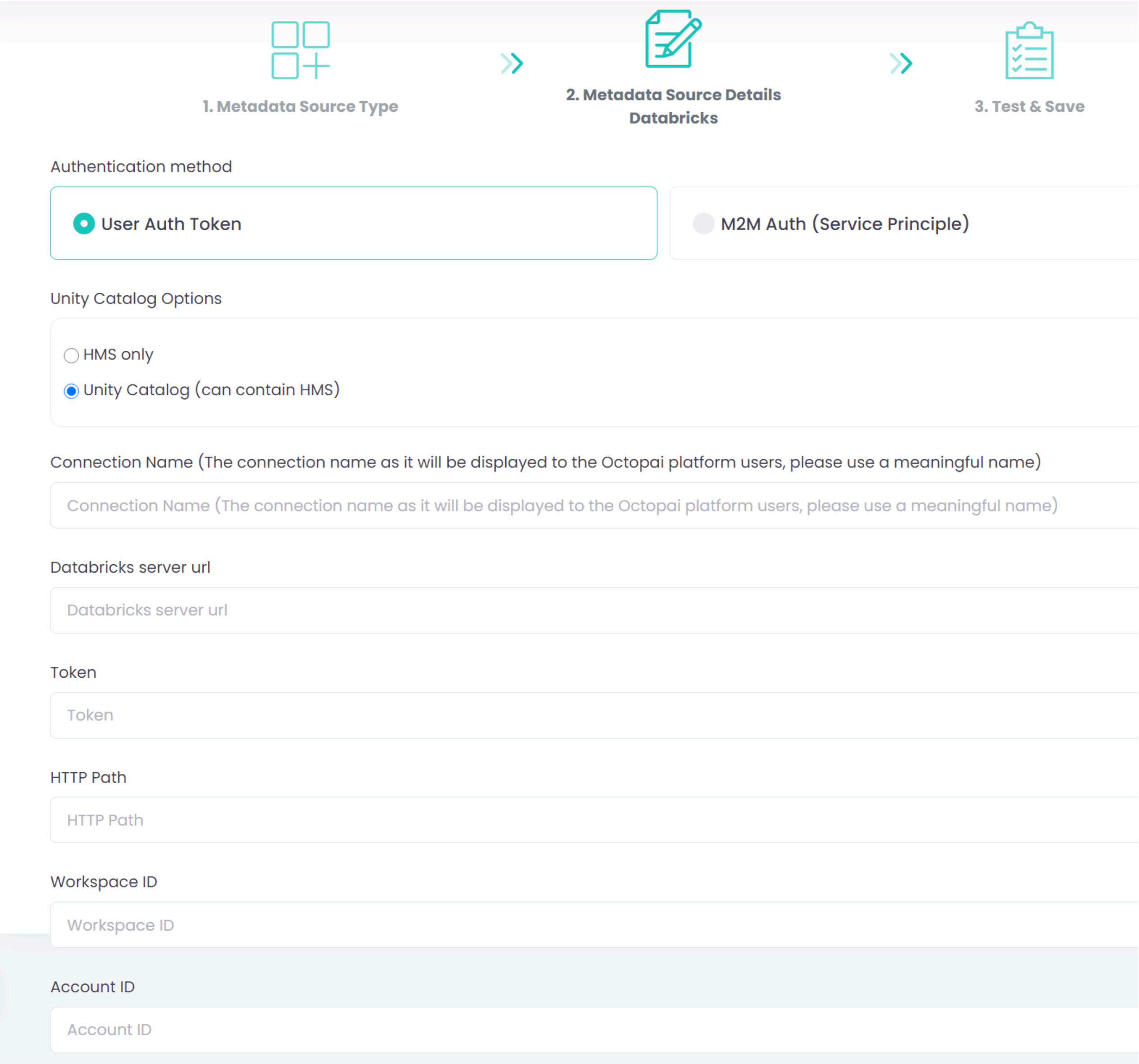
Learn how to configure the Databricks Metadata Source in Cloudera Octopai using either user authentication with Personal Access Tokens or machine-to-machine authentication with service principals.

Cloudera Octopai Data Lineage supports two authentication methods for connecting to Databricks:

- User authentication using a Personal Access Token
- Machine-to-machine (M2M) authentication using a service principal

Important: Select one of the methods when configuring your connection.

Option 1: User authentication token (Personal Access Token)



Authentication method

● User Auth Token ○ M2M Auth (Service Principle)

Unity Catalog Options

○ HMS only

● Unity Catalog (can contain HMS)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Databricks server url

Databricks server url

Token

Token

HTTP Path

HTTP Path

Workspace ID

Workspace ID

Account ID

Account ID

Configure the following settings when using the Personal Access Token authentication method:

1. Unity Catalog Options
   - HMS only – when Databricks uses Hive Metastore without Unity Catalog.
   - Unity Catalog (can contain HMS) – when Databricks uses Unity Catalog. Hive Metastore can also be used (not mandatory).
2. Connection Name

   Assign a clear and meaningful name for the connection. This name will appear to users within the Cloudera Octopai platform.
3. Databricks Server URL

   Enter the customer's Databricks workspace URL.

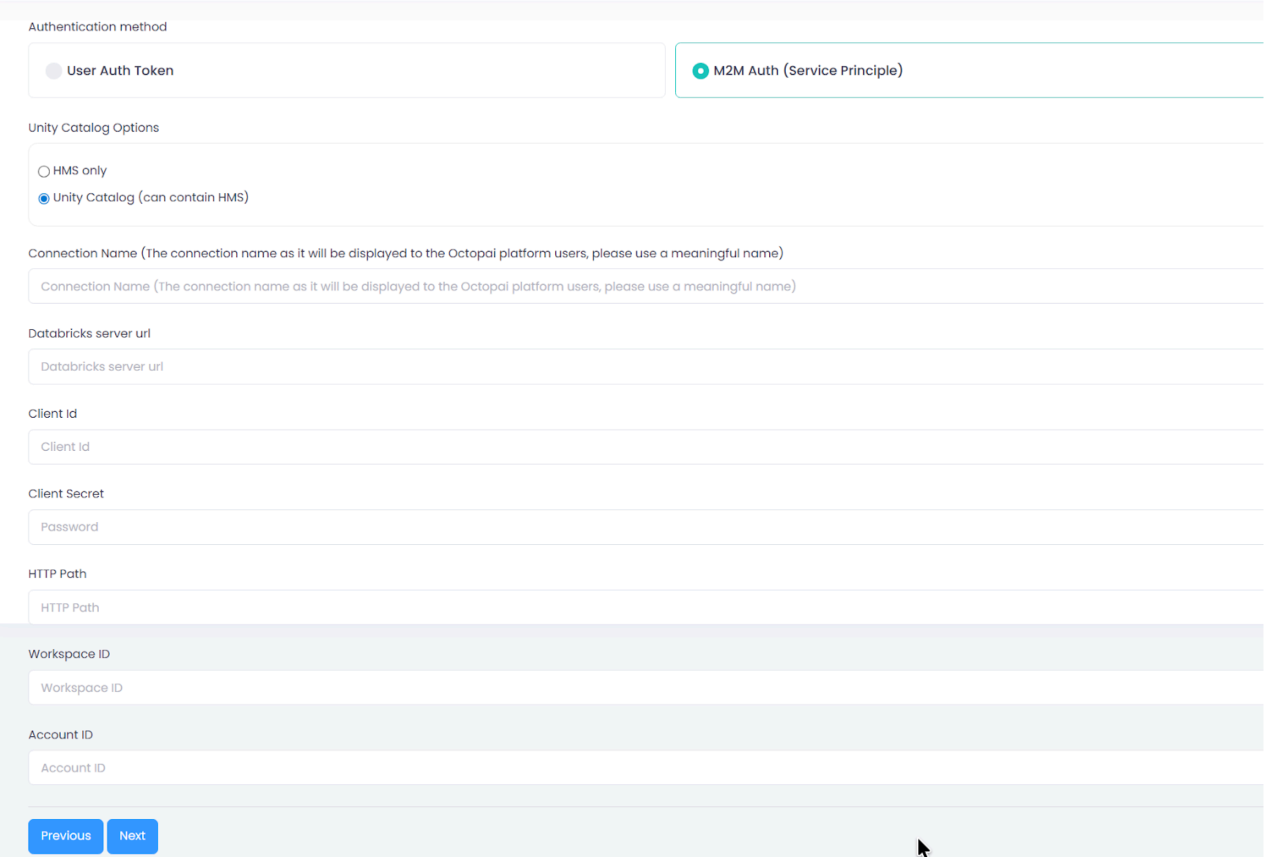
   Example: https://abc-1234.5.azuredatabricks.net
4. Token

   Enter the Personal Access Token generated under SettingsDeveloperAccess Tokens (Manage) in Databricks.
5. HTTP Path

   Paste the HTTP Path copied from the Databricks SQL Warehouse Connection Details field.

   Example: /sql/1.0/warehouses/abc123xyz
6. Workspace ID (for Unity Catalog only)
7. Account ID (for Unity Catalog only, optional)

Option 2: Machine-to-machine authentication (service principal)



Authentication method

○ User Auth Token ● M2M Auth (Service Principle)

Unity Catalog Options

○ HMS only
● Unity Catalog (can contain HMS)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Databricks server url

Databricks server url

Client Id

Client Id

Client Secret

Password

HTTP Path

HTTP Path

Workspace ID

Workspace ID

Account ID

Account ID

Configure the following settings when using the service principal authentication method:

1. **Unity Catalog Options**
   - HMS only – when Databricks uses Hive Metastore without Unity Catalog.
   - Unity Catalog (may include HMS) – when Databricks uses Unity Catalog. Hive Metastore can also be used but is not mandatory.
2. Connection Name

   Assign a clear and meaningful name for the connection. This name will appear to users within the platform.
3. Databricks Server URL

   Enter the customer's Databricks workspace URL.

   Example: https://abc-1234.5.azuredatabricks.net
4. Client ID

   Enter the Client ID of the service principal created in Databricks.
5. Client Secret

   Enter the secret token generated for the service principal.
6. HTTP Path (for Unity Catalog only)

   Paste the HTTP Path copied from the Databricks SQL Warehouse Connection Details field.

   Example: /sql/1.0/warehouses/abc123xyz
7. Workspace ID (for Unity Catalog only)
8. Account ID (for Unity Catalog only, optional)

Related Information

Databricks Lineage in Cloudera Octopai Data Lineage

IICS (Informatica Cloud)

Learn how to configure Informatica Cloud (IICS) permissions, set up metadata sources, verify files, access the Cloudera Octopai Data Lineage target folder, and troubleshoot issues.

Before you begin

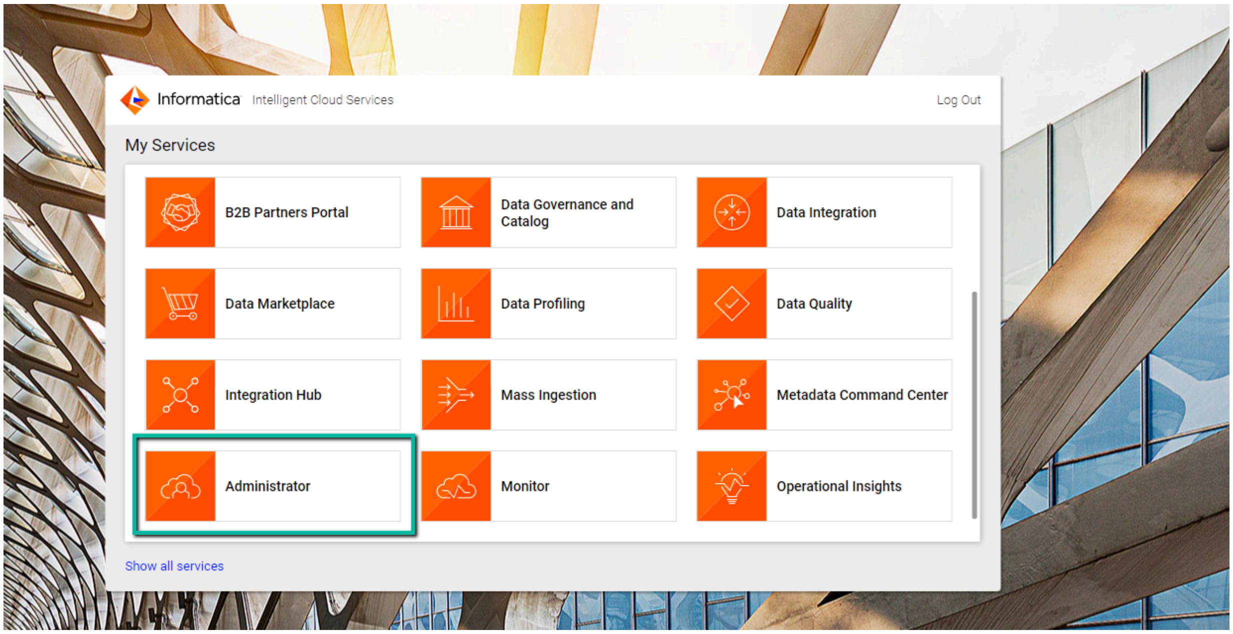
Warning: Missing permissions could end up in broken lineages.

- Add user role:

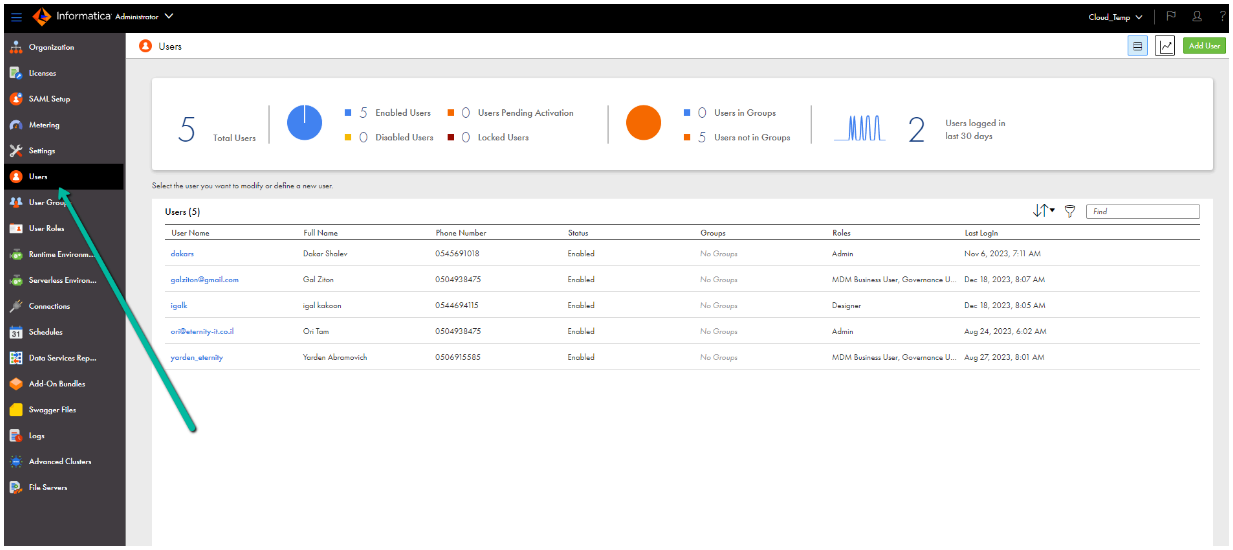
  Log in as admin or any other role that can edit the roles of other users

Procedure

1. Select the Administrator service.



2. On the left panel, select Users.



Select the user you want to modify or define a new user.

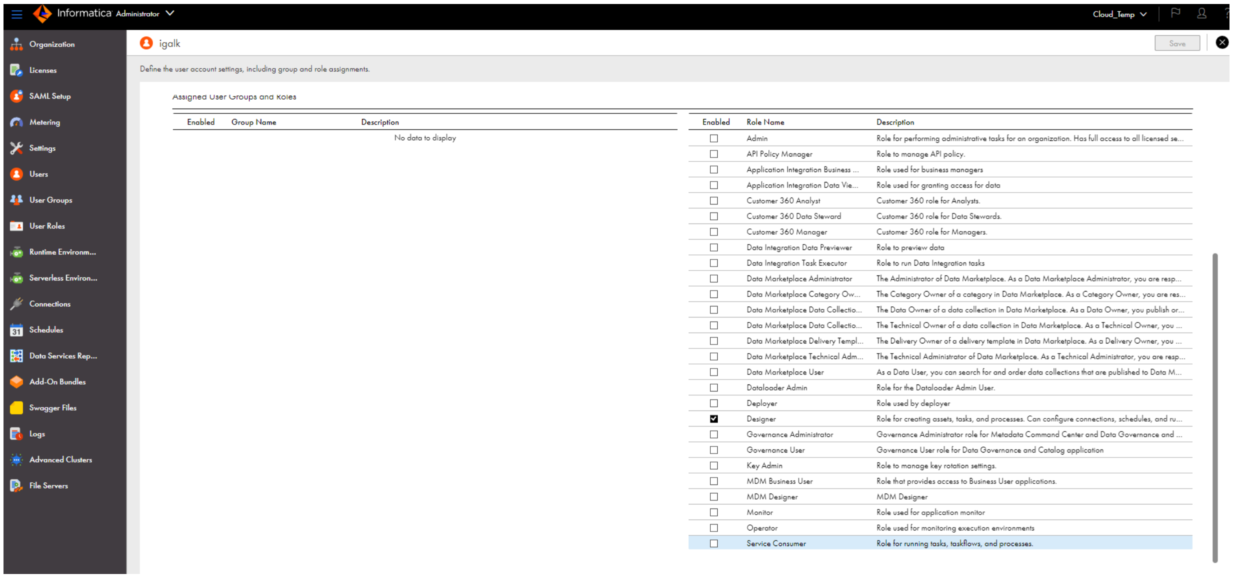
Users (5)

User Name	Full Name	Phone Number	Status	Groups	Roles	Last Login
dakars	Dakar Shalev	0545691018	Enabled	No Groups	Admin	Nov 6, 2023, 7:11 AM
galziton@gmail.com	Gal Ziton	0504938475	Enabled	No Groups	MDM Business User, Governance U...	Dec 18, 2023, 8:07 AM
igalk	igal kakoon	0544694115	Enabled	No Groups	Designer	Dec 18, 2023, 8:05 AM
ori@eternity-it.co.il	Ori Tam	0504938475	Enabled	No Groups	Admin	Aug 24, 2023, 6:02 AM
yarden_eternity	Yarden Abramovich	0506915585	Enabled	No Groups	MDM Business User, Governance U...	Aug 27, 2023, 8:01 AM

3. Scroll down to the role list, and select the "Designer" role.

   Select the user to whom you want to assign a role.

   Click the Save button in the upper right screen.



igalk

Define the user account settings, including group and role assignments.

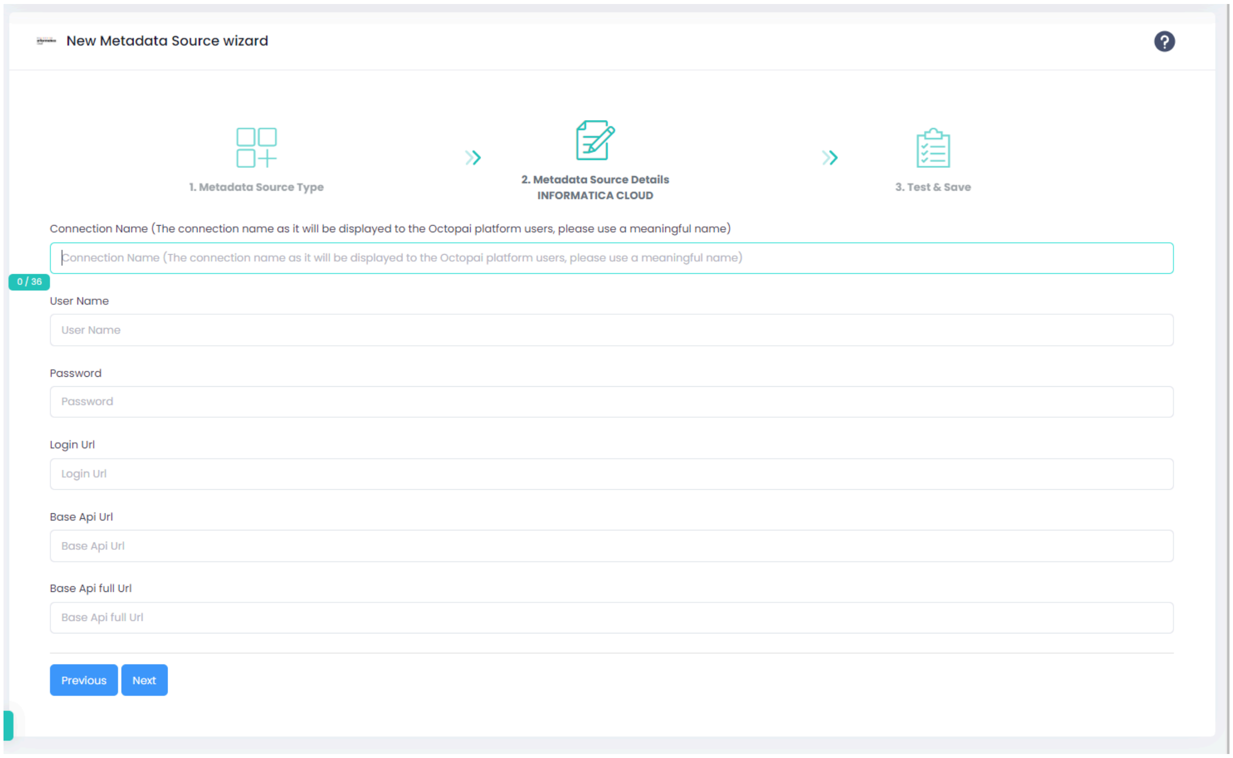
Assigned User Groups and Roles

Enabled	Group Name	Description
	No data to display	

Enabled	Role Name	Description
☐	Admin	Role for performing administrative tasks for an organization. Has full access to all licensed se...
☐	API Policy Manager	Role to manage API policy.
☐	Application Integration Business ...	Role used for business managers
☐	Application Integration Data Vie...	Role used for granting access for data
☐	Customer 360 Analyst	Customer 360 role for Analysts.
☐	Customer 360 Data Steward	Customer 360 role for Data Stewards.
☐	Customer 360 Manager	Customer 360 role for Managers.
☐	Data Integration Data Previewer	Role to preview data
☐	Data Integration Task Executor	Role to run Data Integration tasks
☐	Data Marketplace Administrator	The Administrator of Data Marketplace. As a Data Marketplace Administrator, you are resp...
☐	Data Marketplace Category Ow...	The Category Owner of a category in Data Marketplace. As a Category Owner, you are res...
☐	Data Marketplace Data Collectio...	The Data Owner of a data collection in Data Marketplace. As a Data Owner, you publish or...
☐	Data Marketplace Data Collectio...	The Technical Owner of a data collection in Data Marketplace. As a Technical Owner, you ...
☐	Data Marketplace Delivery Templ...	The Delivery Owner of a delivery template in Data Marketplace. As a Delivery Owner, you ...
☐	Data Marketplace Technical Adm...	The Technical Administrator of Data Marketplace. As a Technical Administrator, you are resp...
☐	Data Marketplace User	As a Data User, you can search for and order data collections that are published to Data M...
☐	Dataloader Admin	Role for the Dataloader Admin User.
☐	Deployer	Role used by deployer
☑	Designer	Role for creating assets, tasks, and processes. Can configure connections, schedules, and ru...
☐	Governance Administrator	Governance Administrator role for Metadata Command Center and Data Governance and ...
☐	Governance User	Governance User role for Data Governance and Catalog application
☐	Key Admin	Role to manage key rotation settings.
☐	MDM Business User	Role that provides access to Business User applications.
☐	MDM Designer	MDM Designer
☐	Monitor	Role used for application monitor
☐	Operator	Role used for monitoring execution environments
☐	Service Consumer	Role for running tasks, taskflows, and processes.

4. Set up Informatica Metadata Source.

   Metadata Sources are set on the Cloudera Octopai Client.



Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

User Name

User Name

Password

Password

Login Url

Login Url

Base Api Url

Base Api Url

Base Api full Url

Base Api full Url

baseApiUrl: The URL of Informatica Cloud. To derive the correct URL, take the application's URL and remove the first segment before the dot. For example, https://emw1.dm-em.informaticacloud.com/saas converts to https://dm-em.informaticacloud.com/saas.

baseApiUrlFull: For example, https://emw1.dm-em.informaticacloud.com/saas is the full URL of the app.

5. Verify the extracted Metadata File.

6. Access the Cloudera Octopai Target Folder (TGT).

a) Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**

b) Open the zip file having the Connector Name.

c) Verify its content: Quantity & Quality of inner files.

Error during the extraction:

- Check the permissions.
- Send the log with the connector number and name to Cloudera Support:

  C:\Program Files (x86)\Octopai\Service\log

SSIS (Files / Integration Services / Integration Services Catalog / Azure)

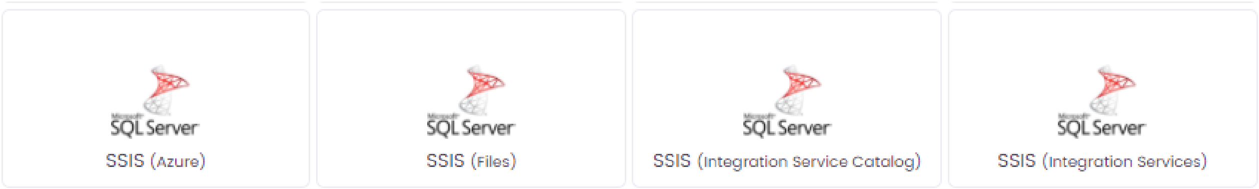
Learn how to configure SSIS (Files, Integration Services, Integration Services Catalog, Azure) as metadata sources for Cloudera Octopai.

Note: Version supported: up to SQL 2022

SSIS method

In the Cloudera Octopai Client, select the SSIS method implemented in your organization.

Figure 13: Select the SSIS method in the Cloudera Octopai Client



SSIS Files

Warning:

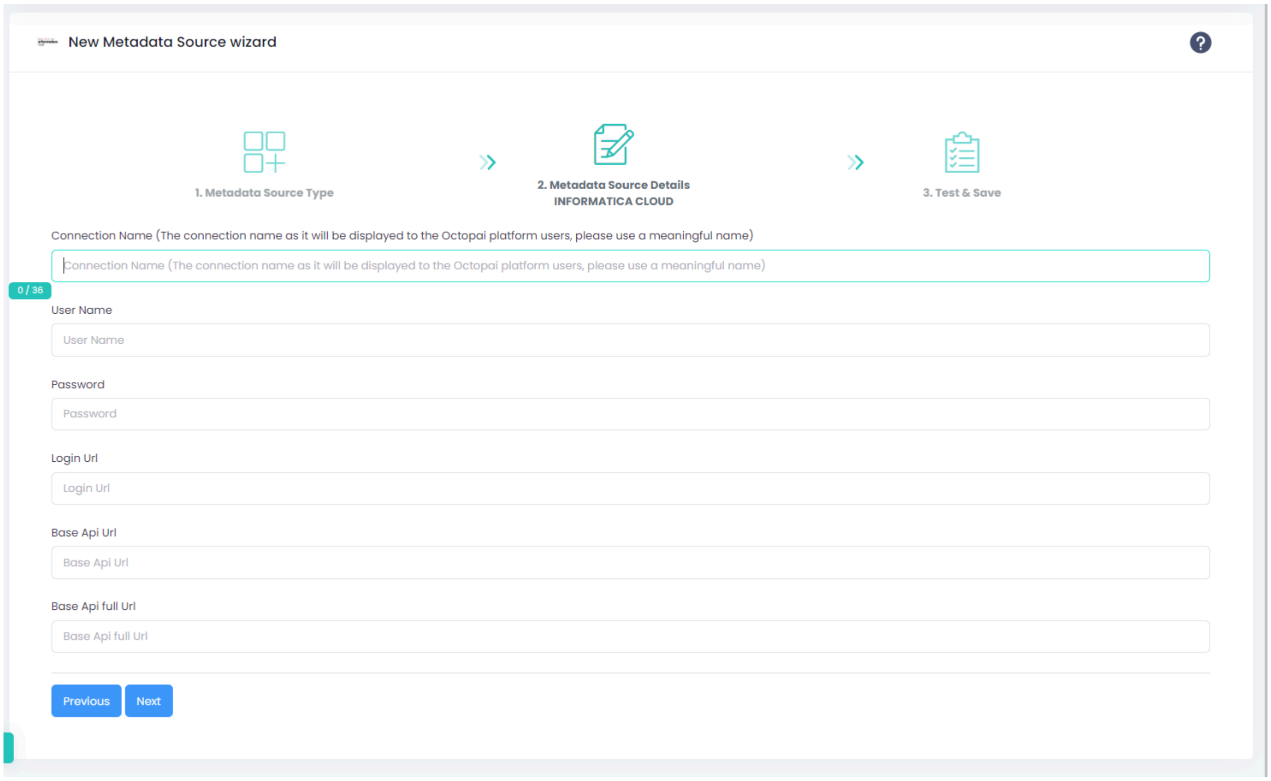
Missing permissions could end up in broken lineages.

Note:

Permissions Prerequisites:

- Read Permission for the Cloudera Octopai Windows NT User to the SSIS folder containing the *.dtsx files.
- Read Permission for the Cloudera Octopai Windows NT User to the SSIS folder containing the *.dtsx files.

Figure 14: Setting up SSIS Files Metadata Source



New Metadata Source wizard

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

User Name

Password

Login Url

Base Api Url

Base Api full Url

SSIS Integration Services

Warning:

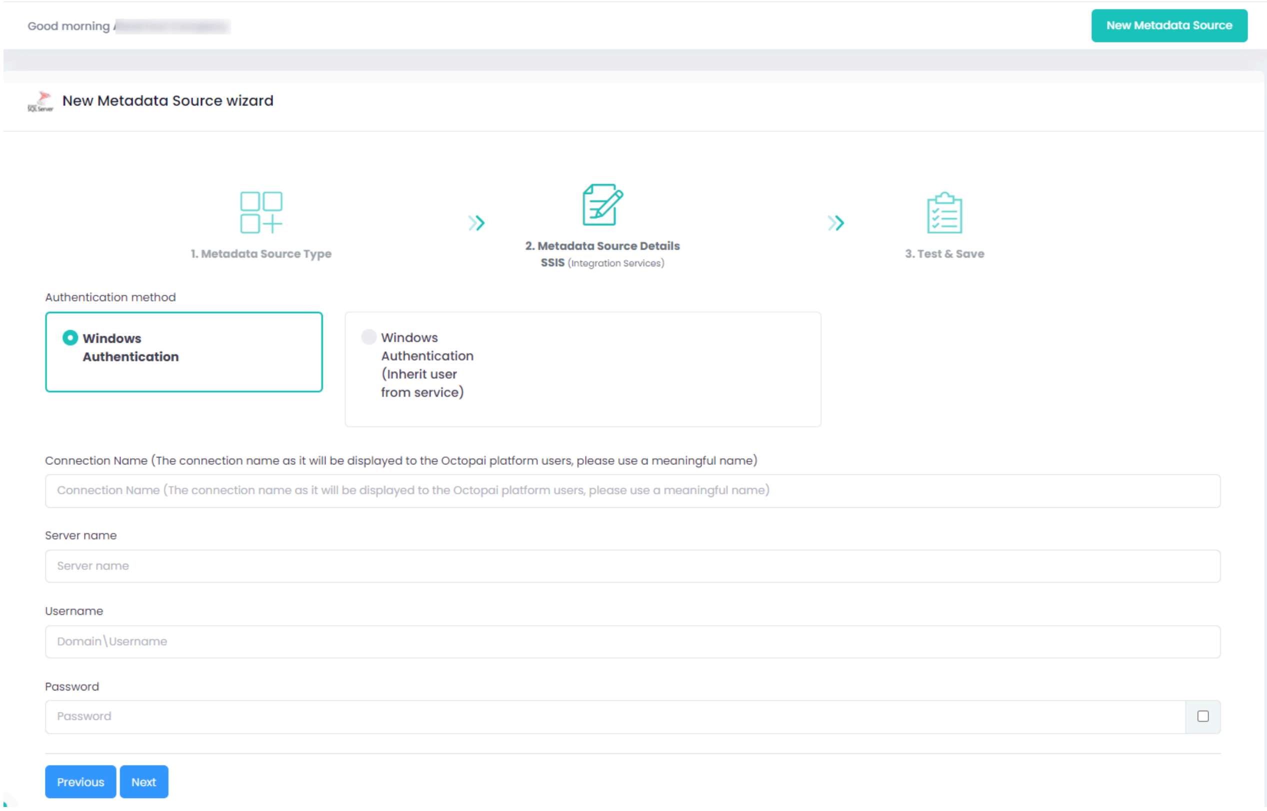
Missing permissions could end up in broken lineages.

Note:

Permissions Prerequisites (Server Type: Integration Services - MSDB):

- Map the following users to the login used for extraction:
  - Database **msdb** with Database role membership **db_datareader**.
  - Database **SSISDB** with Database role membership **public** and **ssis_admin**.

Figure 15: Setting up SSIS Integration Services Metadata Source



Good morning [redacted]

New Metadata Source wizard

Authentication method

● Windows Authentication

○ Windows Authentication (Inherit user from service)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Server name

Server name

Username

Domain\Username

Password

Password □

SSIS Integration Services Catalog

Warning:

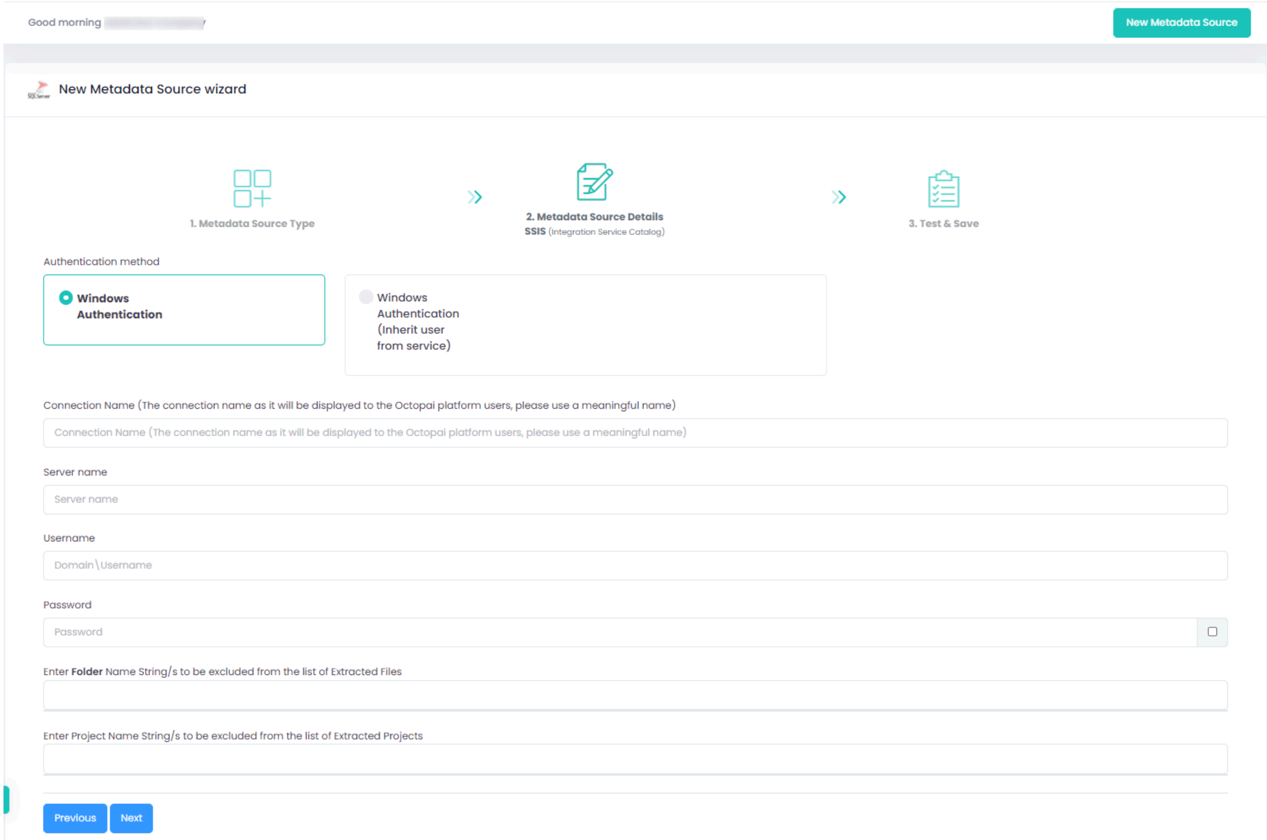
Missing permissions could end up in broken lineages.

Note:

Permissions Prerequisites:

- Grant database role membership to **ssis_admin**.
- Grant read access permission to **SSISDB.catalog.environments**.

Figure 16: Setting up SSIS Integration Services Catalog Metadata Source



Good morning [redacted]

New Metadata Source wizard

1. Metadata Source Type
2. Metadata Source Details
SSIS (Integration Service Catalog)
3. Test & Save

Authentication method

● Windows Authentication

○ Windows Authentication (Inherit user from service)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Server name

Server name

Username

Domain\Username

Password

Password □

Enter **Folder** Name String/s to be excluded from the list of Extracted Files

Enter Project Name String/s to be excluded from the list of Extracted Projects

SSIS Azure

Warning:

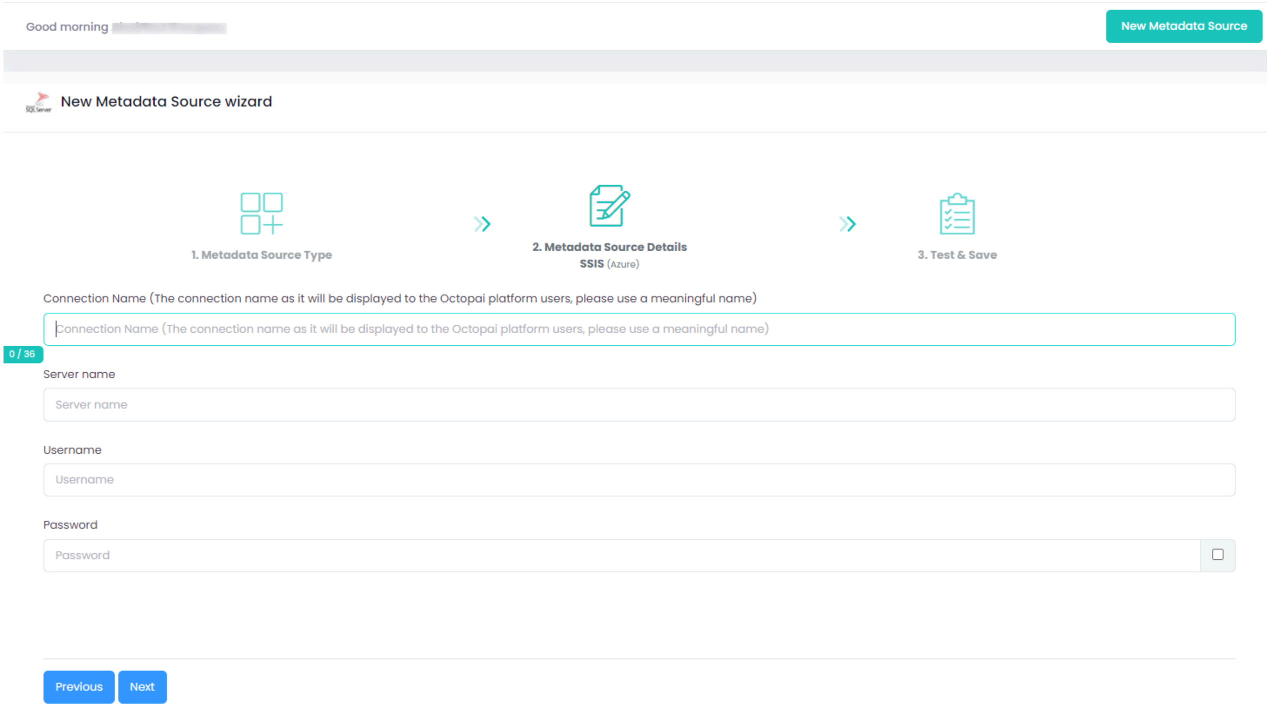
Missing permissions could end up in broken lineages.

Permissions prerequisites

- Define a user with the **db_owner** role on the SSISDB instance for the Integration Services Catalog.

Set up the metadata source

Figure 17: Configure the SSIS Azure metadata source



Good morning [redacted]

New Metadata Source wizard

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Server name

Server name

Username

Username

Password

Password

Verify the extracted metadata file

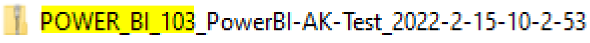
Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT folder located on the server where the Cloudera Octopai Client is installed.

   The default path is **C:\Program Files (x86)\Octopai\Service\TGT**.
2. Open the ZIP file with the connector name.

   Example:

   Figure 18: Example connector archive

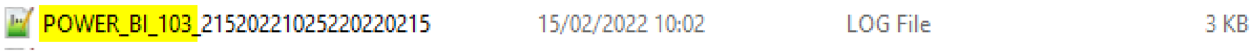
   POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53
3. Verify the quantity and quality of the inner files.

Troubleshoot

If an error occurs during the extraction, check the following:

- Confirm that the required permissions are in place.
- Send the log with the connector number and name to Cloudera Support. The default log path is **C:\Program Files (x86)\Octopai\Service\log**.

Figure 19: Example log file location



POWER_BI_103_2152022102522022015 15/02/2022 10:02 LOG File 3 KB

Azure Data Factory (ADF)

Overview of the Cloudera Octopai Azure Data Factory extraction workflow.

Note: Version supported: Version 2

Tool Permissions Prerequisites

Before proceeding, verify that you meet the prerequisites for tool permissions in Azure Data Factory.

Warning: Missing permissions could end up in broken lineages.

- A dedicated application registered with the 'Reader' role assigned to the relevant Data Factories.
- Valid 'Client Secret' for authentication credentials.

How to set up the permissions

Ensure you have the necessary permissions to set up and manage Azure Data Factory.

Step 1:

Application Setup: Quickstart: Register an app in the Microsoft identity platform

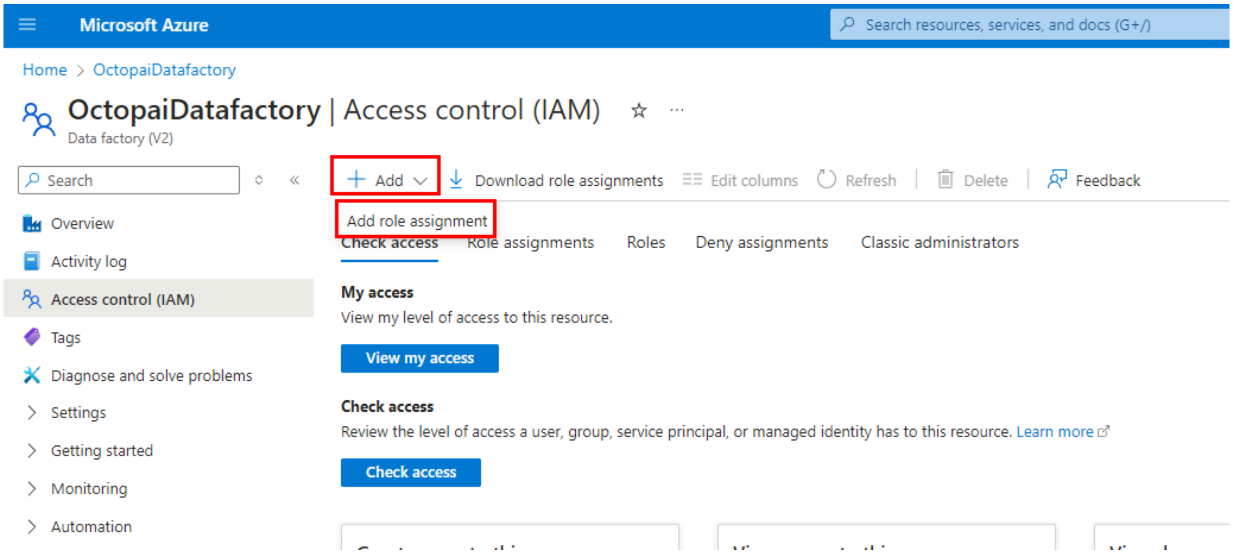
Guidelines for application setup:

- Select ‘Accounts in this organizational directory only’ when creating a new application (registration), under ‘Who can use this application or access this API?’.
- On the same page, leave 'Redirect URI empty'.
- Credentials - On the Application page, under Manage > Certificates & Secrets, use the Client Secret option for Credentials. Copy it immediately, as it won't be fully visible afterward.

Step 2:

Assign the dedicated application a 'Reader' role to the relevant Factory/ies by following the below steps:

1. Under your DataFactory, go to the 'Access control (IAM)' tab and click on 'ADD > Add role assignment'.



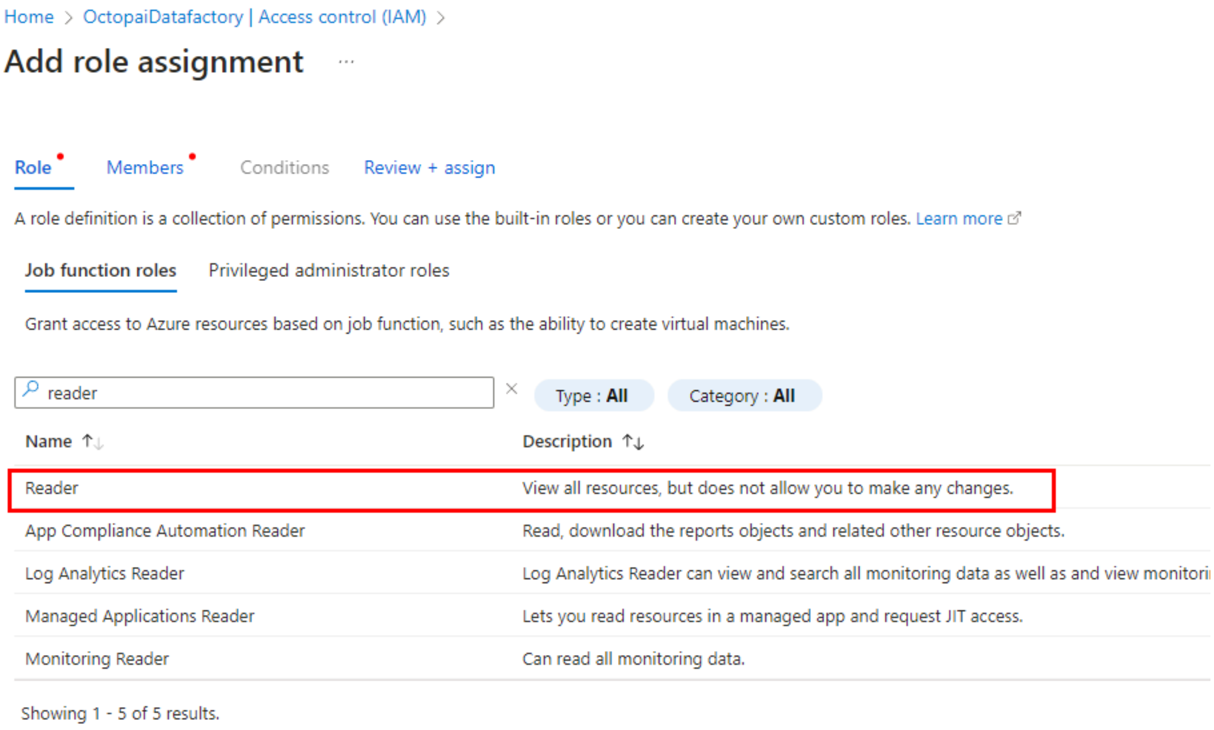
My access

View my level of access to this resource.

Check access

Review the level of access a user, group, service principal, or managed identity has to this resource. Learn more

2. Look for the 'Reader' role and click it.



Home > OctopaiDatafactory | Access control (IAM) >

Add role assignment ...

A role definition is a collection of permissions. You can use the built-in roles or you can create your own custom roles. Learn more

Grant access to Azure resources based on job function, such as the ability to create virtual machines.

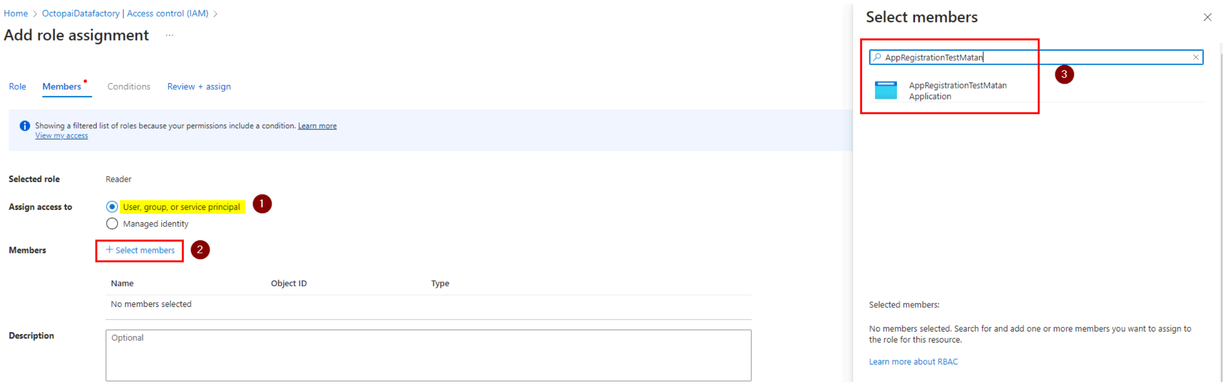
reader

Type : All Category : All

Name ↑↓	Description ↑↓
Reader	View all resources, but does not allow you to make any changes.
App Compliance Automation Reader	Read, download the reports objects and related other resource objects.
Log Analytics Reader	Log Analytics Reader can view and search all monitoring data as well as and view monitori
Managed Applications Reader	Lets you read resources in a managed app and request JIT access.
Monitoring Reader	Can read all monitoring data.

Showing 1 - 5 of 5 results.

3. Under the 'Members' tab, choose 'User, group, or service principal' and click on '+ Select members', then search for your application.



Home > OctopaiDatafactory | Access control (IAM) >

Add role assignment ...

Showing a filtered list of roles because your permissions include a condition. Learn more
View my access

Selected role: Reader

Assign access to:
● User, group, or service principal
○ Managed identity

Members: + Select members

Name	Object ID	Type
No members selected		

Description: Optional

Select members

AppRegistrationTestMatan

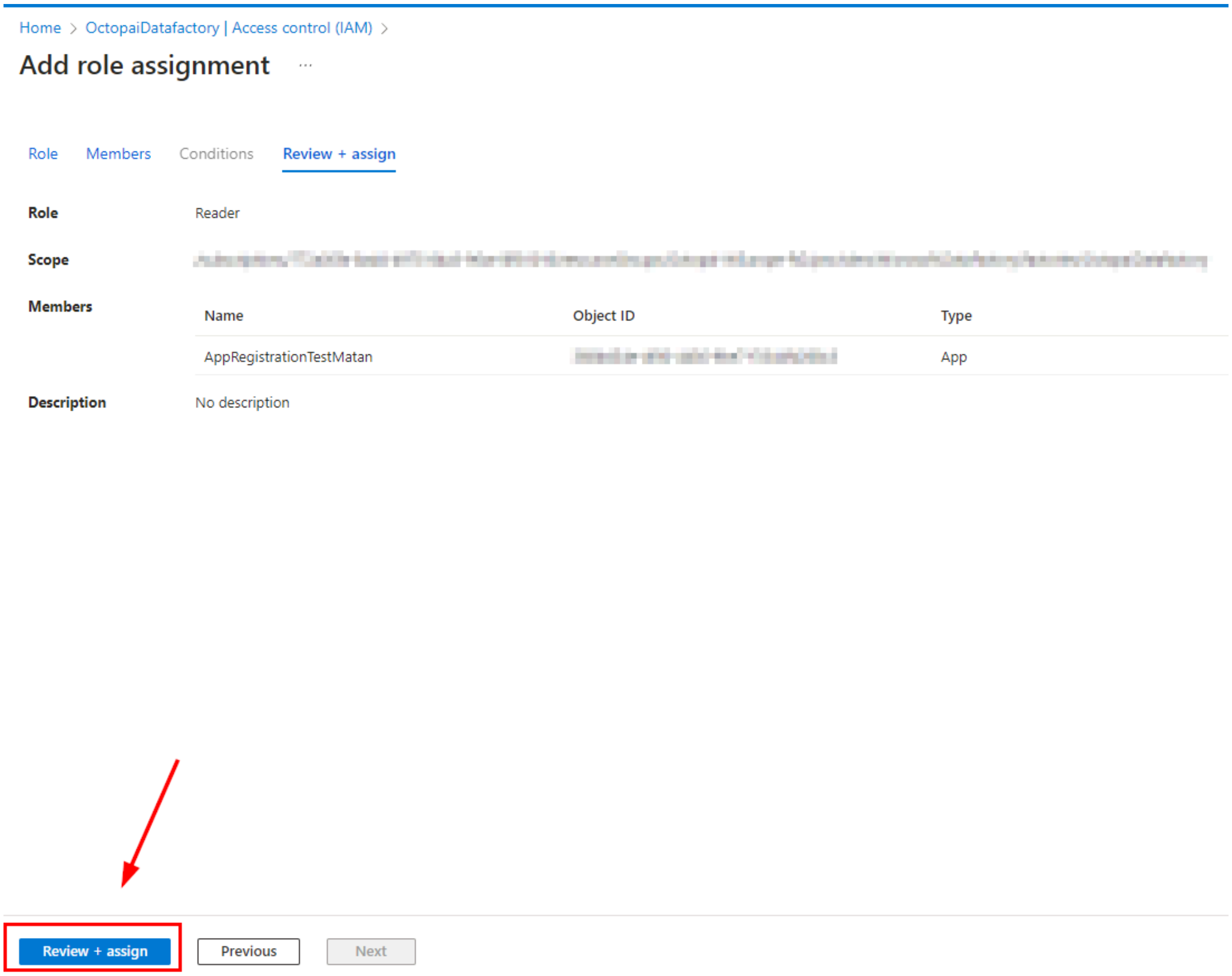
AppRegistrationTestMatan
Application

Selected members:

No members selected. Search for and add one or more members you want to assign to the role for this resource.

Learn more about RBAC

4. The last step will be to review your configuration and assign the role by clicking on 'Review + assign'.



Home > OctopaiDatafactory | Access control (IAM) >

Add role assignment ...

Role Members Conditions **Review + assign**

Role Reader

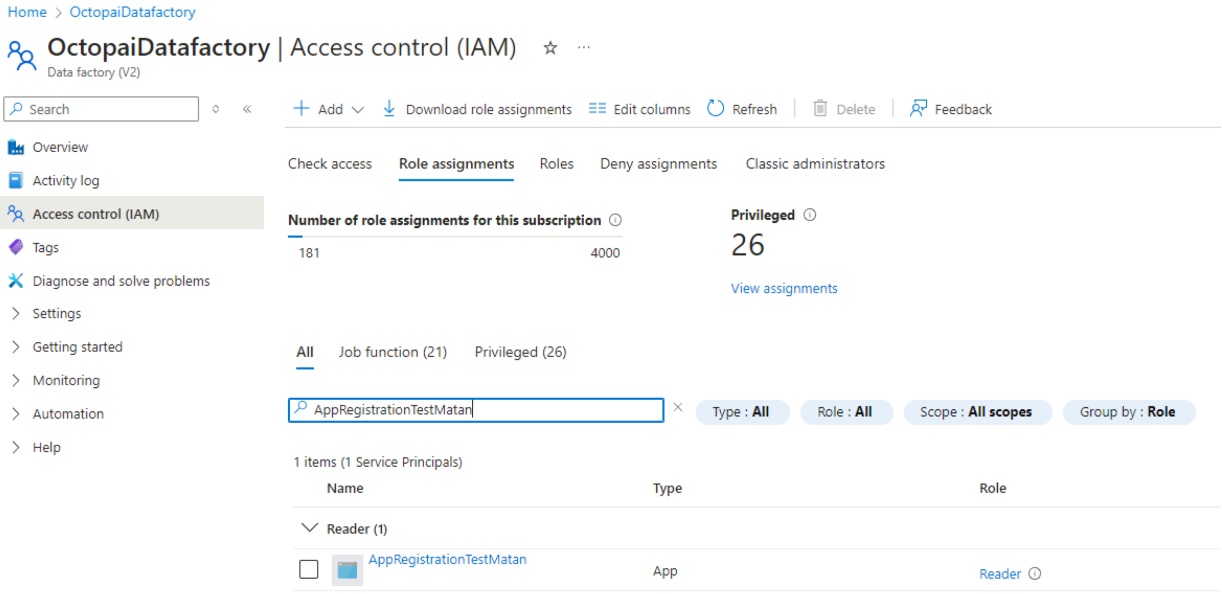
Scope [redacted]

Members

Name	Object ID	Type
AppRegistrationTestMatan	[redacted]	App

Description No description

5. After completing the previous steps, go back to your DataFactory's 'Access control (IAM)' tab > 'Role assignments'. Your application should be there.



Home > OctopaiDatafactory

OctopaiDatafactory | Access control (IAM)

Data factory (V2)

Number of role assignments for this subscription 181 4000

Privileged 26

View assignments

All Job function (21) Privileged (26)

AppRegistrationTestMatan

Type : **All** Role : **All** Scope : **All scopes** Group by : **Role**

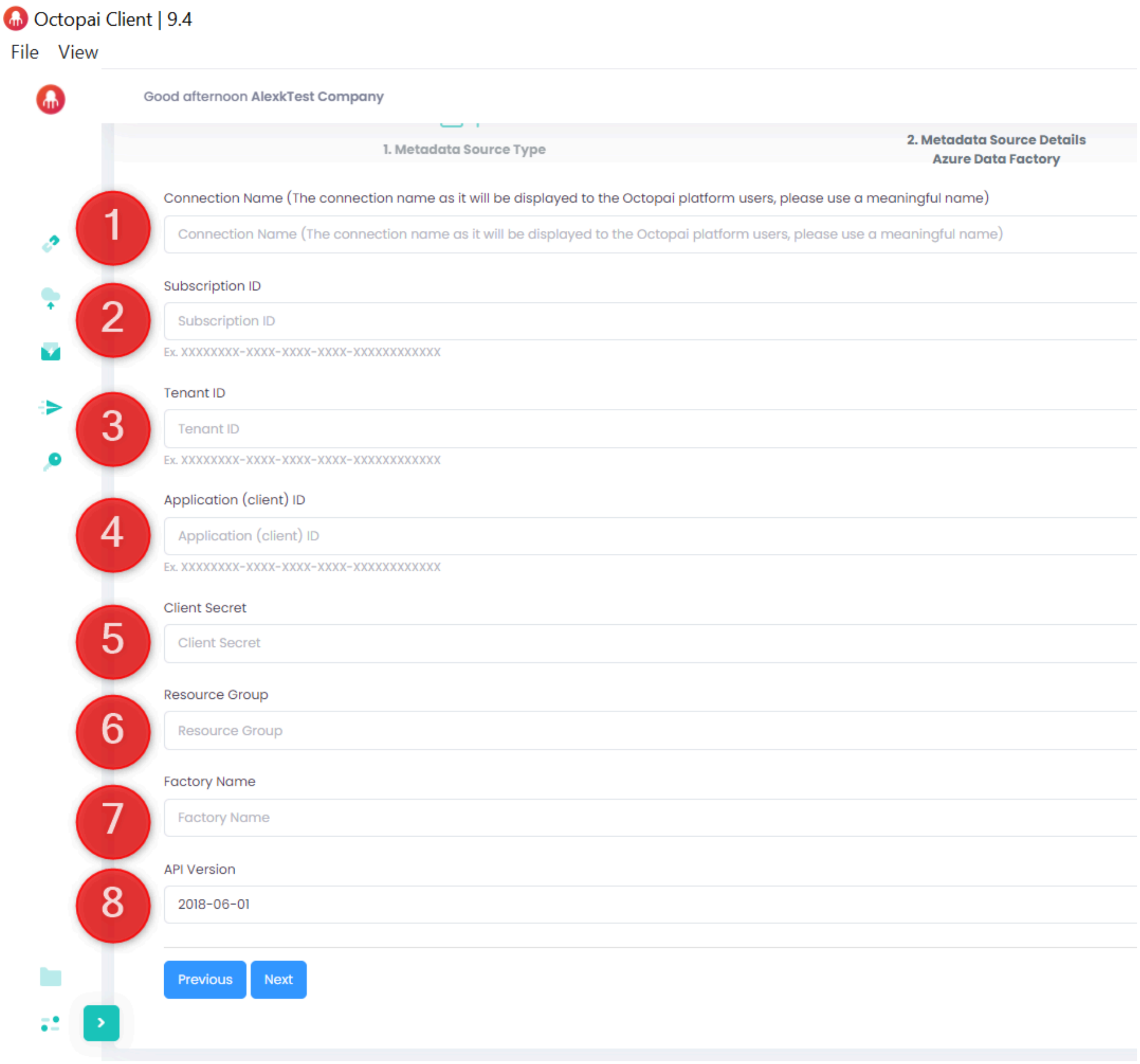
1 items (1 Service Principals)

Name	Type	Role
Reader (1)		
AppRegistrationTestMatan	App	Reader

Setting up ADF Metadata Source

Follow these instructions to configure the metadata source in Azure Data Factory.

Metadata Sources are set on the Cloudera Octopai Client:



Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Subscription ID

Subscription ID

Ex. XXXXXXXX-XXXX-XXXX-XXXX-XXXXXXXXXXXX

Tenant ID

Tenant ID

Ex. XXXXXXXX-XXXX-XXXX-XXXX-XXXXXXXXXXXX

Application (client) ID

Application (client) ID

Ex. XXXXXXXX-XXXX-XXXX-XXXX-XXXXXXXXXXXX

Client Secret

Client Secret

Resource Group

Resource Group

Factory Name

Factory Name

API Version

2018-06-01

Legend:

- Connection Name: Give a meaningful name, as it will be displayed to the Cloudera Octopai platform users.
- Subscription ID: Found in the 'Subscriptions' section of the Azure portal.
- Tenant ID: Available in the 'App registrations' section under the application you created.
- Application (Client) ID: Available in the 'App registrations' section under the application you created.
- Client Secret: Generated in 'App registrations > Certificates & secrets'.
- Resource Group: Found in the 'Resource groups' section where you created or assigned resources for your Data Factory.
- Factory Name: Listed in the 'Data Factory' section under your specific factory instance.
- API Version: Usually specified in the Azure documentation or the REST API version section related to Data Factory.

After completing all the mandatory fields, click on 'Next' > 'Finish' > and 'Run' to extract the metadata from your source.

IBM DataStage

IBM DataStage integration with Cloudera Octopai Data Lineage enables users to export and manage DataStage projects and metadata efficiently. Each project must be exported as a separate file, with specific configurations for ISX or DSX methods, ensuring compatibility with the Cloudera Octopai Client.

Note: Version Supported: up to v11.7

Tool Permissions Prerequisites

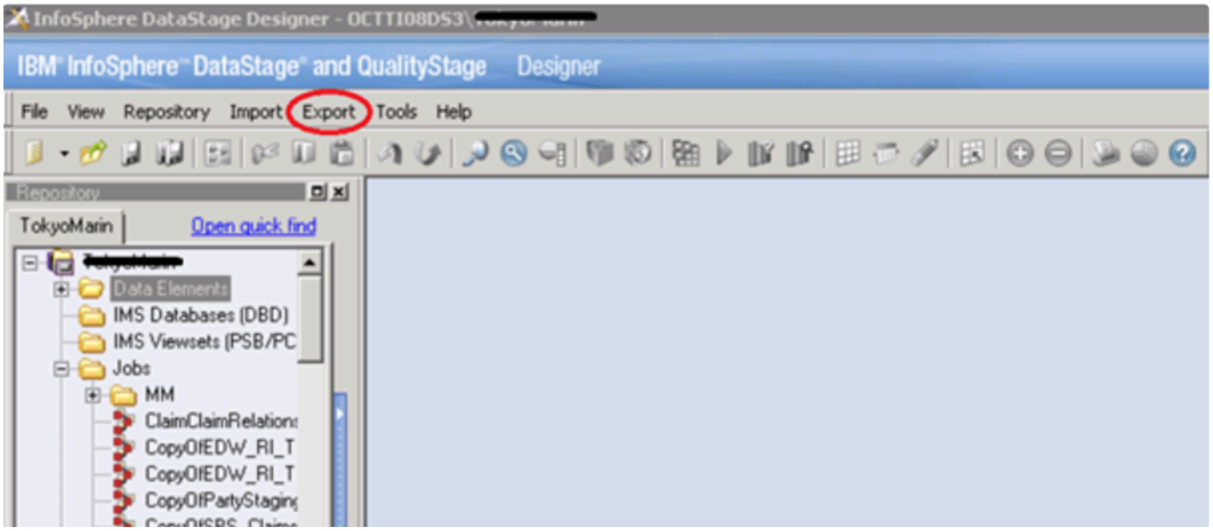
Warning: Missing permissions could end up in broken lineages.

IBM DataStage Client version 8.5 or later installed

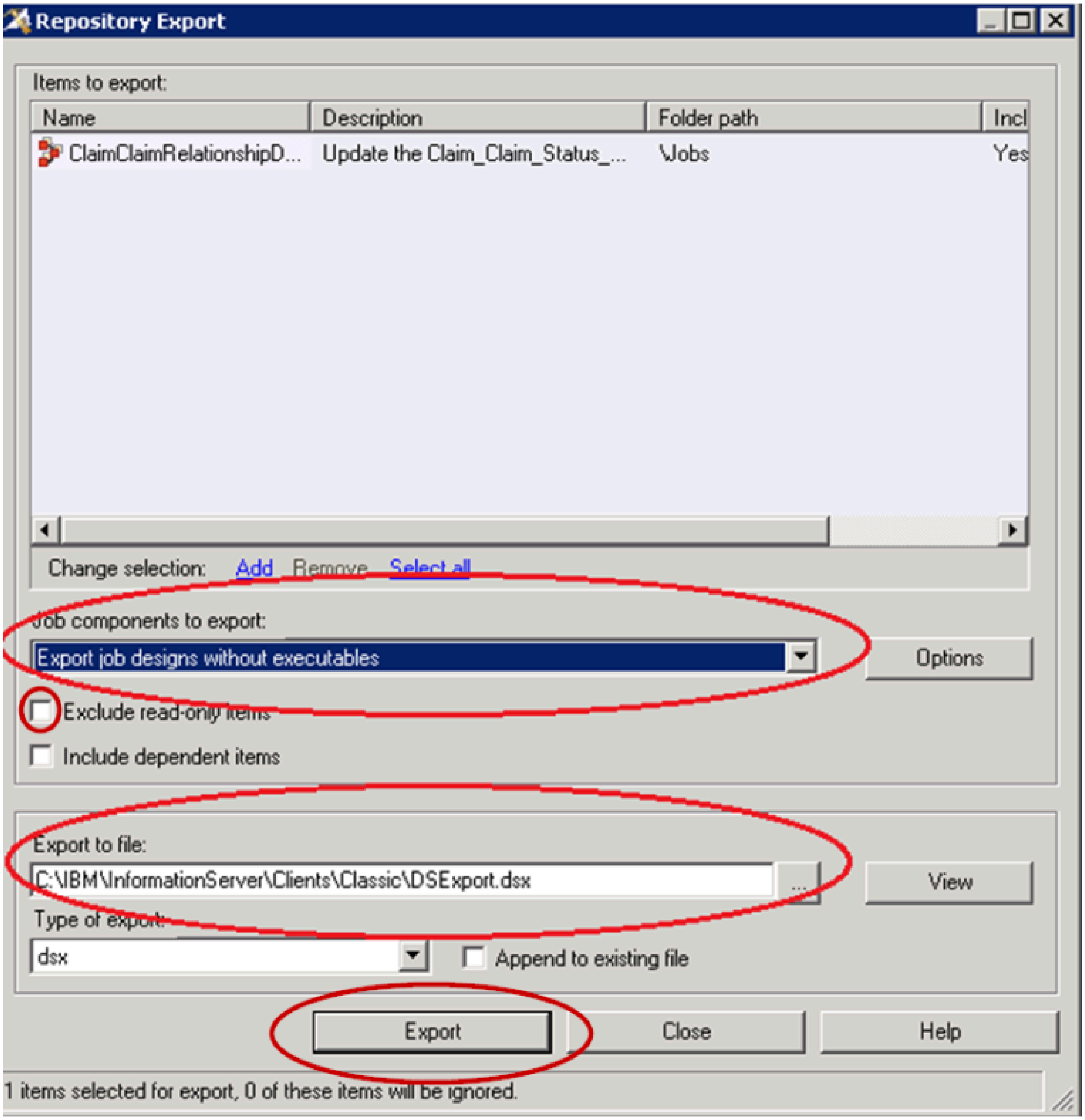
Open Server Port to DataStage Server Machine

How to export IBM DataStage jobs

1. Open DataStage Designer (Client).
2. Choose project and connect.
3. Click **Export** -> **DataStage Components** ->



4. Click **Add** and choose folder jobs from the pop window: **Select Items**.



Items to export:

Name	Description	Folder path	Incl
ClaimClaimRelationshipD...	Update the Claim_Claim_Status_...	\Jobs	Yes

Job components to export: Export job designs without executables

□ Exclude read-only items

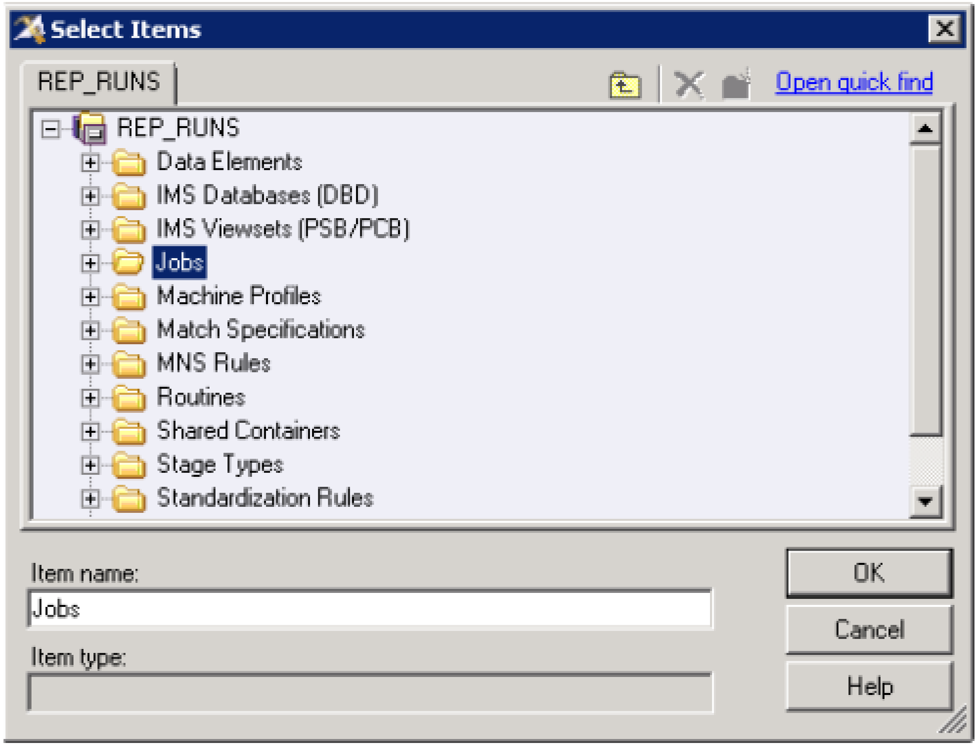
□ Include dependent items

Export to file: C:\IBM\InformationServer\Clients\Classic\DSExport.dsx

Type of export: dsx

□ Append to existing file

1 items selected for export, 0 of these items will be ignored.

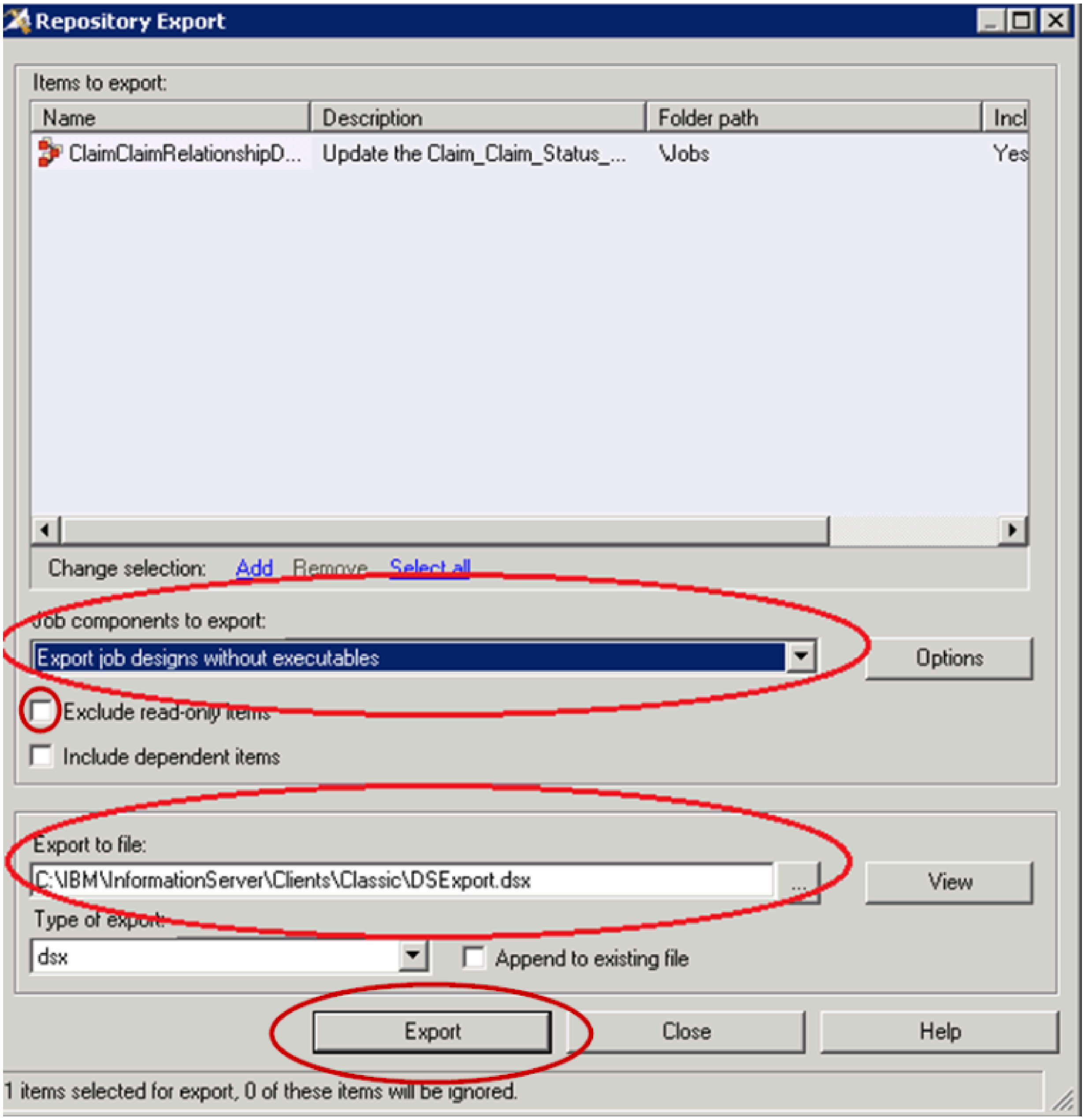


Item name: Jobs

Item type:

5. Configure destination file name and location:
    - At the drop-down list of the **Job components to export:**, choose **Export job designs without executables**.
    - Uncheck the **Exclude read-only items** option.
    - Choose the export file destination and type a name for the DSX file – **Export to file**.

6. Click the **Export** button.



Items to export:

Name	Description	Folder path	Incl
ClaimClaimRelationshipD...	Update the Claim_Claim_Status_...	\Jobs	Yes

Job components to export: Export job designs without executables

□ Exclude read-only items

□ Include dependent items

Export to file: C:\IBM\InformationServer\Clients\Classic\DSExport.dsx

Type of export: dsx

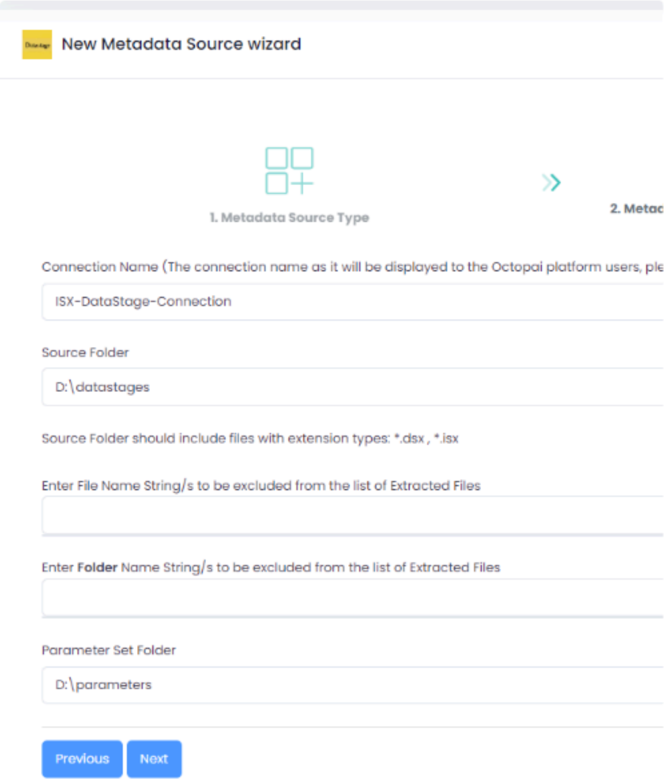
□ Append to existing file

1 items selected for export, 0 of these items will be ignored.

7. The extracted file will appear in the destination defined folder. Use this file as the source for your Cloudera Octopai metadata DataStage connection in the Cloudera Octopai Client.
8. Ensure appropriate permissions to the path to allow the Cloudera Octopai Client to access the file with the user running it.

DataStage automation requires the customer to extract the DataStage project using the following methods.

ISX Method:



Connection Name (The connection name as it will be displayed to the Octopai platform users, ple

ISX-DataStage-Connection

Source Folder

D:\datastages

Source Folder should include files with extension types: *.dsx , *.isx

Enter File Name String/s to be excluded from the list of Extracted Files

Enter **Folder** Name String/s to be excluded from the list of Extracted Files

Parameter Set Folder

D:\parameters

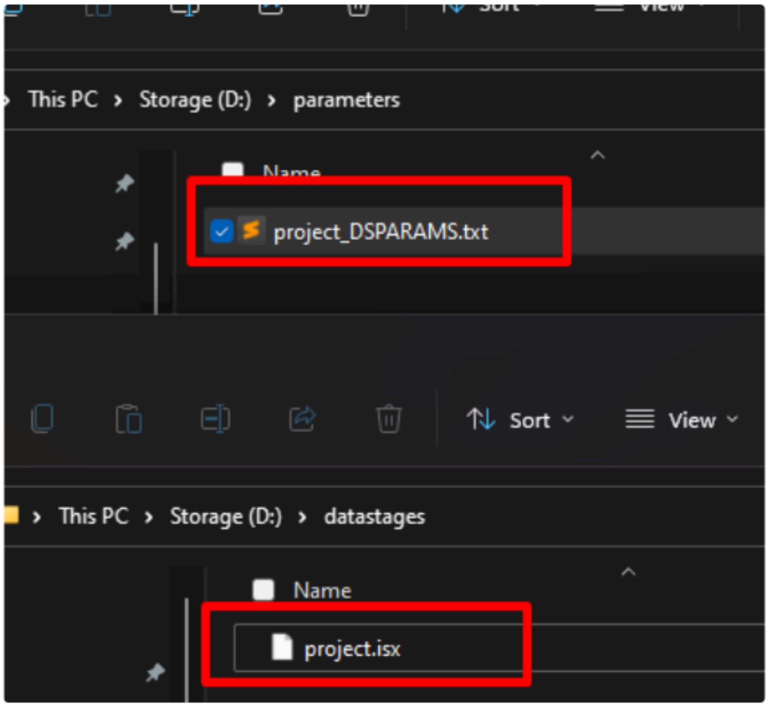
Each **project** for each **Server** needs to be in a separate zip file. Each **project** will be displayed as a different **connection name** in the Cloudera Octopai application:

Octopai Client - DataStage Module’s fields

In the Source Folder, there needs to be only **one** ISX file that resembles **one** project.

In the Parameter Set Folder, there needs to be the parameters file for that **one** project. (the format is TXT)

The Parameter Set file should follow this convention: “Isx_File_Name_DSPARAMS.txt”. (For example, if we have a project named **project.isx** , the Parameter file name will be **project_DSPARAMS.txt**)



This PC › Storage (D:) › parameters

Name

project_DSPARAMS.txt

Sort View

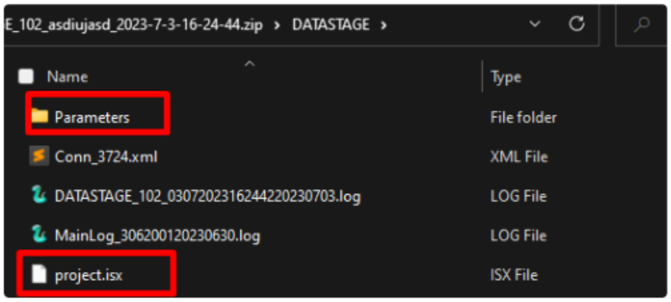
This PC › Storage (D:) › datastages

Name

project.isx

DataStage ISX file + Parameters for ISX Project File

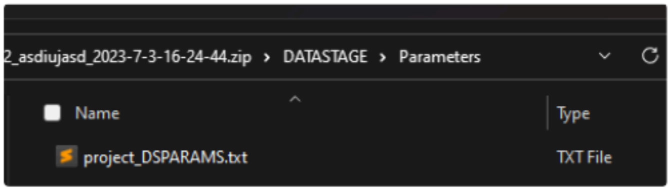
The result from this module should look like the following:



E_102_asdiujasd_2023-7-3-16-24-44.zip › DATASTAGE ›

Name	Type
Parameters	File folder
Conn_3724.xml	XML File
DATASTAGE_102_03072023162442202307 03.log	LOG File
MainLog_306200120230630.log	LOG File
project.isx	ISX File

Inside the zip file

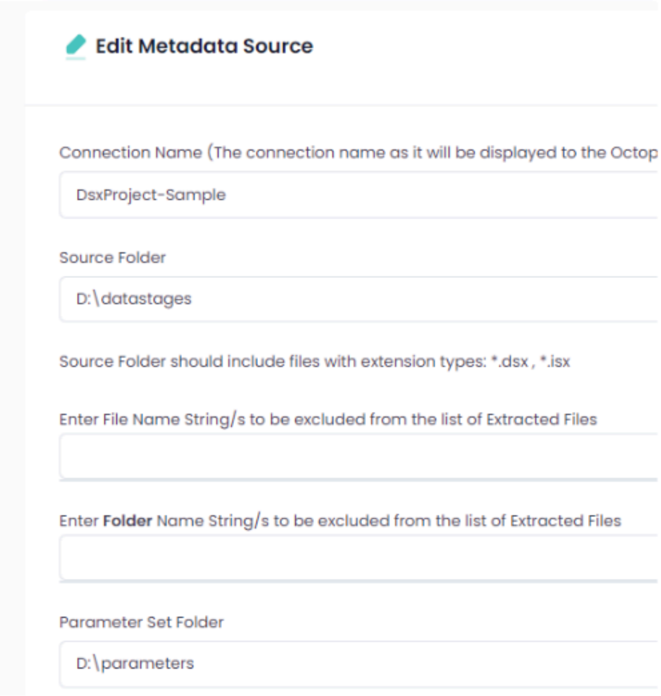


2_asdiujasd_2023-7-3-16-24-44.zip › DATASTAGE › Parameters

Name	Type
project_DSPARAMS.txt	TXT File

Inside the parameters folder

DSX Method:



Edit Metadata Source

Connection Name (The connection name as it will be displayed to the Octop

DsxProject-Sample

Source Folder

D:\datastages

Source Folder should include files with extension types: *.dsx , *.isx

Enter File Name String/s to be excluded from the list of Extracted Files

Enter **Folder** Name String/s to be excluded from the list of Extracted Files

Parameter Set Folder

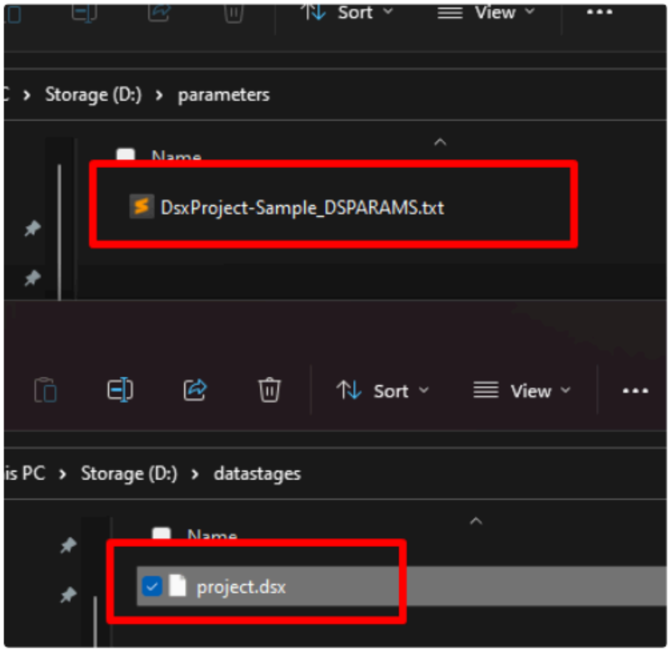
D:\parameters

Octopai Client - DataStage Module's fields

Each **project** for each **Server** needs to be in a separate zip file. The **connection name** specified in the Cloudera Octopai Client will be the **Name** of the parameter file:

In the Source Folder, there needs to be only **one** DSX file that resembles **one** project.

In the Parameter Set Folder, there needs to be the parameters file for that **one** project. (the format is TXT) The Parameter Set file should follow this convention: "ConnectionName_DSPARAMS.txt". (For example, if the **Connection Name** given while creating this connection was **DsxProject-Sample** , the Parameter file name will be **DsxProject-Sample_DSPARAMS.txt**)



› Storage (D:) › parameters

Name

DsxProject-Sample_DSPARAMS.txt

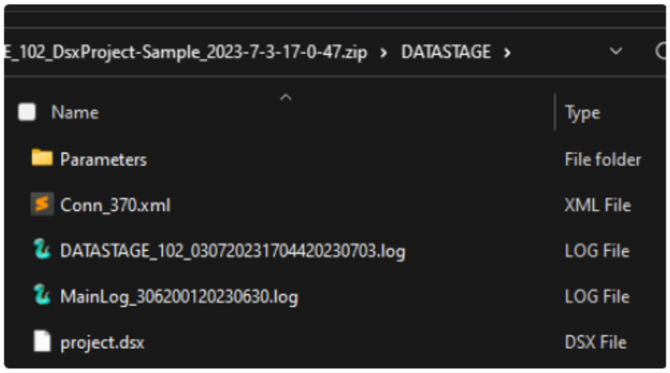
is PC › Storage (D:) › datastages

Name

project.dsx

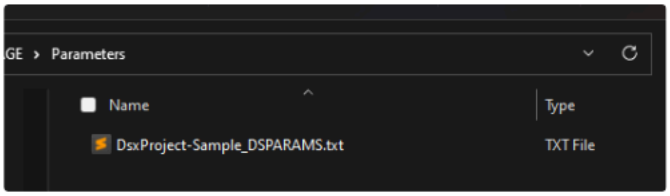
DataStage DSX file + Parameters for DSX Project File

The result from this module should look like the following:


E_102_DsxProject-Sample_2023-7-3-17-0-47.zip › DATASTAGE ›

Name	Type
Parameters	File folder
Conn_370.xml	XML File
DATASTAGE_102_03072023170442023070З.log	LOG File
MainLog_306200120230630.log	LOG File
project.dsx	DSX File

Inside the zip file

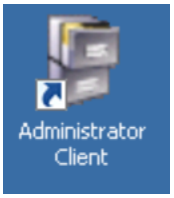

GE › Parameters

Name	Type
DsxProject-Sample_DSPARAMS.txt	TXT File

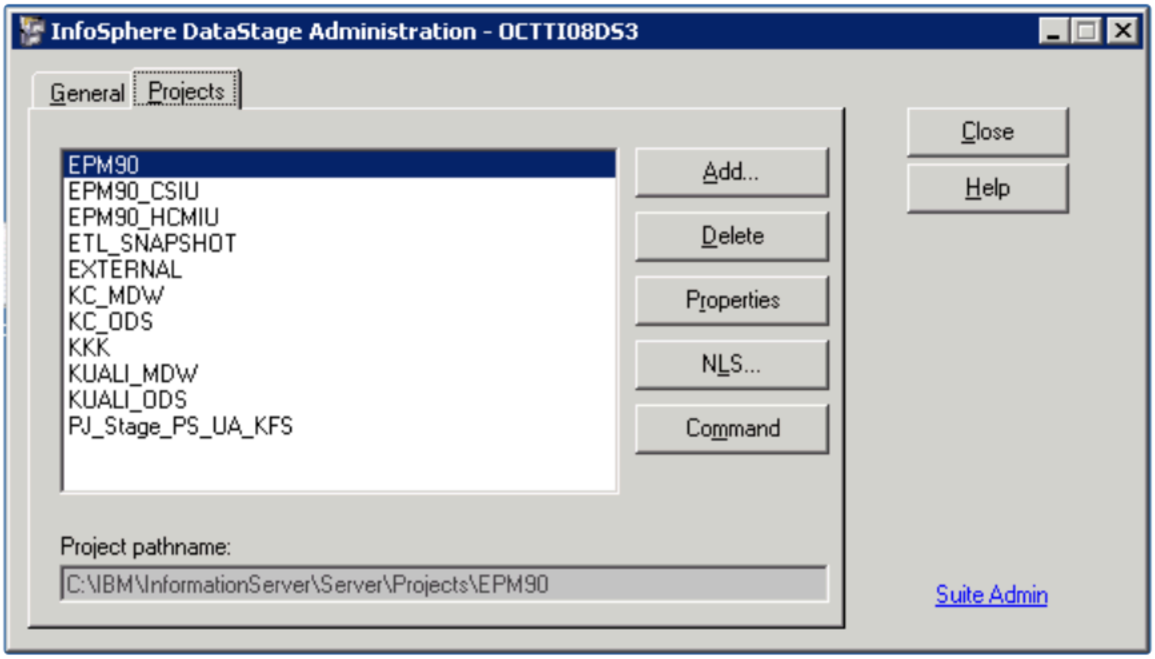
Inside the parameters folder

DataStage: Export of parameter files

1. Go to the Administrator Client.



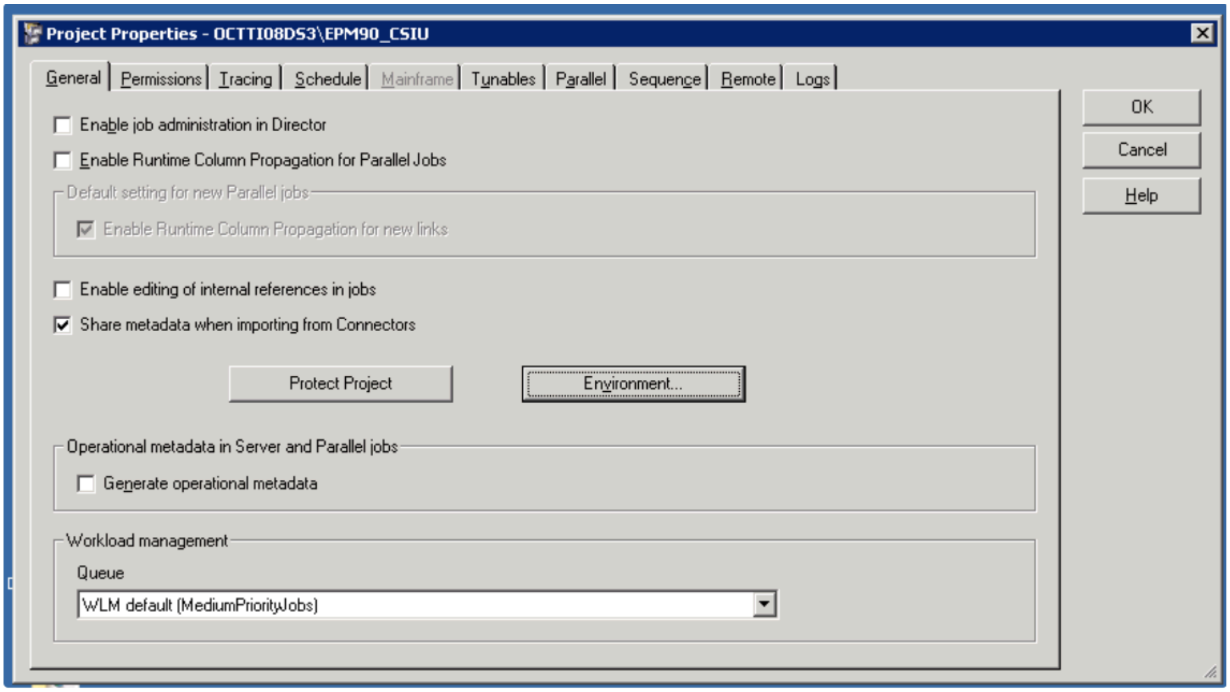
2. Choose the relevant Project and click on **Properties**.



EPM90
EPM90_CSIU
EPM90_HCMIU
ETL_SNAPSHOT
EXTERNAL
KC_MDW
KC_ODS
KKK
KUALI_MDW
KUALI_ODS
PJ_Stage_PS_UA_KFS

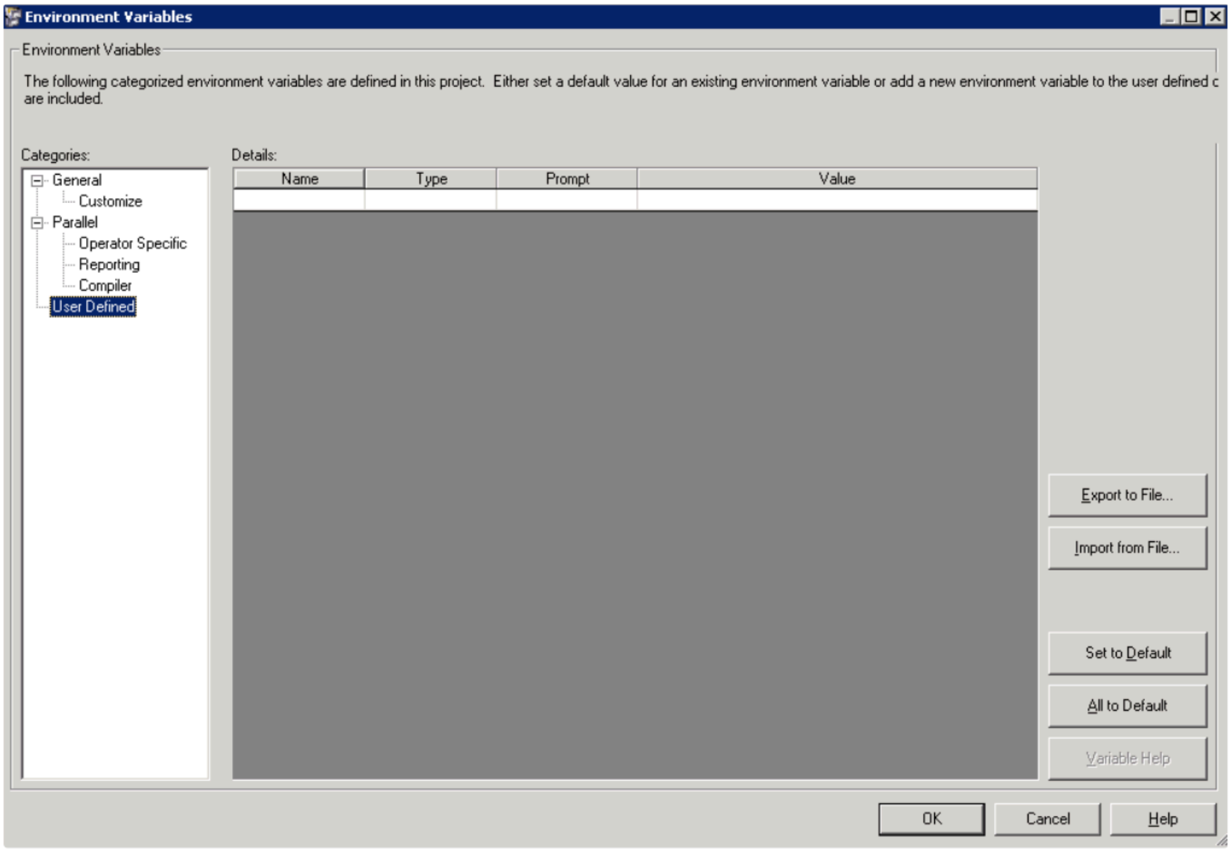
Project pathname: C:\IBM\InformationServer\Server\Projects\EPM90

3. Click on the **Environment** button.



□ Enable job administration in Director
□ Enable Runtime Column Propagation for Parallel Jobs
Default setting for new Parallel jobs
☑ Enable Runtime Column Propagation for new links
□ Enable editing of internal references in jobs
☑ Share metadata when importing from Connectors
Operational metadata in Server and Parallel jobs
□ Generate operational metadata
Workload management
Queue: WLM default (MediumPriorityJobs)

4. Go to **User Defined** and click on **Export to File**.

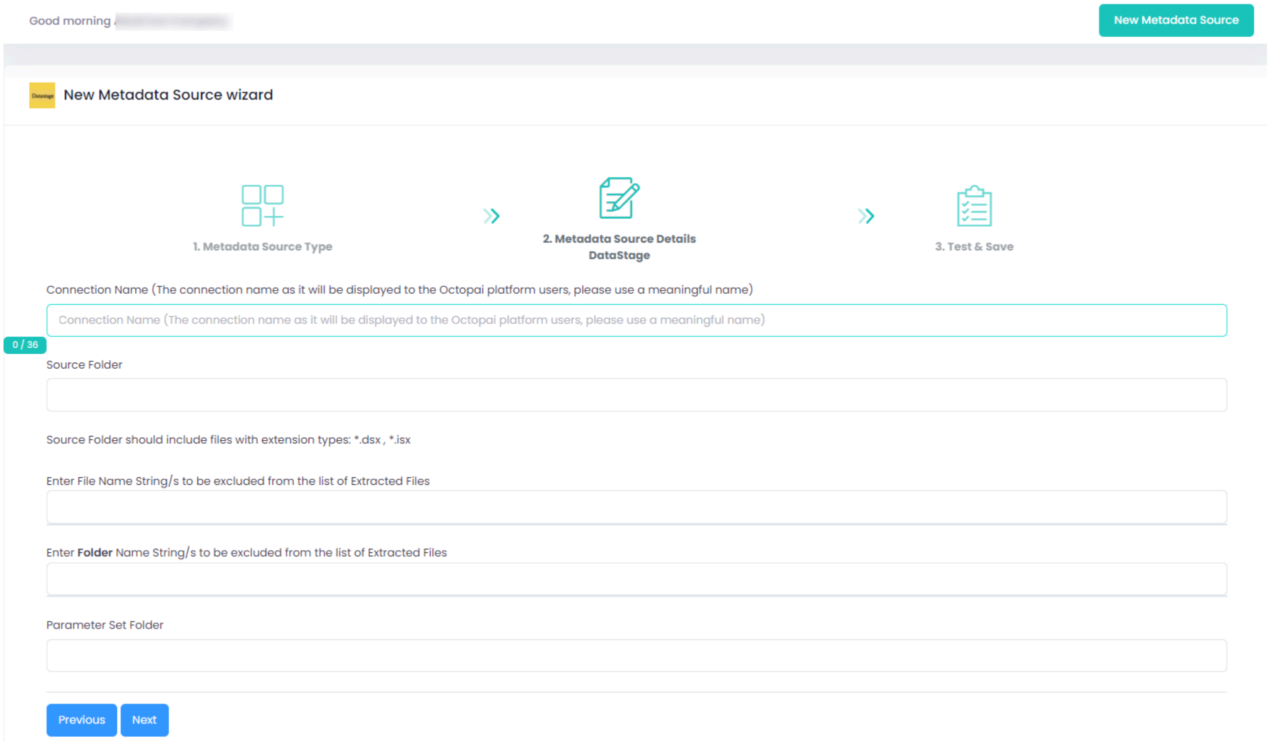


The following categorized environment variables are defined in this project. Either set a default value for an existing environment variable or add a new environment variable to the user defined c are included.

5. The new file name (.env) will be created - save it as **Name of Project _DSPARAMS.txt**.

Setting up DataStage Metadata Source

Metadata Sources are set on the Cloudera Octopai Client



Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Source Folder

Source Folder should include files with extension types: *.dsx , *.isx

Enter File Name String/s to be excluded from the list of Extracted Files

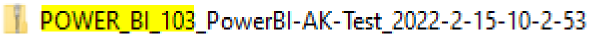
Enter **Folder** Name String/s to be excluded from the list of Extracted Files

Parameter Set Folder

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:



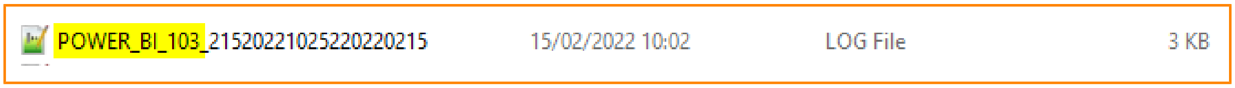
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service\log



```
POWER_BI_103_2152022102522020215    15/02/2022 10:02    LOG File    3 KB
```

Oracle ODI

Learn how to configure Oracle ODI as a metadata source for Cloudera Octopai.

Note: Version Supported: 12c

Tool Permissions Prerequisites

Warning: Missing permissions could end up in broken lineages.

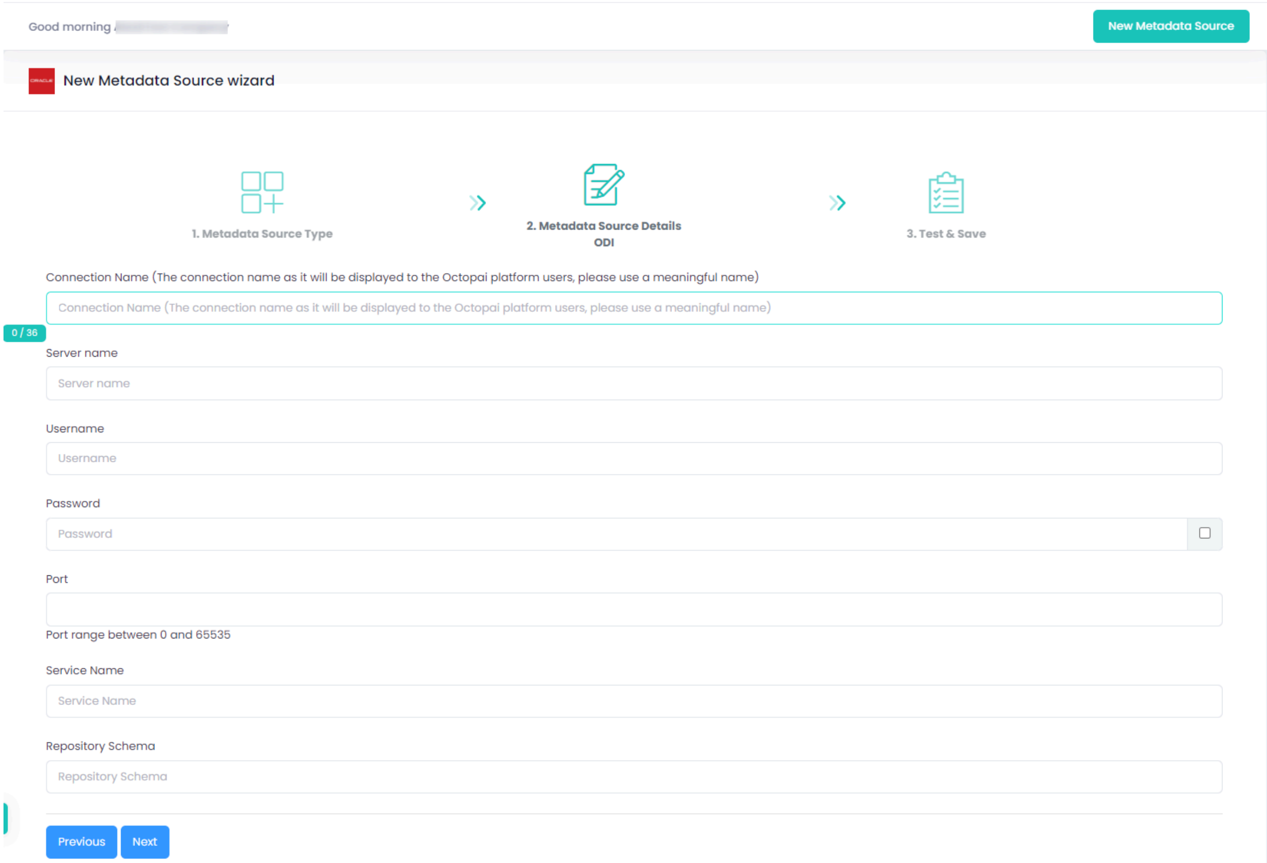
1. Open Server Port for each ODI databases repositorie

2. Existing/New Oracle user for each database with grant read permission for the following dictionary tables:

- SNP_MAP_CONN
- SNP_MAP_CP
- SNP_MAP_ATTR
- SNP_MAP_EXPR
- SNP_PACKAGE
- SNP_MAP_REF
- SNP_TABLE
- SNP_COL
- SNP_DEPLOY_SPEC
- SNP_EXEC_UNIT_GRP
- SNP_LOAD_PLAN
- SNP_EXEC_UNIT
- SNP_PHY_NODE
- SNP_PHY_EXPR
- SNP_MAP_EXPR_REF
- SNP_MAP_PROP
- SNP_SCEN
- SNP_SCEN_FOLDER
- SNP_MAPPING
- SNP_SEQUENCE
- SNP_MODEL
- SNP_FOLDER
- SNP_PROJECT
- SNP_MAP_COMP

Setting up ODI Metadata Source

Metadata Sources are set on the Cloudera Octopai Client



Good morning [redacted]

New Metadata Source wizard

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Server name

Server name

Username

Username

Password

Password □

Port

Port range between 0 and 65535

Service Name

Service Name

Repository Schema

Repository Schema

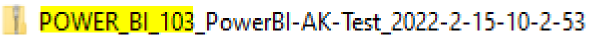
Connection Name: Giving a meaningful name to the connection will be useful for future references.

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:



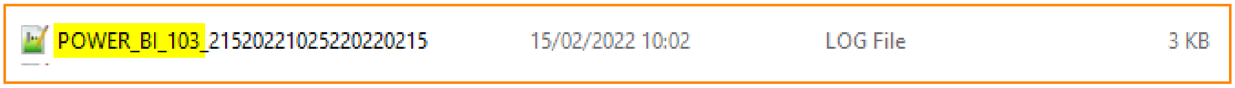
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service\log



POWER_BI_103_2152022102522022O215 15/02/2022 10:02 LOG File 3 KB

Informatica Power Center

Learn how to configure and set up Informatica Power Center as a metadata source for Cloudera Octopai Data Lineage.

Note: **Version Supported:** up to v10.5

Tool Permissions Prerequisites

Note:

Warning: Missing permissions could end up in broken lineages.

1. Open Server Port for each Informatica repository database.

2. Existing/New (Oracle/SQL) user for each database with granted read permission for the following dictionary tables:

    - OPB_SESS_EXTNS
    - OPB_MMD_SESS_EXTNS
    - OPB_SESS_CNX_REFS
    - OPB_CNX_ATTR
    - OPB_CNX
    - OPB_MMD_CNX
    - OPB_EXPRESSION
    - OPB_EXTN_ATTR
    - OPB_MAPPING
    - OPB_OBJECT_TYPE
    - OPB_SWIDGET_ATTR
    - OPB_SWIDGET_INST
    - OPB_TABLE_GROUP
    - OPB_TASK
    - OPB_TASK_ATTR
    - OPB_TASK_INST
    - OPB_TASK_VAL_LIST
    - OPB_WIDGET
    - OPB_WIDGET_INST
    - REP_ALL_MAPPINGS
    - REP_ALL_MAPPLETS
    - REP_ALL_SOURCE_FLDS
    - REP_ALL_SOURCES
    - REP_ALL_TARGET_FLDS
    - REP_ALL_TARGETS
    - REP_ALL_TASKS
    - REP_COMPONENT
    - REP_METADATA_EXTNS
    - REP_SESS_WIDGET_CNXS
    - REP_SUBJECT
    - REP_TASK_ATTR
    - REP_TASK_INST
    - REP_WIDGET_ATTR
    - REP_WIDGET_DEP
    - REP_WIDGET_FIELD
    - REP_WIDGET_INST
    - REP_WORKFLOWS

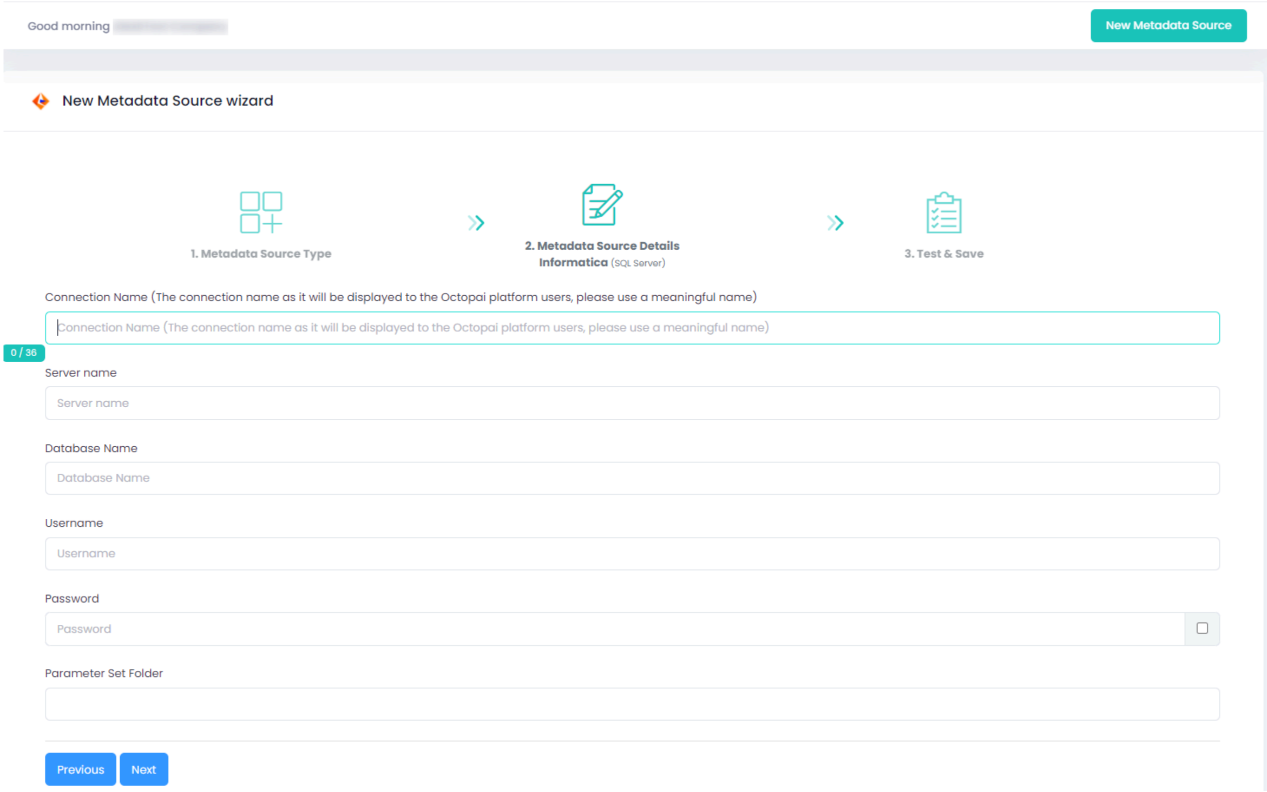
If you would like to build a loop for this action, we offer to use this guide by Aron Kumar.

Setting up Informatica Metadata Source

Metadata Sources are set on the Cloudera Octopai Client.

Note: If you're using parameters for Informatica workflows, it's recommended to organize them in a dedicated folder. To upload these parameters, simply navigate to the designated folder and select it from the "Parameter Set Folder" box.

SQL Server:



Good morning [redacted]

New Metadata Source wizard

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Server name

Server name

Database Name

Database Name

Username

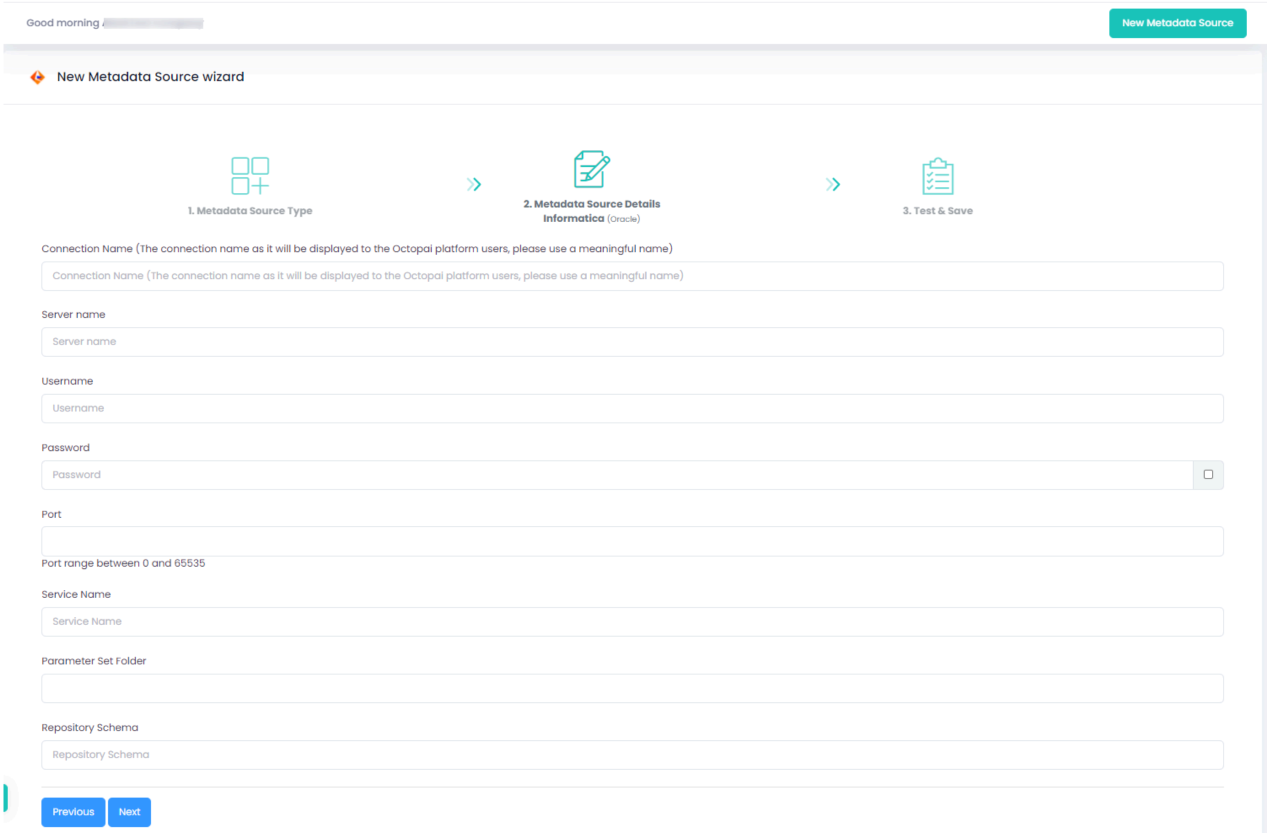
Username

Password

Password □

Parameter Set Folder

Oracle:



Good morning [redacted]

New Metadata Source wizard

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Server name

Server name

Username

Username

Password

Password □

Port

Port range between 0 and 65535

Service Name

Service Name

Parameter Set Folder

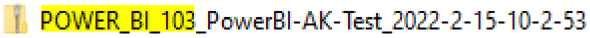
Repository Schema

Repository Schema

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:

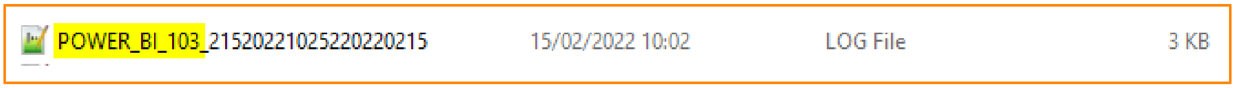
   

   POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53
3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service\log



POWER_BI_103_21520221025220220215 15/02/2022 10:02 LOG File 3 KB

Impala

How to set up and configure the Impala ODBC driver.

Tool Permissions Prerequisites

Warning: Missing permissions could result in broken connections or metadata extraction failures.

Download and Install Impala ODBC Driver

- Download the ODBC driver for Impala from your vendor's website: Impala ODBC Driver Download
- Ensure that you choose the correct driver version (**32-bit or 64-bit**) based on your environment and client tool.

Configure the DSN (Data Source Name)

1. **Open the ODBC Data Source Administrator**
   - Search for **"ODBC Data Source"** in the Start menu.
   - Select **either 32-bit or 64-bit** , depending on the installed driver.
2. **Create a New System DSN**
   - **Name:** Provide a friendly name (e.g., Impala_ODBC).
   - **Description:** (Optional) Add details (e.g., "Impala ODBC connection").
   - **Host:** Enter the **hostname** (or **IP address**) of the Impala service.
   - **Port:** Default port is **21050** for Impala (confirm in your cluster setup).
   - **Database:** Specify a default database to connect to (e.g., default).
   - **Authentication Mechanism:** Use **User Name and Password** authentication.
   - **SSL/TLS:** If required, configure SSL settings for a secure connection.

Verify the Impala ODBC Connection

- **Test the connection** and ensure the setup is successful.
- If the connection fails, verify the **host, port, database, and authentication settings** .

How to Verify the Extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Navigate to the **TGT Folder** on the server where the **Octopai Client** is installed.
    - **Default location:** C:\Program Files (x86)\Octopai\Service\TGT
2. Locate the ZIP file with the **Impala Connector Name** .
    - Example: Impala_Metadata_Export.zip
3. **Verify its content:**
    - Check the **quantity and quality** of inner files.

Troubleshoot Extraction Errors

# **Error during extraction?** Try the following steps:

- **Check permissions** on the Impala server and ODBC connection.
- **Verify the DSN configuration** (correct hostname, port, authentication).
- **Ensure that the Impala service is running** and accessible.
- **Send the log file** with the connector name and number to Cloudera Support:
    - C:\Program Files (x86)\Octopai\Service\log

Setting up Apache Impala Connector with Kerberos Authentication

Learn how to configure the Apache Impala connector in Cloudera Octopai Client using Kerberos authentication.

Before you begin

Before configuring the Apache Impala connector in Cloudera Octopai, ensure the following components are available and properly configured:

- **MIT Kerberos for Windows:** Download and install from the official MIT Kerberos download page. The default installation path is C:\Program Files\MIT\Kerberos\. To verify the installation, ensure that the following executable file exists in your environment: C:\Program Files\MIT\Kerberos\bin\kinit.exe. This path is configured by default in the kerberos.settings.json file used by the Cloudera Octopai Client.
- **Kerberos Configuration File (**krb5.ini**):** Obtain this file from the Hadoop or Impala cluster administrator and place it under C:\ProgramData\MIT\Kerberos5\krb5.ini. The configuration must include the following sections and values adjusted to the actual cluster environment:

```
[libdefaults]
                        default_realm = ROOT.COMOPS.SITE
                        dns_lookup_realm = false
                        dns_lookup_kdc = false
                        ticket_lifetime = 24h
                        renew_lifetime = 7d
                        forwardable = true

                        [realms]
                        ROOT.COMOPS.SITE = {
                        kdc = ccycloud-1.cdp.root.comops.site
                        admin_server = ccycloud-1.cdp.root.comops.site
                        }

                        [domain_realm]
                        .root.comops.site = ROOT.COMOPS.SITE
```

```
                                root.comops.site = ROOT.COMOPS.SITE
```

- **Kerberos Keytab File:** Obtain the keytab file from the Hadoop or Hive cluster administrator. The keytab contains encrypted credentials used for Kerberos authentication and enables non-interactive authentication. The file format is binary with .keytab extension. Securely store the keytab file in a location accessible to the Cloudera Octopai Client, for example, at C:\Octopai\keytabs\impala.keytab.
- **Impala ODBC Driver:** Download and install an Impala ODBC driver that supports Kerberos authentication. Driver options include the Cloudera or Hortonworks Impala ODBC driver, or a vendor-specific equivalent. Ensure the driver architecture (32-bit or 64-bit) matches the Cloudera Octopai Client installation.

Procedure

1. Install and configure the Impala ODBC driver.

a) Open the ODBC Data Source Administrator:

- Search for "ODBC Data Source" in the Windows Start menu.
- Select either 32-bit or 64-bit, depending on the installed driver.

b) Create a new system DSN:

- Navigate to the System DSN tab and click Add.
- Select the Impala ODBC driver (for example, Cloudera ODBC Driver for Impala) and click Finish.

c) Configure the DSN basic settings:

- **Data Source Name:** Provide a user-friendly name (for example, Impala_Kerberos_Prod).
- **Description:** (Optional) Add details (for example, "Production Impala with Kerberos authentication").
- **Host:** Enter the hostname or IP address of the Impala service (for example, ccycloud-1.cdp.root.comops.site).
- **Port:** Default port is 21050. Confirm the correct value in the cluster configuration.
- **Database:** Specify a default database to connect to (for example, default).

d) Configure the Kerberos authentication parameters:

- **Authentication Mechanism:** Kerberos
- **Service Name:** impala
- **Realm:** Kerberos realm (for example, ROOT.COMOPS.SITE)
- **Host FQDN:** Fully qualified domain name of the Impala service host
- **Kerberos Configuration Path:** C:\ProgramData\MIT\Kerberos5\krb5.ini

e) Configure SSL or TLS (Optional):

If the cluster requires secure connectivity, enable SSL, and configure the truststore path and password according to the security policy of the cluster.

f) Configure advanced settings if required by the environment:

- Connection timeout values
- Thrift transport, typically SASL for Kerberos
- Native query execution

g) Test and save the DSN:

- Test the connection by clicking Test. A valid Kerberos ticket may be required for the test to succeed.
- If the test is successful, click OK to save the DSN.

2. Configure Cloudera Octopai Client for Impala with Kerberos.
   a) Add an Impala connection to Cloudera Octopai Client:

      Launch the Cloudera Octopai Client and click Add New Connection or use the connection wizard.

   b) Configure the Impala connection parameters:
      - **Connection Name:** Provide a descriptive name (for example, Production Impala Kerberos)
      - Authentication settings:
        - **Authentication Type:** Kerberos (Kerberos-specific fields will be displayed)
        - **Kerberos Principal:** for example, impala@ROOT.COMOPS.SITE
        - **Keytab Path:** Full path to the keytab file (for example, C:\Octopai\keytabs\impala.keytab)
      - If using ODBC, specify the DSN name created earlier (for example, Impala_Kerberos_Prod)
3. Test the connection.
   - Click Test Connection.
   - During testing, the Cloudera Octopai Client performs the following steps:
     - Acquire a Kerberos ticket using kinit and the provided keytab
     - Attempt to connect to Impala
     - Display the connection status

If the test fails, check the error message and verify the following:

- Kerberos configuration
- DSN settings
- Service availability
- File permissions

Apache Hive

Learn how to set up and configure Apache Hive ODBC connections, including driver installation, DSN creation, and metadata verification.

Before you begin

Warning: Missing permissions could result in broken connections or metadata extraction failures.

Procedure

1. Download and Install Hive ODBC Driver.
   a) Download the ODBC driver for Hive from the Hive ODBC Driver Download website.
   b) Ensure that you choose the correct driver version, either 32-bit or 64-bit, based on your environment and client tool.

2. Configure the DSN (Data Source Name).
   a) Open the ODBC Data Source Administrator.
      1. Search for ODBC Data Source in the Start menu.
      2. Select either 32-bit or 64-bit, depending on the installed driver.

   b) Create a New System DSN.
      1. Provide a friendly name, for example Hive_ODBC.
      2. (Optional) Add details, for example Hive ODBC connection.
      3. Enter the hostname or IP address of the Hive service.
      4. Set the port. The default port is 10000 for Hive but you must confirm the value in your cluster setup.
      5. Specify a default database, for example default.
      6. Use the User Name and Password authentication method.
      7. If required, configure SSL settings for a secure connection.
3. Verify the extracted metadata file. Access the Cloudera Octopai Target Folder (TGT) and troubleshoot issues as needed.
   a) Navigate to the TGT Folder on the server where the Cloudera Octopai Client is installed.

      The default location is C:\Program Files (x86)\Octopai\Service\TGT.
   b) Locate the ZIP file with the Hive Connector name.

      For example, Hive_Metadata_Export.zip.
   c) Verify the file content by checking the quantity and quality of the included files.

If an error occurred during the extraction, perform the following troubleshooting steps:

1. Check permissions on the Hive server and ODBC connection.
2. Verify the DSN configuration, including the correct hostname, port, and authentication.
3. Send the log file with the connector name and number to Cloudera Support.

   You can find the log at C:\Program Files (x86)\Octopai\Service\log.

Setting up Apache Hive Connector with Kerberos Authentication

Learn how to configure the Apache Hive connector in Cloudera Octopai Client using Kerberos authentication.

Before you begin

Before configuring the Apache Hive connector in Cloudera Octopai, ensure the following components are available and properly configured:

- **MIT Kerberos for Windows:** Download and install from the official MIT Kerberos download page. The default installation path is C:\Program Files\MIT\Kerberos\. To verify the installation, ensure that the following executable file exists in your environment: C:\Program Files\MIT\Kerberos\bin\kinit.exe. This path is configured by default in the kerberos.settings.json file used by the Cloudera Octopai Client.
- **Kerberos Configuration File (**krb5.ini**):** Obtain this file from the Hadoop or Hive cluster administrator and place it under C:\ProgramData\MIT\Kerberos5\krb5.ini. The configuration must include the following sections and values adjusted to the actual cluster environment:

  ```
  [libdefaults]
                          default_realm = ROOT.COMOPS.SITE
                          dns_lookup_realm = false
                          dns_lookup_kdc = false
                          ticket_lifetime = 24h
                          renew_lifetime = 7d
                          forwardable = true
  ```

```
[realms]
ROOT.COMOPS.SITE = {
kdc = ccycloud-1.cdp.root.comops.site
admin_server = ccycloud-1.cdp.root.comops.site
}

[domain_realm]
.root.comops.site = ROOT.COMOPS.SITE
root.comops.site = ROOT.COMOPS.SITE
```

- **Kerberos Keytab File:** Obtain the keytab file from the Hadoop or Hive cluster administrator. The keytab contains encrypted credentials used for Kerberos authentication and enables non-interactive authentication. The file format is binary with .keytab extension. Securely store the keytab file in a location accessible to the Cloudera Octopai Client, for example, at C:\Octopai\keytabs\hive.keytab.
- **Hive ODBC Driver:** Download and install a Hive ODBC driver that supports Kerberos authentication. Driver options include the Cloudera or Hortonworks Hive ODBC driver, or a vendor-specific equivalent. Ensure the driver architecture (32-bit or 64-bit) matches the Cloudera Octopai Client installation.

Procedure

1. Install and configure the Hive ODBC driver.
   a) Open the ODBC Data Source Administrator:
      - Search for "ODBC Data Source" in the Windows Start menu.
      - Select either 32-bit or 64-bit, depending on the installed driver.

   b) Create a new system DSN:
      - Navigate to the System DSN tab and click Add.
      - Select the Hive ODBC driver (for example, Cloudera ODBC Driver for Hive) and click Finish.

   c) Configure the DSN basic settings:
      - **Data Source Name:** Provide a user-friendly name (for example, Hive_Kerberos_Prod).
      - **Description:** (Optional) Add details (for example, "Production Hive with Kerberos authentication").
      - **Host:** Enter the hostname or IP address of the HiveServer2 service (for example, ccycloud-1.cdp.root.comops.site).
      - **Port:** Default port is 10000. Confirm the correct value in the cluster configuration.
      - **Database:** Specify a default database to connect to (for example, default).

   d) Configure the Kerberos authentication parameters:
      - **Authentication Mechanism:** Kerberos
      - **Service Name:** hive
      - **Realm:** Kerberos realm (for example, ROOT.COMOPS.SITE)
      - **Host FQDN:** Fully qualified domain name of the HiveServer2 service host
      - **Kerberos Configuration Path:** C:\ProgramData\MIT\Kerberos5\krb5.ini

   e) Configure SSL or TLS (Optional):

      If the cluster requires secure connectivity, enable SSL, and configure the truststore path and password according to the security policy of the cluster.

   f) Configure advanced settings if required by the environment:
      - Connection timeout values
      - Thrift transport, typically SASL for Kerberos
      - Native query execution

   g) Test and save the DSN:
      - Test the connection by clicking Test. A valid Kerberos ticket may be required for the test to succeed.
      - If the test is successful, click OK to save the DSN.

2. Configure Cloudera Octopai Client for Hive with Kerberos.
   a) Add a Hive connection to Cloudera Octopai Client:

      Launch the Cloudera Octopai Client and click Add New Connection or use the connection wizard.

   b) Configure the Hive connection parameters:

      - **Connection Name:** Provide a descriptive name (for example, Production Hive Kerberos)
      - Authentication settings:
        - **Authentication Type:** Kerberos (Kerberos-specific fields will be displayed)
        - **Kerberos Principal:** for example, hive@ROOT.COMOPS.SITE
        - **Keytab Path:** Full path to the keytab file (for example, C:\Octopai\keytabs\hive.keytab)
      - If using ODBC, specify the DSN name created earlier (for example, Hive_Kerberos_Prod)
3. Test the connection.
   - Click Test Connection.
   - During testing, the Cloudera Octopai Client performs the following steps:
     - Acquire a Kerberos ticket using kinit and the provided keytab
     - Attempt to connect to Hive
     - Display the connection status

If the test fails, check the error message and verify the following:

- Kerberos configuration
- DSN settings
- Service availability
- File permissions

DB2 iSeries

Configure Cloudera Octopai Data Lineage to extract metadata from IBM DB2 iSeries, from client setup through ODBC configuration and required permissions.

Tool Permissions Prerequisites

- Warning: Missing permissions could end up in broken lineages.
- Microsoft Visual C++ Redistributable 2013 x64: Essential for running applications developed with Visual C++ on Windows systems. Download from this link.
- IBM i Access Client Solutions: A suite for accessing and managing IBM iSeries systems.
- Network Configuration: Ensure the server port is open for each DB2 iSeries database connection to facilitate communication.

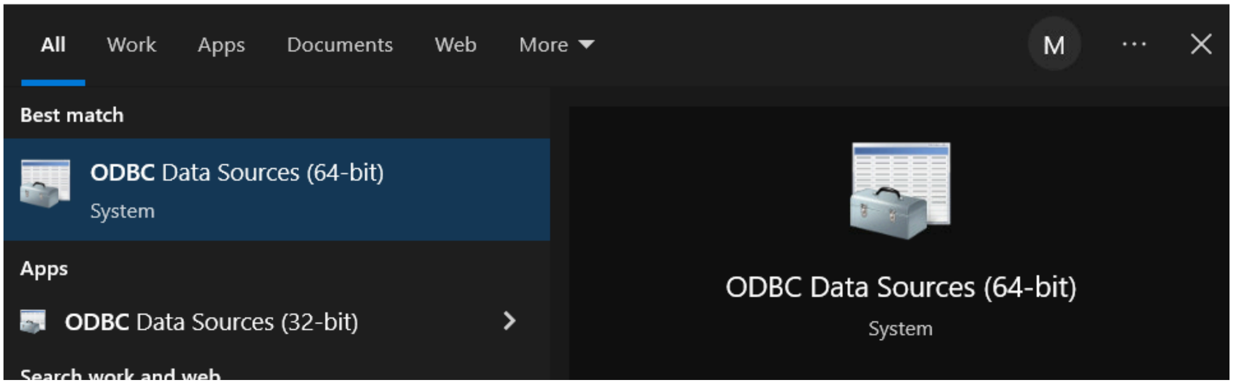
Configuring ODBC Data Sources for DB2 iSeries

IBM i Access Client Solutions Installation:

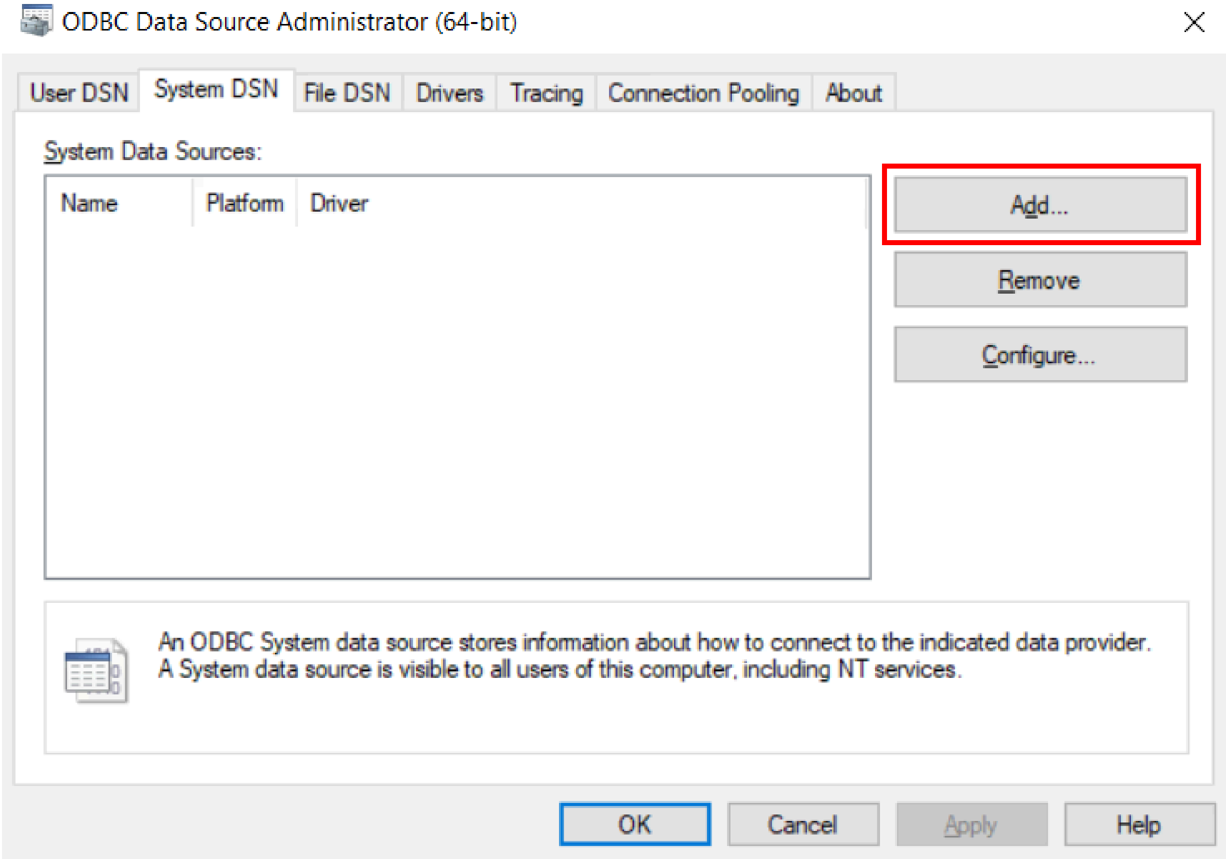
1. Navigate to the IBM website's login page, and log in using your IBM ID and password.
2. Download the "ACS Windows App Pkg English (64bit)".
3. Execute the installer located at IBMiAccess_v1r1_WindowsAP_English\Image64a\setup.exe to extract and install the ACS application.

Post-installation Configuration:

1. Access the ODBC Data Source Administrator by navigating through Control Panel > Administrative Tools > ODBC Data Sources (64-bit).


All
Work
Apps
Documents
Web
More
M
Best match
ODBC Data Sources (64-bit)
System
Apps
ODBC Data Sources (32-bit)
Search work and web
ODBC Data Sources (64-bit)
System

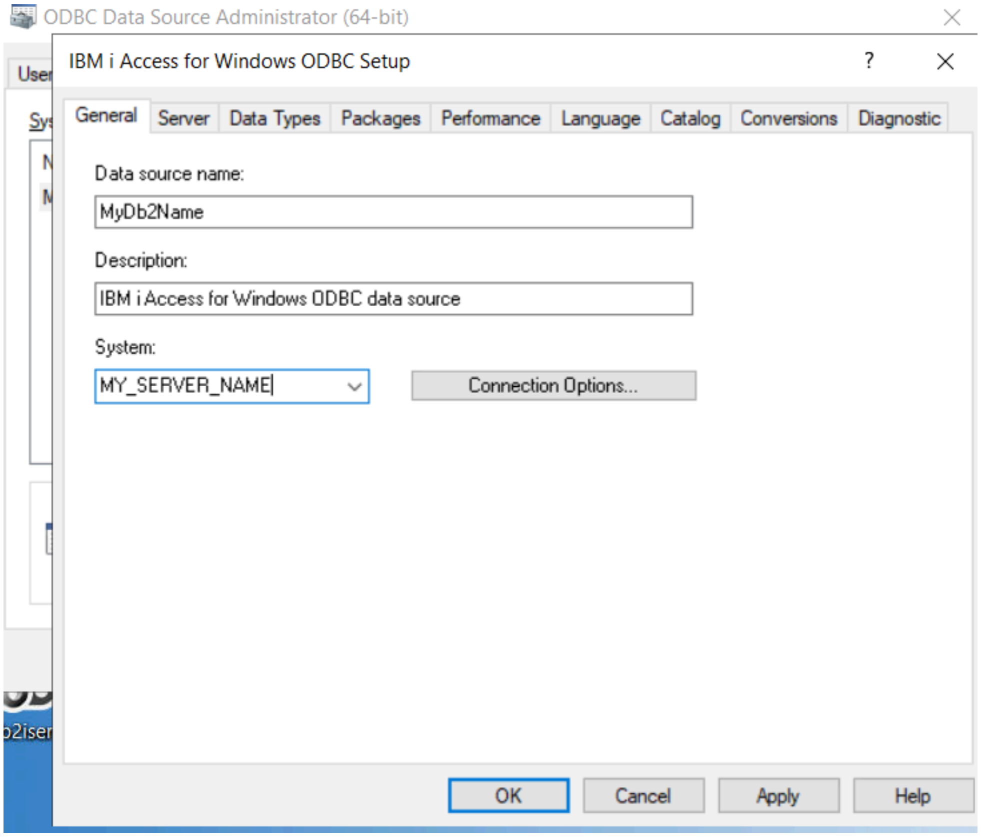
2. In the System DSN tab, click on Add to initiate the creation of a new data source.



An ODBC System data source stores information about how to connect to the indicated data provider. A System data source is visible to all users of this computer, including NT services.

3. Select iSeries Access ODBC Driver and click the Finish button. The General tab of the IBM i Access for Windows ODBC Setup dialog is displayed.

4. Recommended Settings for the General tab:
    - Data source name – Enter any unique name to identify the ODBC data source.
    - System – Enter the hostname where LMi resides.



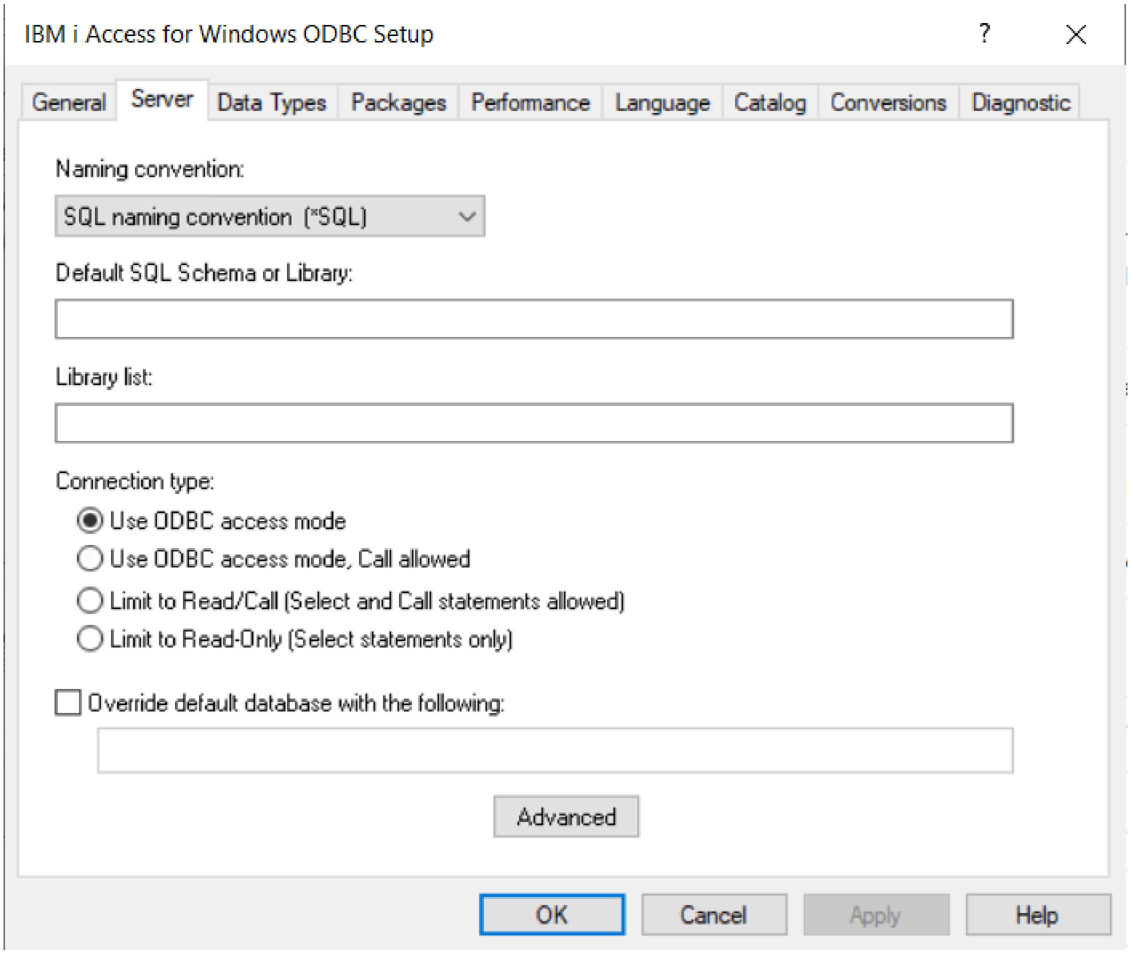
Data source name: MyDb2Name

Description: IBM i Access for Windows ODBC data source

System: MY_SERVER_NAME

5. Click the Connection Options button to display the Connection Options dialog. Recommended settings:
    - Default user ID – Use iSeries Navigator default
    - Sign-on dialog prompting – Prompt for SQL Connect if needed
    - Security – Use the same security as the iSeries Navigator connection
6. Click OK.

7. On the IBM i Access for Windows ODBC Setup dialog box, select the Server tab. Recommended settings:
   a. In the Naming convention combo box, select SQL naming convention (*SQL).
   b. Leave Default SQL Schema or Library blank.
   c. Leave the Library list blank.
   d. For the Connection type, select Use ODBC access mode.
   e. Leave Override default database with the following unchecked.



Naming convention: SQL naming convention (*SQL)

Default SQL Schema or Library:

Library list:

Connection type:

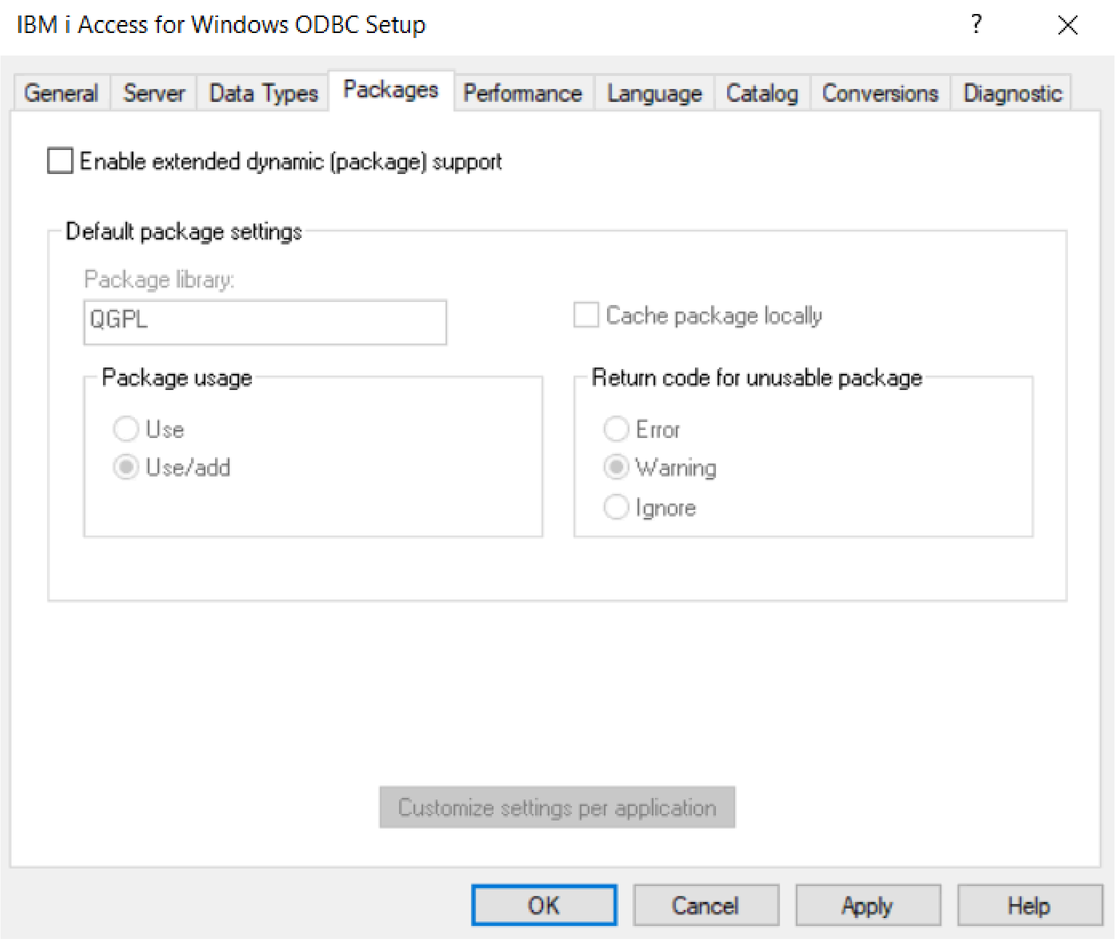
● Use ODBC access mode
○ Use ODBC access mode, Call allowed
○ Limit to Read/Call (Select and Call statements allowed)
○ Limit to Read-Only (Select statements only)

□ Override default database with the following:

   f. Click OK.

8. On the IBM i Access for Windows ODBC Setup dialog box, select the Packages tab. Packages Tab Adjustment:

   In the Packages tab, it is crucial to ensure the 'Enable Extended Dynamic Support' checkbox is unchecked, deactivating other controls on the page.



□ Enable extended dynamic (package) support

Default package settings

Package library: QGPL

□ Cache package locally

Package usage

○ Use

● Use/add

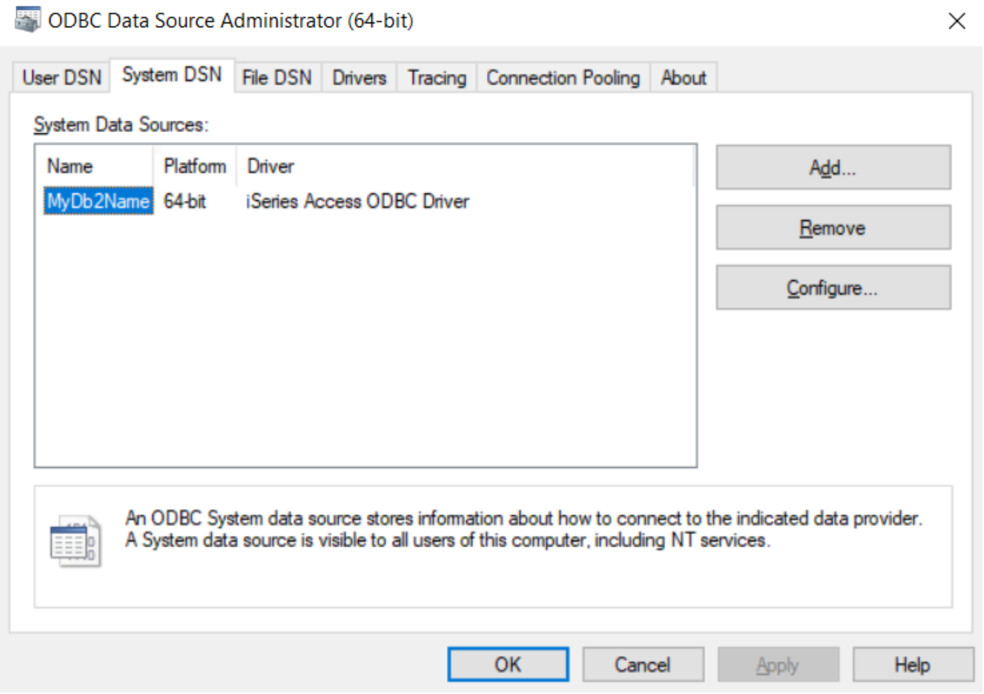
Return code for unusable package

○ Error

● Warning

○ Ignore

Finally, this is what we will see:



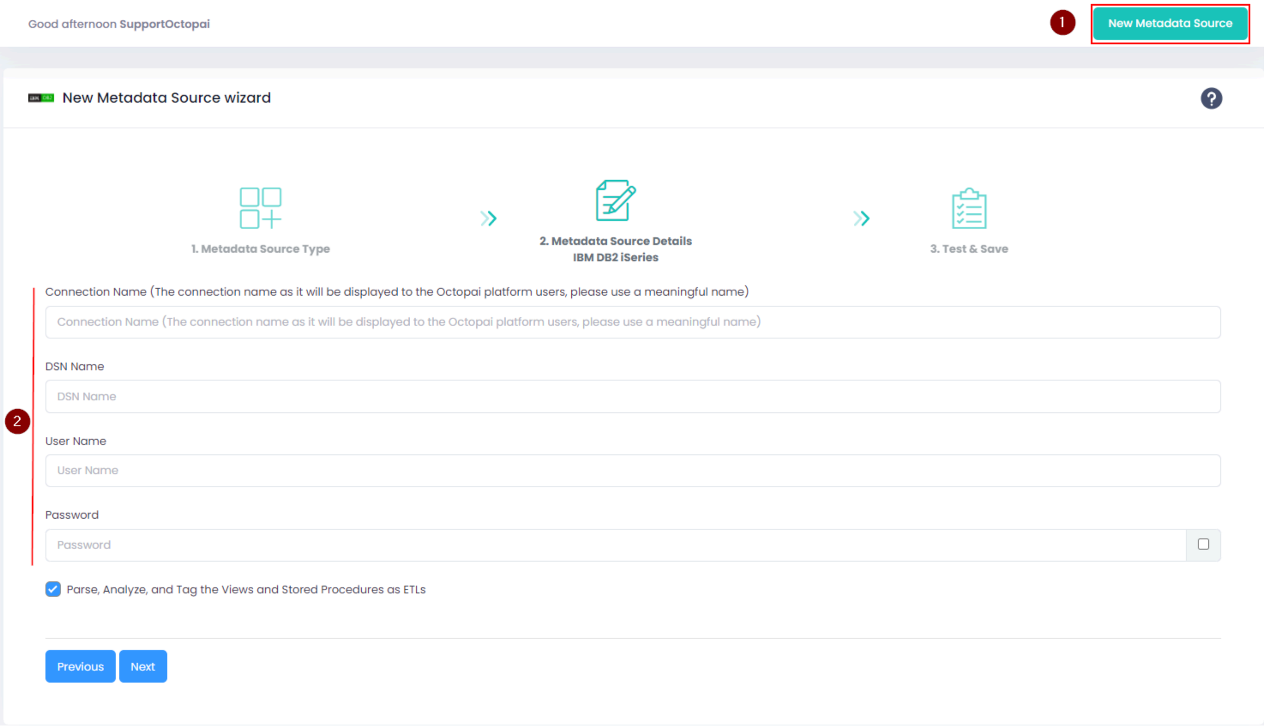
System Data Sources:

Name	Platform	Driver
MyDb2Name	64-bit	iSeries Access ODBC Driver

An ODBC System data source stores information about how to connect to the indicated data provider. A System data source is visible to all users of this computer, including NT services.

Setting up IBM Iseries Metadata Source in Cloudera Octopai Client

First, press "New Metadata Source," then fill in the relevant information for your new connection.



Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

DSN Name

DSN Name

User Name

User Name

Password

Password

☑ Parse, Analyze, and Tag the Views and Stored Procedures as ETLs

Google BigQuery

Learn how to configure Google BigQuery as a metadata source for Cloudera Octopai.

Note: **Version Supported:** up to 2023

Tool Permissions Prerequisites

Warning:

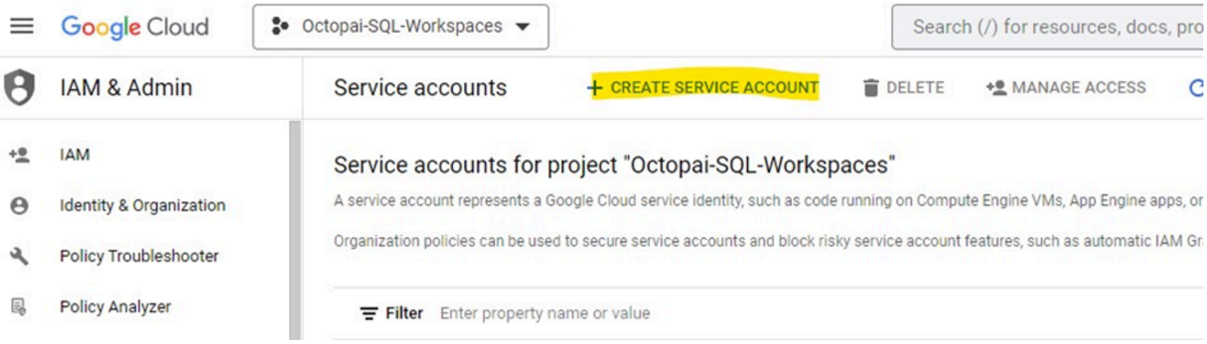
Missing permissions could end up in broken lineages.

Google Service Account (SA) integrated with Cloudera Octopai.

How to establish a Google Cloud Service Account

Creating a Google Cloud Service Account (SA)

1. Open the Google Cloud Console and navigate to your desired project.
2. Access the IAM & Admin menu and select 'Service Accounts'.
3. Click on 'Create Service Account' to initiate the creation process.



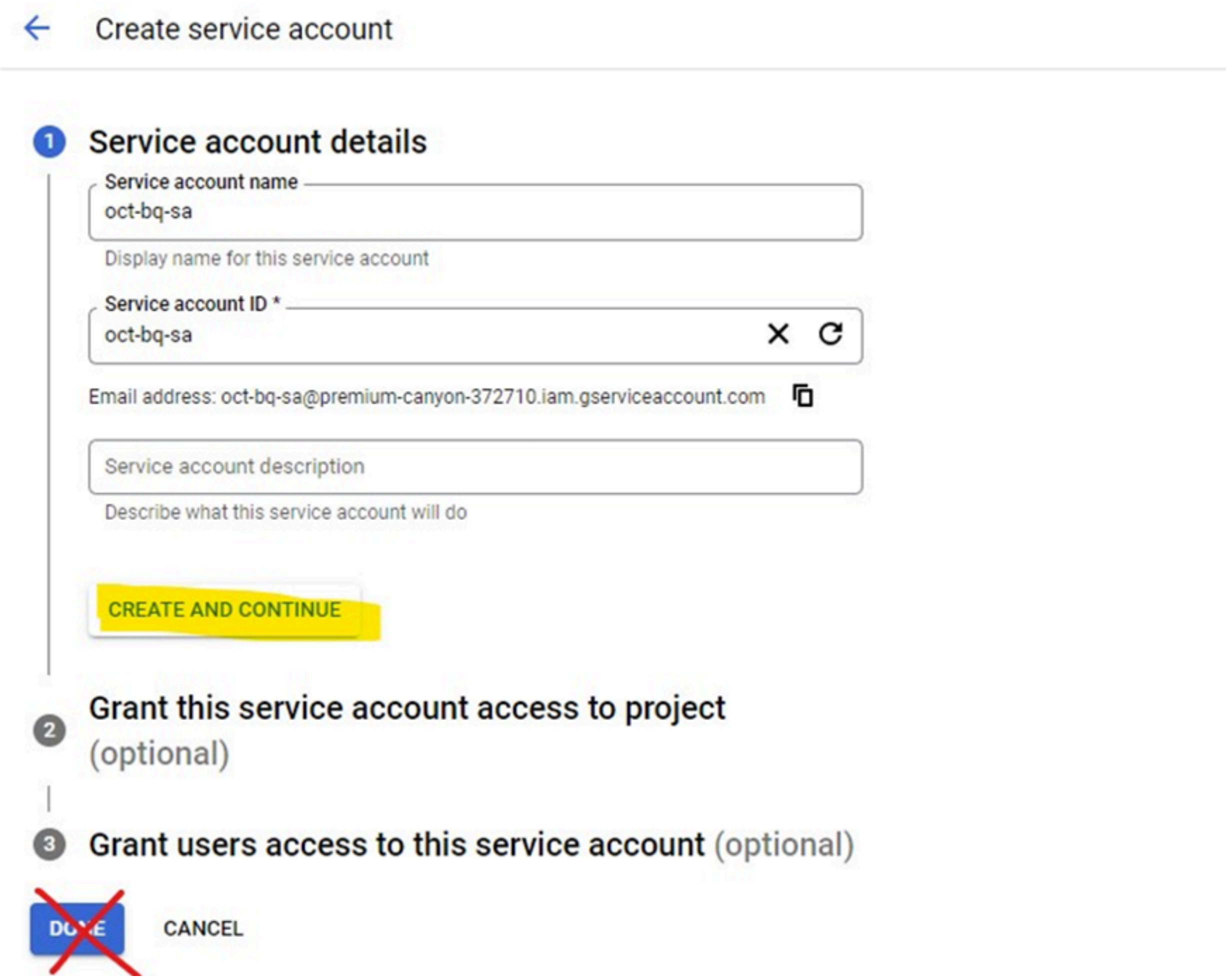
Service accounts for project "Octopai-SQL-Workspaces"

A service account represents a Google Cloud service identity, such as code running on Compute Engine VMs, App Engine apps, or

Organization policies can be used to secure service accounts and block risky service account features, such as automatic IAM Gr

4. Assign a unique, identifiable name to your Service Account. Upon clicking 'Create and Continue', an Identity and Access Management (IAM) principal will be instantiated automatically.

   ⚠ **Important:** Important: Do not click 'Done' at this point.



Create service account

1 Service account details

Service account name: oct-bq-sa

Display name for this service account

Service account ID *: oct-bq-sa

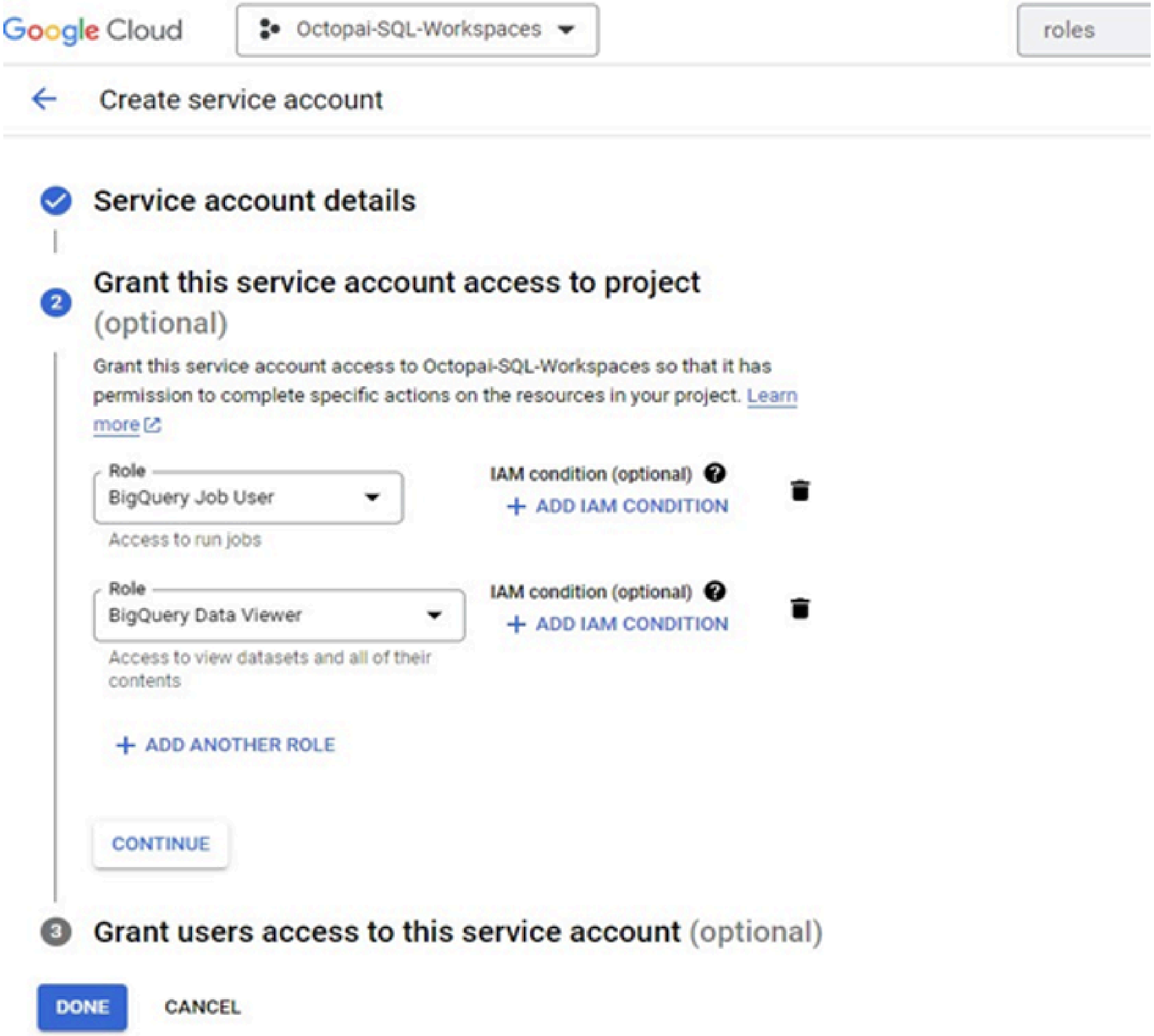
Email address: oct-bq-sa@premium-canyon-372710.iam.gserviceaccount.com

Service account description

Describe what this service account will do

CREATE AND CONTINUE
2 Grant this service account access to project (optional)
3 Grant users access to this service account (optional)
DONE
CANCEL

5. From the 'roles' dropdown menu, select both 'BigQuery Data Viewer' and 'BigQuery Job User' to assign necessary permissions to your Service Account.



Service account details

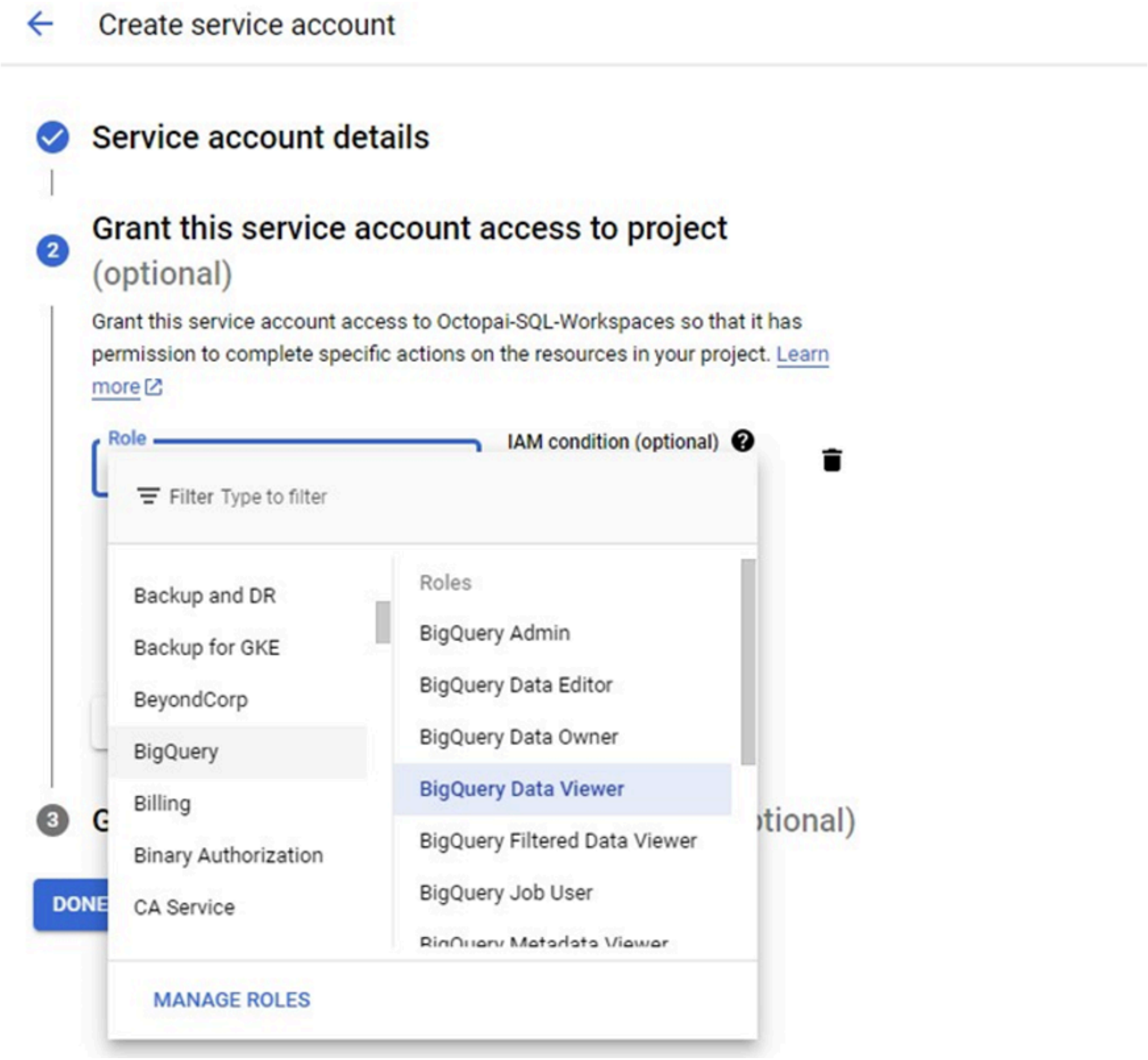
Grant this service account access to project (optional)

Grant this service account access to Octopai-SQL-Workspaces so that it has permission to complete specific actions on the resources in your project. Learn more

Role: BigQuery Job User
Access to run jobs

Role: BigQuery Data Viewer
Access to view datasets and all of their contents

Grant users access to this service account (optional)



Create service account

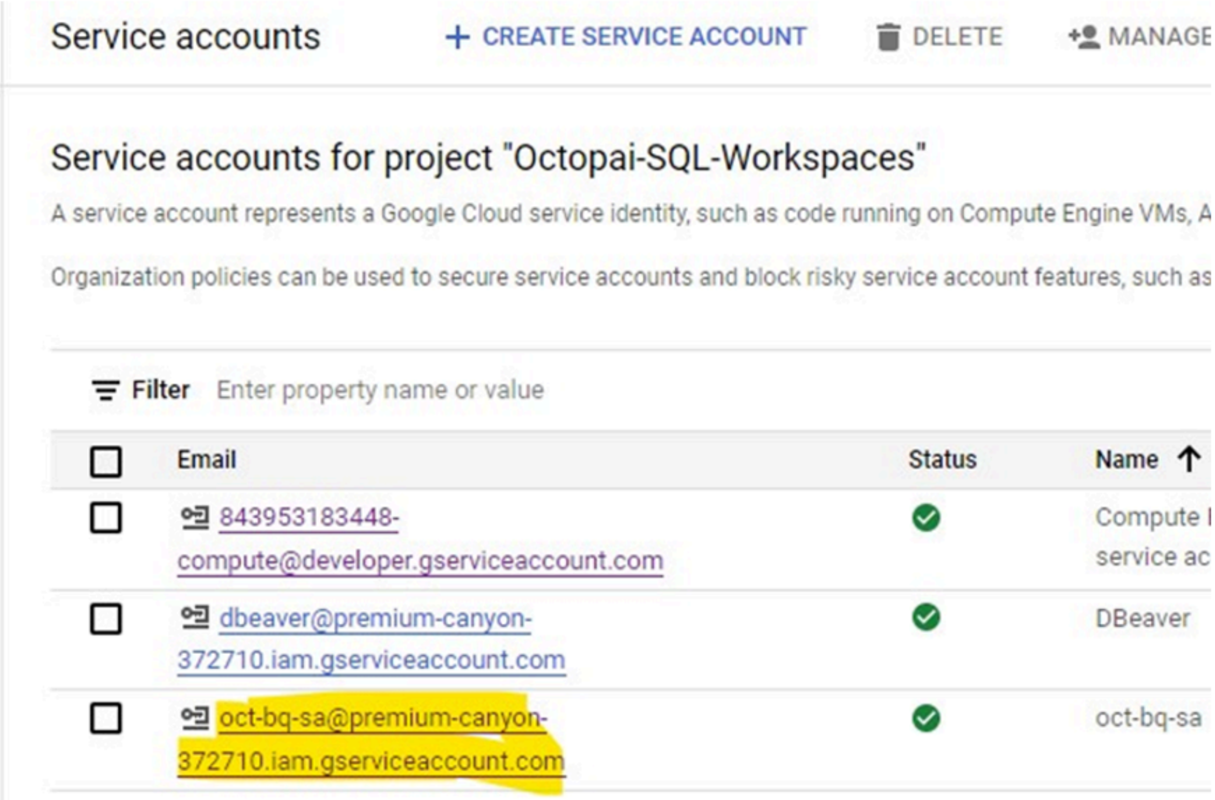
Service account details

Grant this service account access to project (optional)

Grant this service account access to Octopai-SQL-Workspaces so that it has permission to complete specific actions on the resources in your project. Learn more

6. Complete the Service Account creation process by clicking 'Done'.

7. Access the newly created Service Account and navigate to the 'KEYS' tab.

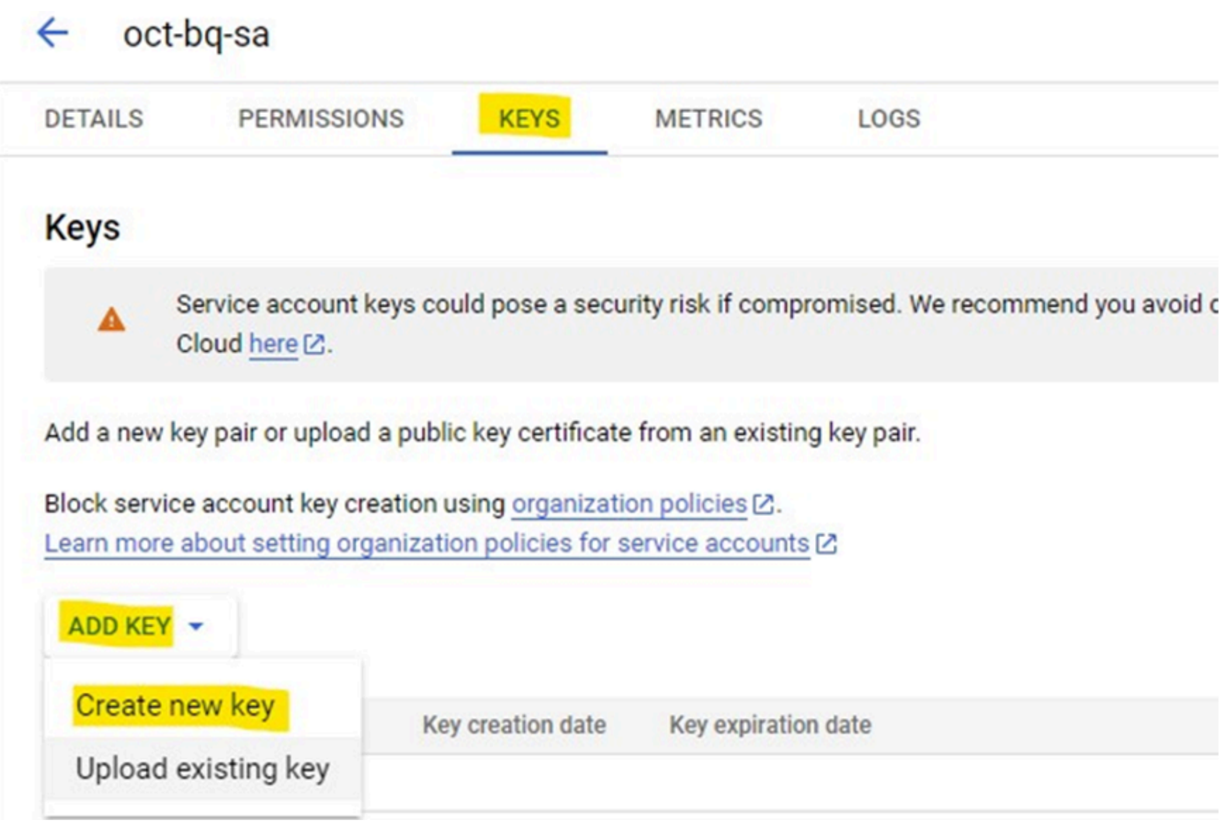


Service accounts for project "Octopai-SQL-Workspaces"

A service account represents a Google Cloud service identity, such as code running on Compute Engine VMs, A

Organization policies can be used to secure service accounts and block risky service account features, such as

☐	Email	Status	Name ↑
☐	843953183448-compute@developer.gserviceaccount.com	✓	Compute service ac
☐	dbeaver@premium-canyon-372710.iam.gserviceaccount.com	✓	DBeaver
☐	oct-bq-sa@premium-canyon-372710.iam.gserviceaccount.com	✓	oct-bq-sa



Keys

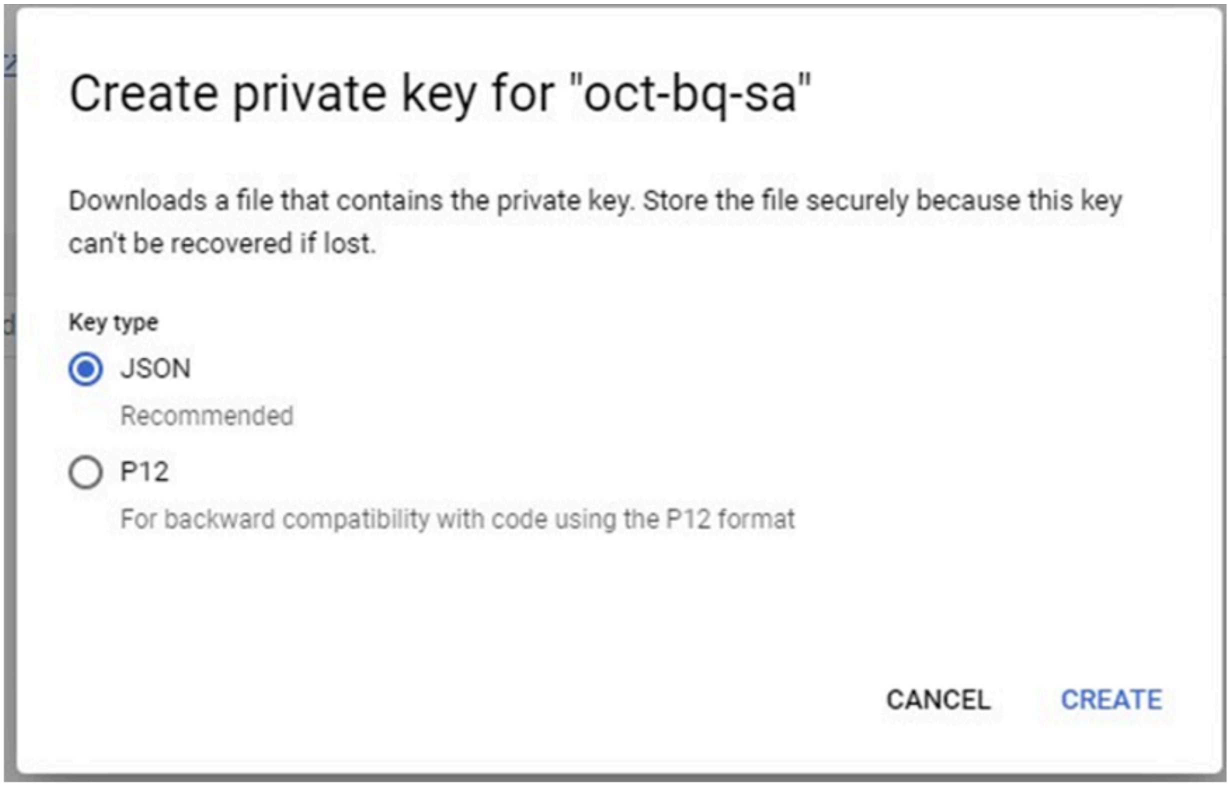
Service account keys could pose a security risk if compromised. We recommend you avoid c Cloud here.

Add a new key pair or upload a public key certificate from an existing key pair.

Block service account key creation using organization policies.
Learn more about setting organization policies for service accounts

Generating a Service Account Key

- Click on 'ADD KEY', then 'Create new key'. Make sure to select the JSON format. The key will automatically be downloaded to your local system.



Create private key for "oct-bq-sa"

Downloads a file that contains the private key. Store the file securely because this key can't be recovered if lost.

Key type

● JSON
Recommended

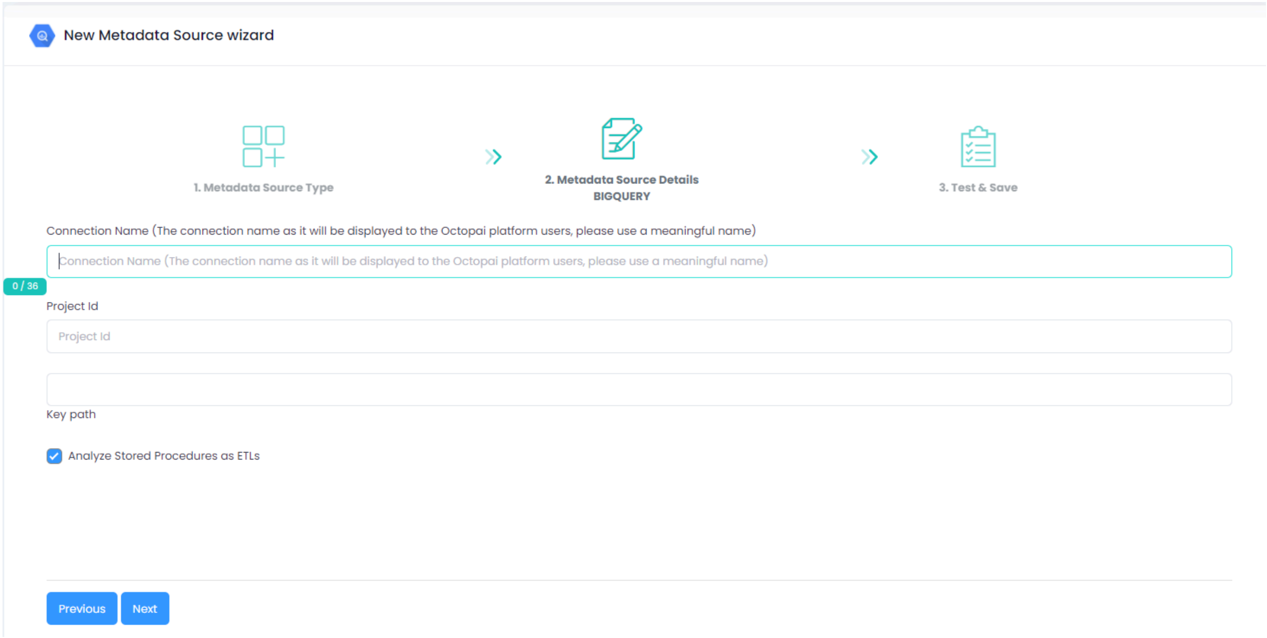
○ P12
For backward compatibility with code using the P12 format

Configuring Cloudera Octopai with BigQuery:

1. Open Cloudera Octopai and start the creation of a new metadata source, choosing 'BigQuery' as your source.
2. Assign a descriptive name to this metadata source for easy reference in the future.
3. Input the Project ID associated with your Google Cloud project. This can be found within the downloaded JSON file (under the 'project_id' field) or in the project selector on the Google Cloud Console.
4. Specify the file path where your downloaded JSON key file is stored in the 'Key Path' field.
5. Save your settings and initiate the connection by clicking 'Save and Run'.

Setting up Google BigQuery Metadata Source

Metadata Sources are set on the Cloudera Octopai Client


New Metadata Source wizard

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Project Id

Project Id

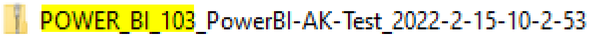
Key path

☑ Analyze Stored Procedures as ETLs

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having your **Connector Name** Example:

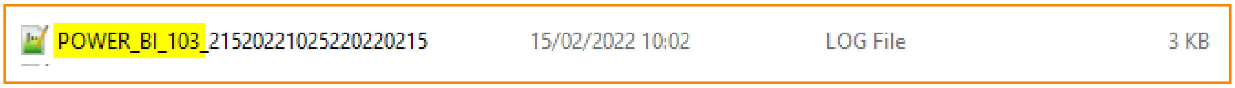
   
   POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service\log


POWER_BI_103_21520221025220220215 15/02/2022 10:02 LOG File 3 KB

GreenPlum

Learn how to configure GreenPlum as a metadata source for Cloudera Octopai.

Tool Permissions Prerequisites

⚠ **Warning:** Missing permissions could end up in broken lineages.

Add the following line to the **pg_hba.conf** file located under **/data/master/gpseg-1**.

Enable remote access, allowing the Cloudera Octopai Client to perform the MetaData extraction:

TYPE	DATABASE	USER	ADDRESS	METHOD
host	databasename	octopai-user	octopai-client-ip-address	md5

Replace **"databasename"**, **"octopai-user"**, and **"octopai-client-ip-address"** with the actual values for your database and client. This line will allow the Cloudera Octopai Client to access the specified database using **md5** authentication.

Ensure the following prerequisites are met:

- Postgres ODBC driver installed on the machine running the Cloudera Octopai Client.
- Open Server Port for each Postgre Database Connection.
- Existing/New Postgres user (**OCTOPAI_USER**) for each connection with grant select permission for the following dictionary tables:
  - information_schema.tables
  - information_schema.views
  - information_schema.columns
  - pg_catalog.pg_proc
  - pg_catalog.pg_namespace
  - pg_catalog.pg_attribute
  - pg_catalog.pg_constraint
  - pg_catalog.pg_class
  - pg_catalog.pg_database

How to set up the permissions

psqlodbc - PostgreSQL ODBC driver

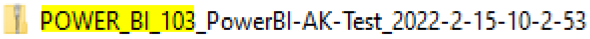
Setting up GreenPlum Metadata Source

Metadata Sources are set on the Cloudera Octopai Client

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:



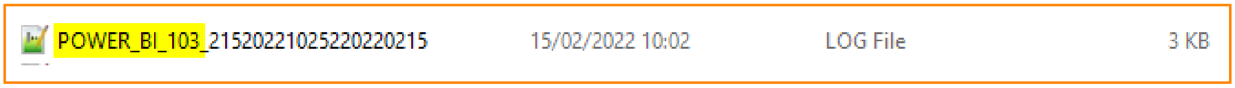
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service\log



POWER_BI_103_21520221025220220215 15/02/2022 10:02 LOG File 3 KB

Oracle Autonomous Data Warehouse (OAD/ADW)

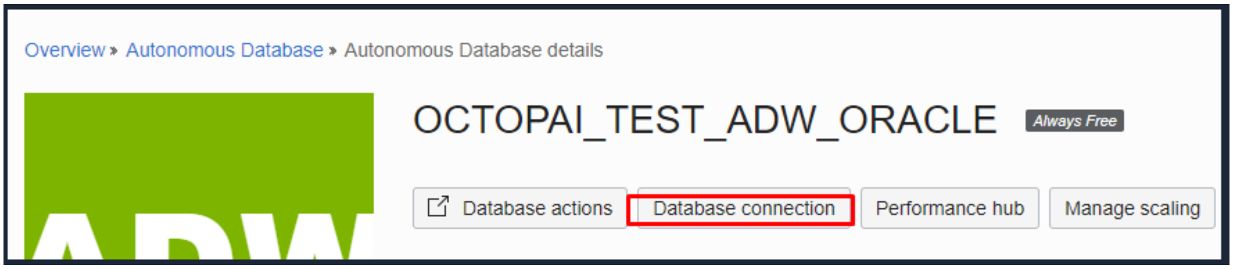
Learn how to configure Oracle Autonomous Data Warehouse (OAD/ADW) as a metadata source for Cloudera Octopai.

Note: Version Supported: 19c / 20c / 21c

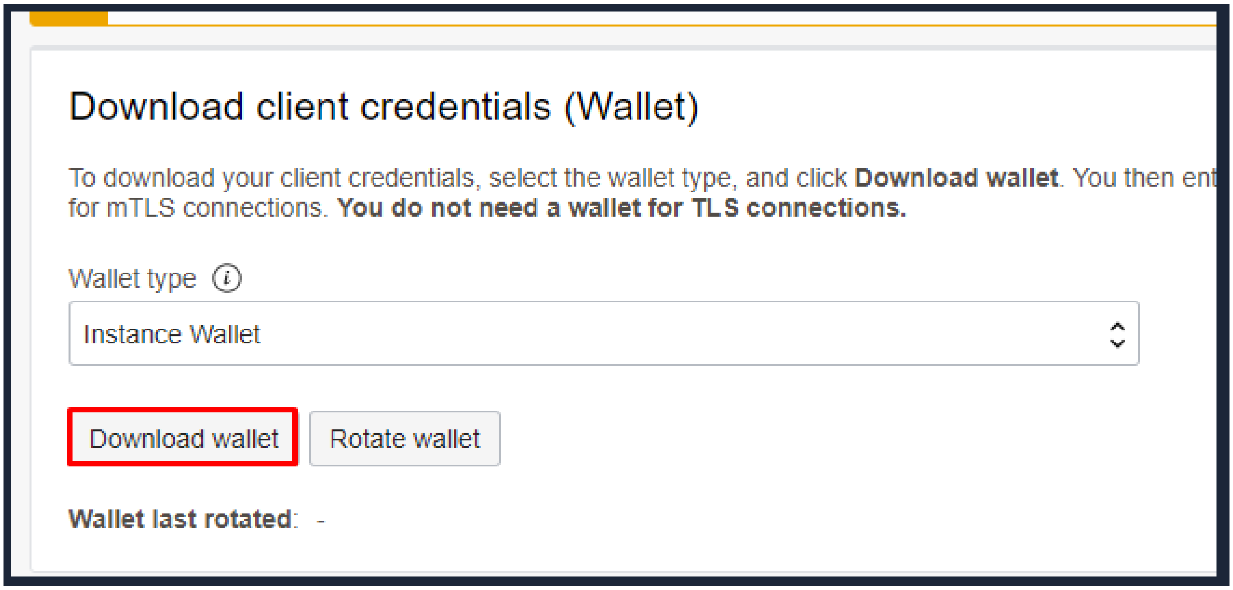
Tool Permissions Prerequisites

Warning: Missing permissions could end up in broken lineages.

1. Log into OCI, go to databases.
2. Select the database you want to connect and click on “Database connection”.



3. Set the wallet type to “Instance Wallet”, and download the file.



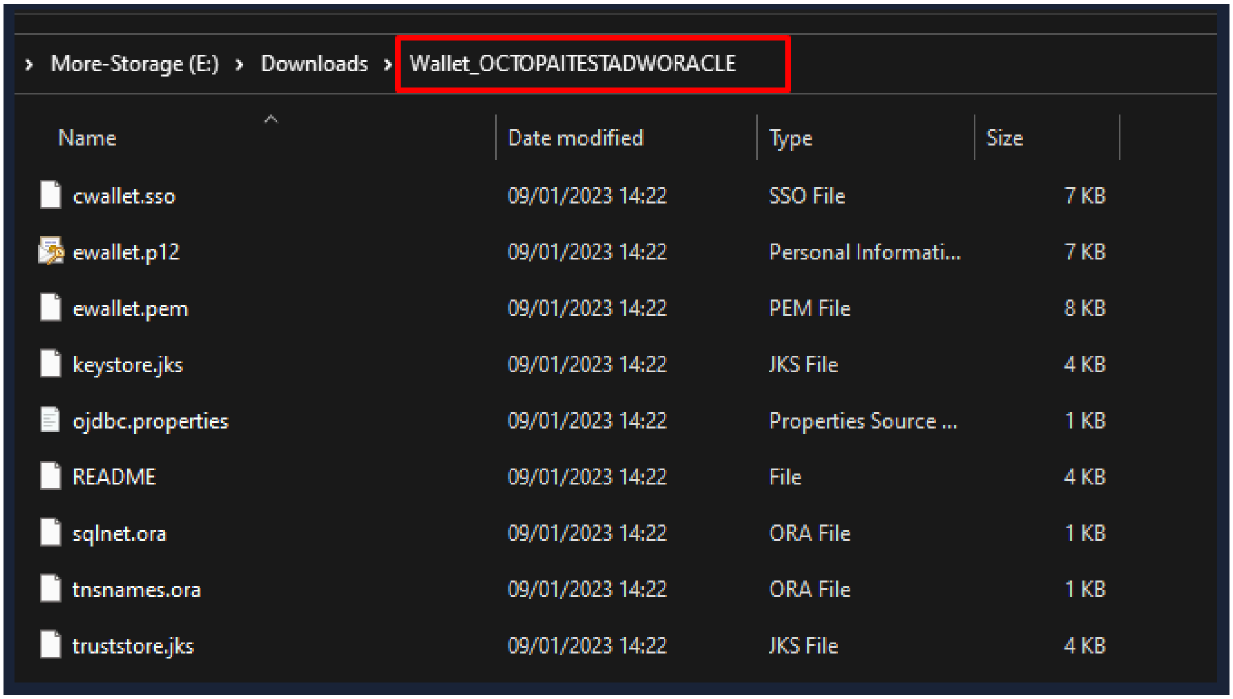
Download client credentials (Wallet)

To download your client credentials, select the wallet type, and click **Download wallet**. You then ent
for mTLS connections. **You do not need a wallet for TLS connections.**

Wallet type: Instance Wallet

Wallet last rotated: -

4. Extract the content of the zip into a folder, it should look like this:

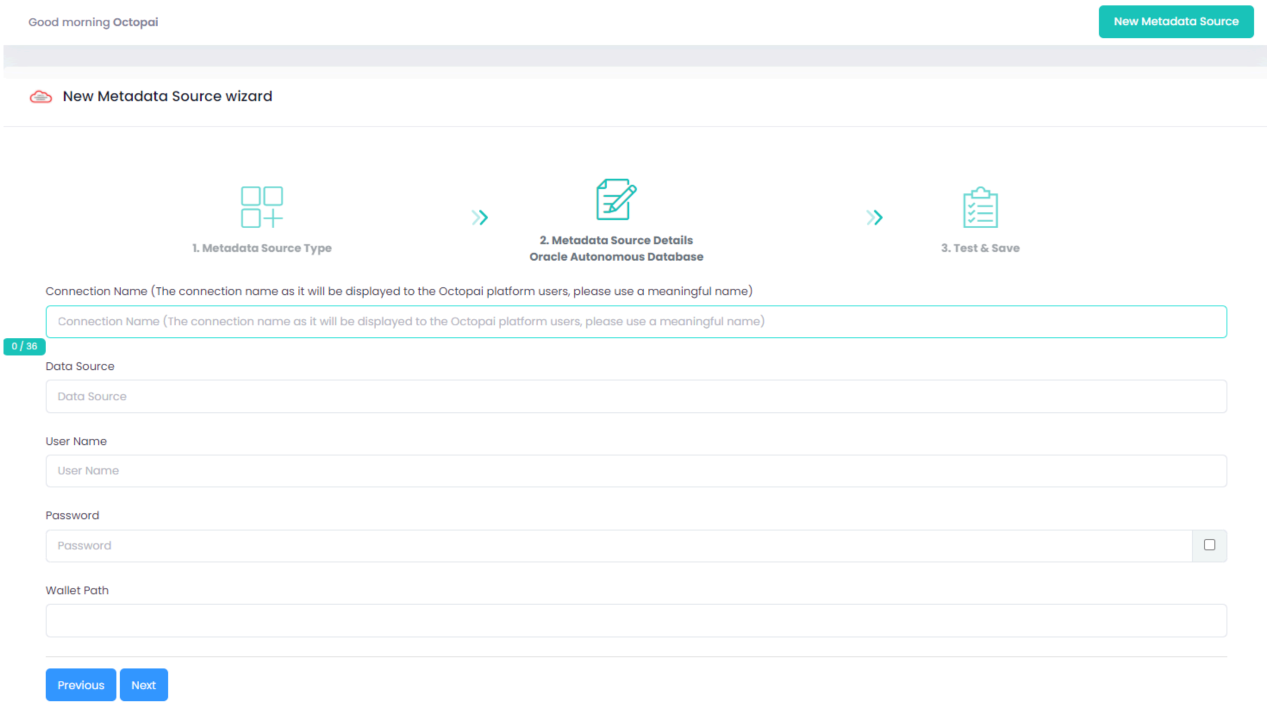


› More-Storage (E:) › Downloads › Wallet_OCTOPAITESTADWORACLE

Name	Date modified	Type	Size
cwallet.sso	09/01/2023 14:22	SSO File	7 KB
ewallet.p12	09/01/2023 14:22	Personal Informati...	7 KB
ewallet.pem	09/01/2023 14:22	PEM File	8 KB
keystore.jks	09/01/2023 14:22	JKS File	4 KB
ojdbc.properties	09/01/2023 14:22	Properties Source ...	1 KB
README	09/01/2023 14:22	File	4 KB
sqlnet.ora	09/01/2023 14:22	ORA File	1 KB
tnsnames.ora	09/01/2023 14:22	ORA File	1 KB
truststore.jks	09/01/2023 14:22	JKS File	4 KB

Setting up OAD Metadata Source

Metadata Sources are set on the Cloudera Octopai Client



Good morning Octopai

New Metadata Source wizard

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Data Source

Data Source

User Name

User Name

Password

Password □

Wallet Path

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:

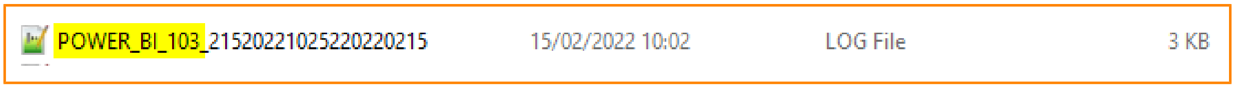
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service\log



POWER_BI_103_21520221025220220215 15/02/2022 10:02 LOG File 3 KB

IBM DB2 LUW

Learn how to integrate IBM DB2 LUW with Cloudera Octopai.

Tool Permissions Prerequisites

⚠ **Warning:** Missing permissions could end up in broken lineages.

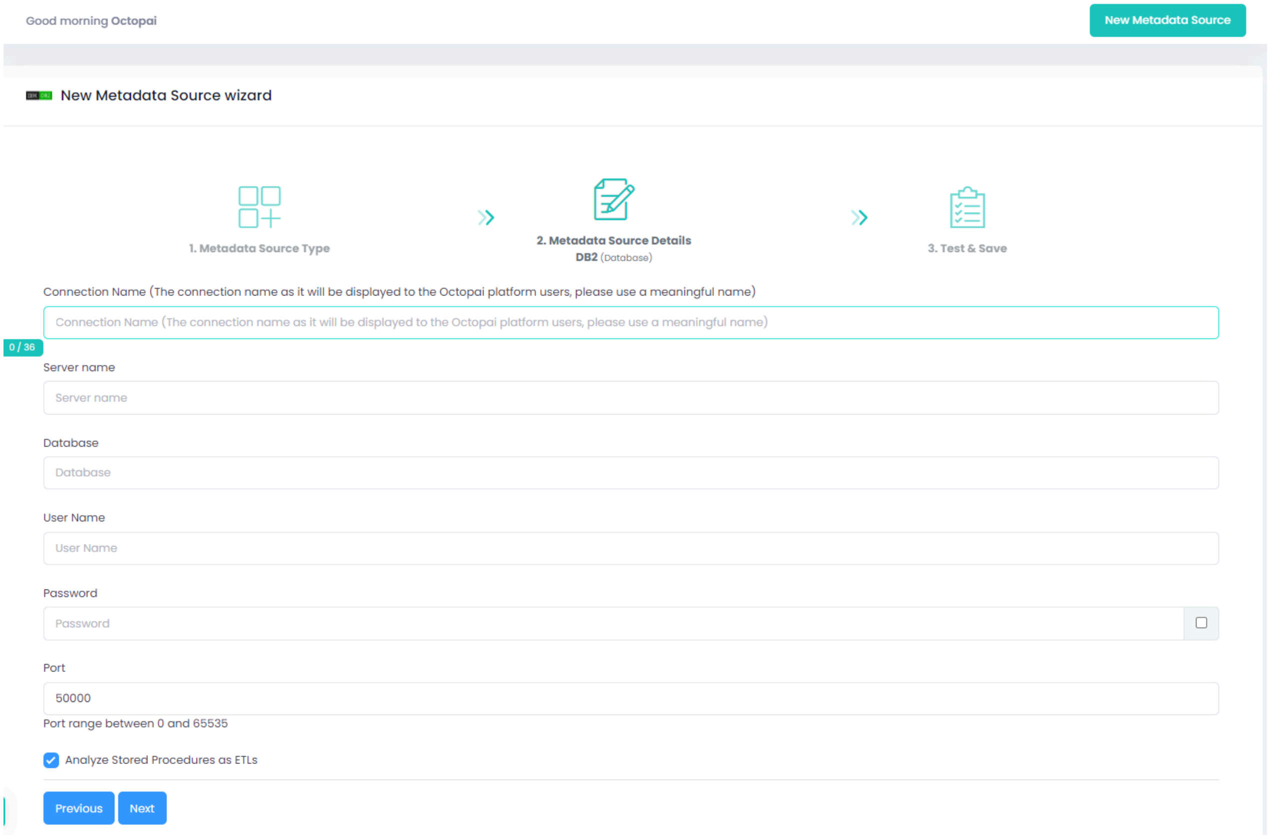
- DB2 ODBC driver installed on the machine running the Cloudera Octopai Client
- Microsoft Visual C++ Redistributable - Visual Studio 2013 (VC++ 12.0) - 64X, installed on the machine running the Cloudera Octopai Client
- Open Server Port for each DB2 Database Connection

- Existing/New DB2 user(OCTOPAI_USER) for each connection with:
  - Grant connect on DATABASE ;
  - Grant Execute on function SYSPROC.ENV_GET_INST_INFO;
  - Grant Execute on function SYSPROC.MON_GET_CONTAINER;
  - Grant select permission for the following dictionary tables:
    - SYSIBM.SYSTABAUTH;
    - SYSIBM.SYSCOLAUTH;
    - SYSIBM.SYSDBAUTH;
    - SYSIBMADM.DBMCFG;
    - SYSIBMADM.DBCFG;
    - SSYSIBM.SYSDBAUTH;
    - SYSCAT.PACKAGEAUTH;
    - SYSCAT.TBSPACEAUTH;
    - SSYSCAT.DBAUTH;
    - SYSCAT.SEQUENCEAUTH;
    - SYSCAT.INDEXAUTH;
    - SYSCAT.TABLES;
    - SYSIBM.SYSSCHEMAAUTH ;
    - SSYSIBM.SYSTABAUTH;
    - SYSCAT.LIBRARYAUTH;
    - SYSCAT.TABAUTH;
    - SYSIBM.SYSROUTINEAUTH;
    - SYSIBM.ROUTINES;
    - SYSCAT.INDEXES;
    - SSYSCAT.SCHEMAAUTH;
    - SYSCAT.PACKAGES;
    - SYSCAT.VIEWS;
    - SYSCAT.TRIGGERS;
    - SYSCAT.PASSTHRUAUTH;
    - SYSCAT.ROUTINEAUTH;
    - SYSCAT.TABLESPACES;
    - SYSCAT.SEQUENCES;
    - SYSCAT.ROUTINES;
    - SYSCAT.INDEXES;
    - SYSCAT.PACKAGES;
    - SYSCAT.SCHEMATA;
    - SYSCAT.TRIGGERS;
    - SYSCAT.TABLES;
    - SYSCAT.VIEWS;
    - SYSCAT.ROUTINES;

Setting up IBM - DB2 Metadata Source

Metadata Sources are set on the Cloudera Octopai Client

Figure 20: New Metadata Source Wizard



Good morning Octopai

New Metadata Source

New Metadata Source wizard

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Server name

Server name

Database

Database

User Name

User Name

Password

Password

Port

50000

Port range between 0 and 65535

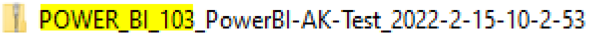
☑ Analyze Stored Procedures as ETLs

Previous Next

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:



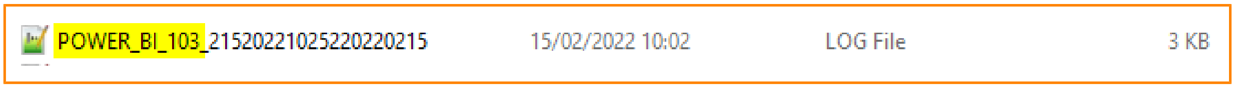
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service\log



POWER_BI_103_21520221025220220215 15/02/2022 10:02 LOG File 3 KB

MySQL

How to set up MySQL metadata sources, configure permissions, and verify extracted files.

Tool Permissions Prerequisites

Warning: Missing permissions could end up in broken lineages.

MySQL ODBC driver installed on the machine running the Cloudera Octopai Client

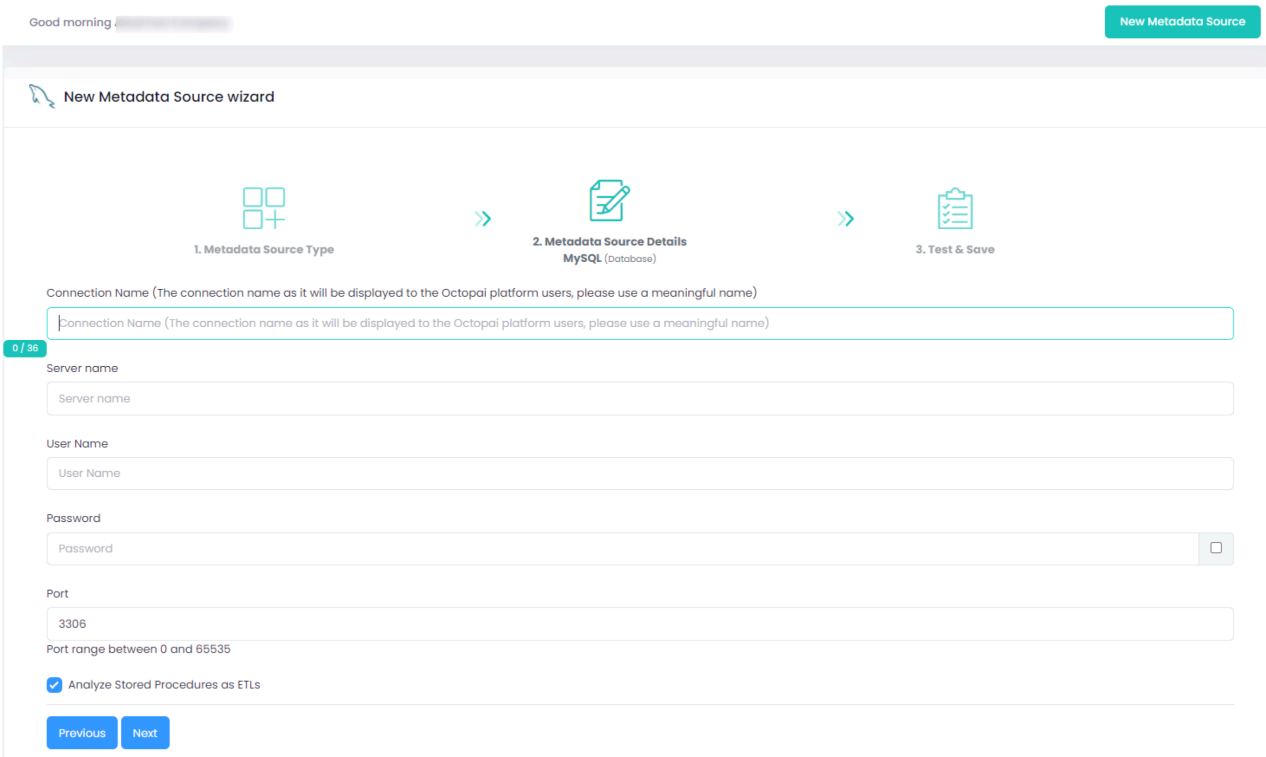
Open Server Port for each MySQL Database Connection

Existing/New MySQL user (OCTOPAI_USER) for each connection with grant select permission for the following dictionary tables:

- information_schema.tables;
- information_schema.columns;
- information_schema.key_column_usage;
- information_schema.table_constraints;
- information_schema.routines;

Setting up MySQL Metadata Source

Metadata Sources are set on the Cloudera Octopai Client



Good morning [redacted]

New Metadata Source wizard

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Server name

Server name

User Name

User Name

Password

Password

Port

3306

Port range between 0 and 65535

☑ Analyze Stored Procedures as ETLs

How to set up the permissions

MySQL :: Download Connector/ODBC

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:

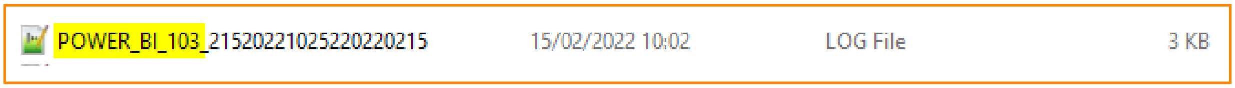
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service \log



POWER_BI_103_2152022102522022015 15/02/2022 10:02 LOG File 3 KB

PostgreSQL

How to set up PostgreSQL metadata sources in Cloudera Octopai Data Lineage, including prerequisites, verification, and troubleshooting.

Note: Version supported: up to 16.x

Tool Permissions Prerequisites

Warning: Missing permissions could end up in broken lineages.

- Postgres ODBC driver installed on the machine running the Cloudera Octopai Client
- Open Server Port for each Postgre Database Connection
- Existing/New Postgres user (OCTOPAI_USER) for each connection with grant select permission for the following dictionary tables:
  - information_schema.tables;
  - information_schema.views;
  - information_schema.columns;
  - pg_catalog.pg_proc;
  - pg_catalog.pg_namespace;
  - pg_catalog.pg_attribute;
  - pg_catalog.pg_constraint;
  - pg_catalog.pg_class;
  - pg_catalog.pg_database;

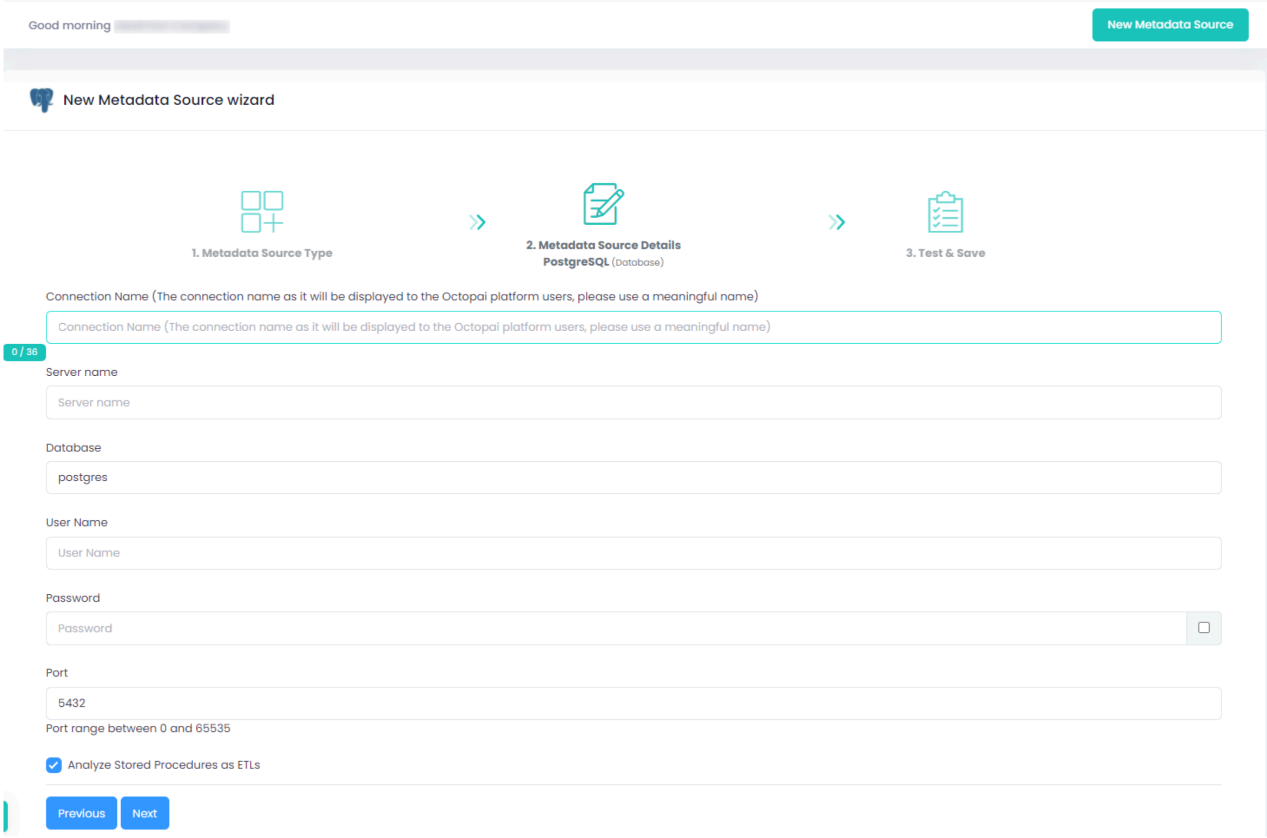
How to set up the permissions

psqlodbc - PostgreSQL ODBC driver

Setting up PostgreSQL Metadata Source

Metadata Sources are set on the Cloudera Octopai Client

Figure 21: New Metadata Source wizard



Good morning [redacted]

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Server name

Server name

Database

postgres

User Name

User Name

Password

Password

Port

5432

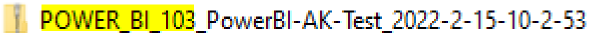
Port range between 0 and 65535

☑ Analyze Stored Procedures as ETLs

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:



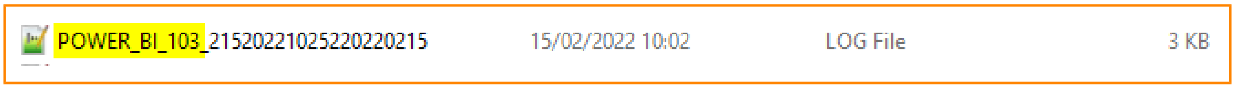
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service\log



POWER_BI_103_2152022102522020215 15/02/2022 10:02 LOG File 3 KB

Amazon Redshift

Learn how to integrate Amazon Redshift with Cloudera Octopai.

Note: **Version supported:** up to 2023

Tool Permissions Prerequisites

Warning:

Missing permissions could end up in broken lineages.

Amazon Redshift ODBC driver installed on the machine running the Cloudera Octopai Client https://docs.aws.amazon.com/redshift/latest/mgmt/configure-odbc-connection.html#install-odbc-driver-windows

Open Server Port for each Redshift Database Connection

Exiting/New Redshift user (OCTOPAI_USER) for each connection with grant select permission for the following dictionary tables:

- pg_catalog.pg_database
- pg_catalog.pg_proc
- pg_catalog.pg_table_def
- pg_catalog.pg_tables
- pg_catalog.pg_views
- pg_catalog.pg_namespace
- pg_catalog.pg_attribute
- pg_catalog.pg_class
- pg_catalog.pg_constraint
- pg_catalog.svv_table_info
- pg_catalog.svv_all_tables
- pg_catalog.svv_all_columns
- pg_catalog.pg_user
- pg_catalog.svv_external_schemas
- pg_catalog.svv_external_tables
- pg_catalog.svv_external_columns

Grant select on all tables in schema dbo to the group assigned to the Redshift user.

Grant usage on schema dbo to the group assigned to the Redshift user.

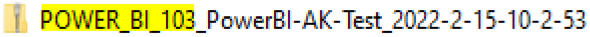
Setting up Redshift Metadata Source

Metadata Sources are set on the Cloudera Octopai Client

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:



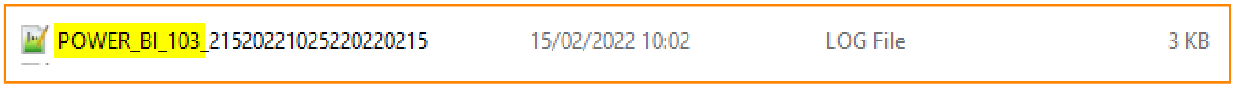
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service \log



```
POWER_BI_103_2152022102522022021 5    15/02/2022 10:02    LOG File    3 KB
```

Snowflake Tool Permissions Prerequisites

Learn how to configure Snowflake permissions to enable seamless integration with Cloudera Octopai.

Note: Version supported: 2025

Permissions overview

Warning:

Missing permissions can result in incomplete or broken lineage. Proper role assignments are critical for a successful Cloudera Octopai–Snowflake integration.

This guide explains how to create a dedicated user and role, grant the required permissions, validate configuration, and revoke access if needed. Replace placeholder values such as <username> and <warehouse_name> with your actual Snowflake details.

General guidelines

- Create the Snowflake user while signed in with an **ACCOUNTADMIN** role.
- Assign required roles and a **DEFAULT_WAREHOUSE** to ensure efficient query processing.
- Grant **USAGE** on all schemas in the target databases to simplify access management.

Note:

Snowflake object names and settings (including warehouses) are case sensitive. Keep capitalization consistent to avoid errors.

Database and warehouse roles

Database roles

- USAGE: Lists the database and allows metadata queries.
- CREATE SCHEMA: Creates schemas within the database.
- CREATE TABLE: Creates tables within the database.
- SELECT: Queries data from tables.
- INSERT: Inserts data into tables.
- UPDATE: Updates table data.
- DELETE: Deletes table data.
- REFERENCES: Creates foreign-key relationships.

Warehouse roles

- USAGE: Runs queries on the warehouse.
- MONITOR: Reviews warehouse usage and performance.
- OPERATE: Starts, stops, and resizes the warehouse.
- OWNERSHIP: Grants full control over the warehouse.

Set up Snowflake permissions

Follow these steps to provision a dedicated Cloudera Octopai user and assign the necessary Snowflake roles and privileges.

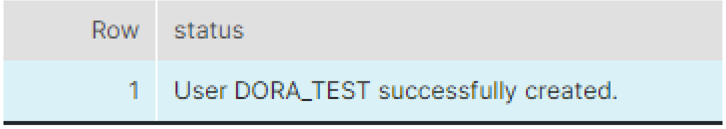
1. Create a dedicated user

Run the following SQL to create a Snowflake user for Cloudera Octopai metadata extraction:

```
CREATE USER <username>
  PASSWORD = '<password>'
  DEFAULT_WAREHOUSE = '<warehouse>'
  MUST_CHANGE_PASSWORD = false;
```

Result: the user appears in Snowflake.

Figure 22: Sample user creation output



Row	status
1	User DORA_TEST successfully created.

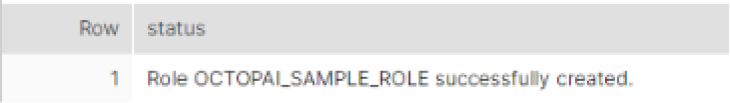
2. Create a dedicated role

Create a role to encapsulate Cloudera Octopai permissions:

```
CREATE ROLE <role_name>;
```

Result: the role is registered in Snowflake.

Figure 23: Role creation confirmation



Row	status
1	Role OCTOPAI_SAMPLE_ROLE successfully created.

3. Assign the role to the user

Grant the role and set it as the default:

```
GRANT ROLE <role_name> TO USER <username>;
```

```
ALTER USER <username> SET DEFAULT_ROLE = <role_name>;
```

4. Grant warehouse usage

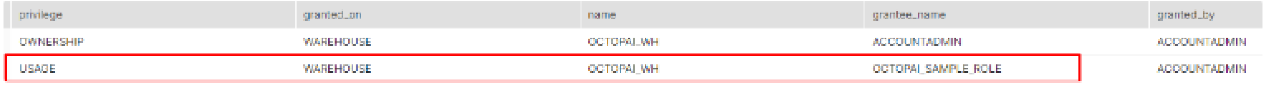
Allow the role to run workloads on the chosen warehouse:

```
GRANT USAGE ON WAREHOUSE <warehouse_name> TO ROLE <role_name>;
```

Verify warehouse grants when needed:

```
SHOW GRANTS ON WAREHOUSE <warehouse_name>;
```

Figure 24: Warehouse grants report



privilege	granted_on	name	grantee_name	granted_by
OWNERSHIP	WAREHOUSE	OCTOPAI_WH	ACCOUNTADMIN	ACCOUNTADMIN
USAGE	WAREHOUSE	OCTOPAI_WH	OCTOPAI_SAMPLE_ROLE	ACCOUNTADMIN

5. Grant imported privileges on the Snowflake database

Provide access to the shared SNOWFLAKE database:

```
GRANT IMPORTED PRIVILEGES ON DATABASE SNOWFLAKE TO ROLE <role_name>;
```

6. Grant object-specific permissions

Use the following commands as required for your environment:

Functions

```
GRANT USAGE ON FUNCTION <db_name>.<schema_name>.<function_name>(<datatype1>,
<datatype2>, ...) TO ROLE <role_name>;
```

Procedures

```
GRANT USAGE ON PROCEDURE <db_name>.<schema_name>.<procedure_name>(<datatype1
>,<datatype2>, ...) TO ROLE <role_name>;
```

Pipes

```
GRANT MONITOR, OPERATE ON PIPE <database_name>.<schema_name>.<pipe_name> TO
ROLE <role_name>;
```

Dynamic tables

```
GRANT USAGE ON DATABASE <db_name> TO ROLE <role_name>;
GRANT USAGE ON SCHEMA <schema_name> TO ROLE <role_name>;
GRANT SELECT ON ALL DYNAMIC TABLES IN SCHEMA <schema_name> TO ROLE <role_na
me>;
```

Optional: enable access to future dynamic tables.

```
GRANT SELECT ON FUTURE DYNAMIC TABLES IN SCHEMA <schema_name> TO ROLE <role_
name>;
```

Troubleshooting checks

Use these queries to validate role assignments and investigate issues:

- Check database-level grants:

```
SHOW GRANTS TO ROLE <role_name> ON DATABASE <database_name>;
```

- Review default warehouse and role settings:

```
SHOW USERS LIKE '%<username>%';
```

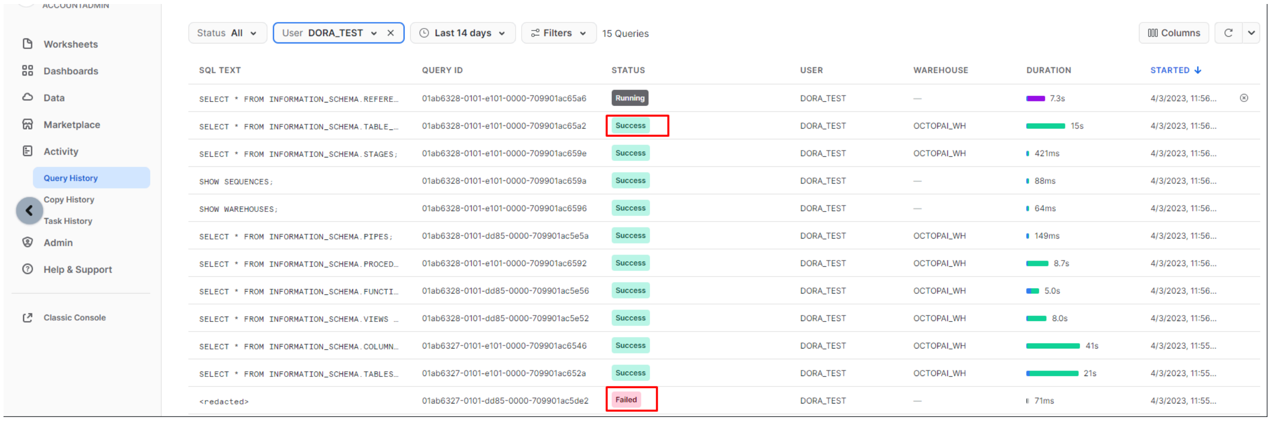
```
SHOW ROLES LIKE '%<role_name>%';
```

- Confirm warehouse grants:

```
SHOW GRANTS ON WAREHOUSE <warehouse_name>;
```

- Inspect query history logs through the Snowflake UI.

Figure 25: Query history filtered by role



Status All | User DORA_TEST | Last 14 days | Filters | 15 Queries

SQL TEXT	QUERY ID	STATUS	USER	WAREHOUSE	DURATION	STARTED
SELECT * FROM INFORMATION_SCHEMA.REFERE…	01ab6328-0101-e101-0000-709901ac65a6	Running	DORA_TEST	—	7.3s	4/3/2023, 11:56…
SELECT * FROM INFORMATION_SCHEMA.TABLE_…	01ab6328-0101-e101-0000-709901ac65a2	Success	DORA_TEST	OCTOPAI_WH	15s	4/3/2023, 11:56…
SELECT * FROM INFORMATION_SCHEMA.STAGES;	01ab6328-0101-e101-0000-709901ac659e	Success	DORA_TEST	OCTOPAI_WH	421ms	4/3/2023, 11:56…
SHOW SEQUENCES;	01ab6328-0101-e101-0000-709901ac659a	Success	DORA_TEST	—	88ms	4/3/2023, 11:56…
SHOW WAREHOUSES;	01ab6328-0101-e101-0000-709901ac6596	Success	DORA_TEST	—	64ms	4/3/2023, 11:56…
SELECT * FROM INFORMATION_SCHEMA.PIPES;	01ab6328-0101-dd85-0000-709901ac5e5a	Success	DORA_TEST	OCTOPAI_WH	149ms	4/3/2023, 11:56…
SELECT * FROM INFORMATION_SCHEMA.PROCED…	01ab6328-0101-e101-0000-709901ac6592	Success	DORA_TEST	OCTOPAI_WH	8.7s	4/3/2023, 11:56…
SELECT * FROM INFORMATION_SCHEMA.FUNCTI…	01ab6328-0101-dd85-0000-709901ac5e56	Success	DORA_TEST	OCTOPAI_WH	5.0s	4/3/2023, 11:56…
SELECT * FROM INFORMATION_SCHEMA.VIEWS …	01ab6328-0101-dd85-0000-709901ac5e52	Success	DORA_TEST	OCTOPAI_WH	8.0s	4/3/2023, 11:56…
SELECT * FROM INFORMATION_SCHEMA.COLUMN…	01ab6327-0101-e101-0000-709901ac6546	Success	DORA_TEST	OCTOPAI_WH	41s	4/3/2023, 11:55…
SELECT * FROM INFORMATION_SCHEMA.TABLES…	01ab6327-0101-e101-0000-709901ac652a	Success	DORA_TEST	OCTOPAI_WH	21s	4/3/2023, 11:55…
<redacted>	01ab6327-0101-dd85-0000-709901ac5de2	Failed	DORA_TEST	—	71ms	4/3/2023, 11:55…

Revoke warehouse permissions

Use revoke statements when access must be removed:

```
REVOKE <privilege> ON WAREHOUSE <warehouse_name> FROM ROLE <role_name>;
```

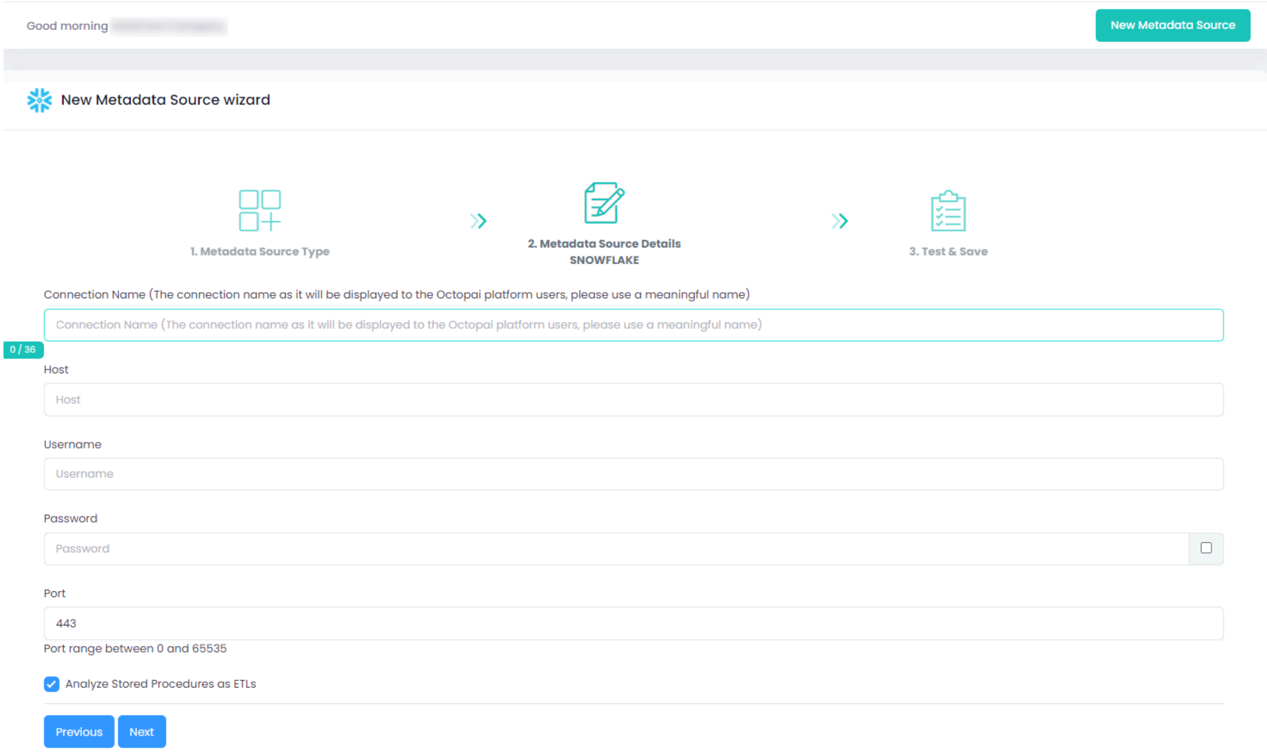
Example:

```
REVOKE MONITOR, OPERATE, USAGE ON WAREHOUSE my_warehouse FROM ROLE my_role;
```

Set up the Snowflake metadata source

Configure metadata sources through the Cloudera Octopai Client.

Figure 26: Metadata source configuration in the OC



Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Host

Username

Password

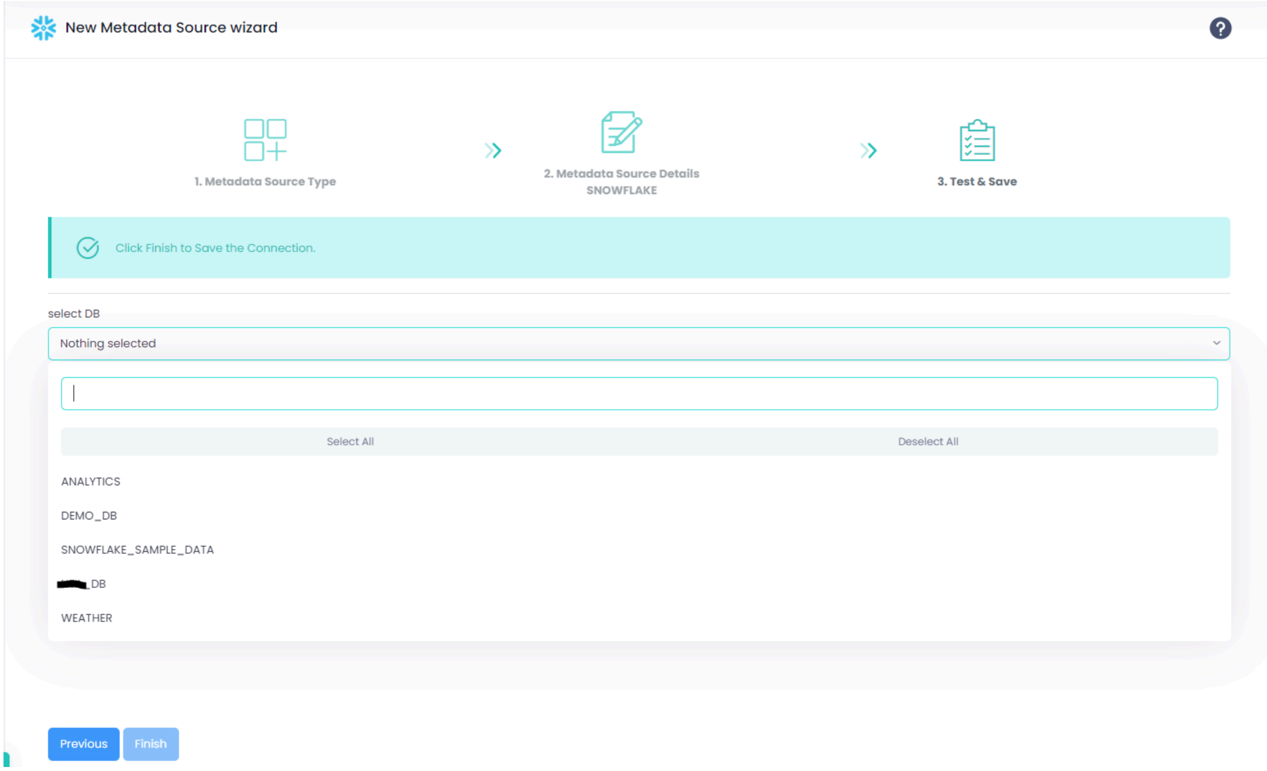
Port: 443

Port range between 0 and 65535

☑ Analyze Stored Procedures as ETLs

After clicking Next, select the databases to scan using the **DB List** parameter. The list reflects the Snowflake access you granted.

Figure 27: Database selection list



Click Finish to Save the Connection.

select DB: Nothing selected

ANALYTICS

DEMO_DB

SNOWFLAKE_SAMPLE_DATA

[redacted]_DB

WEATHER

Enhanced Snowflake connector: key pair authentication

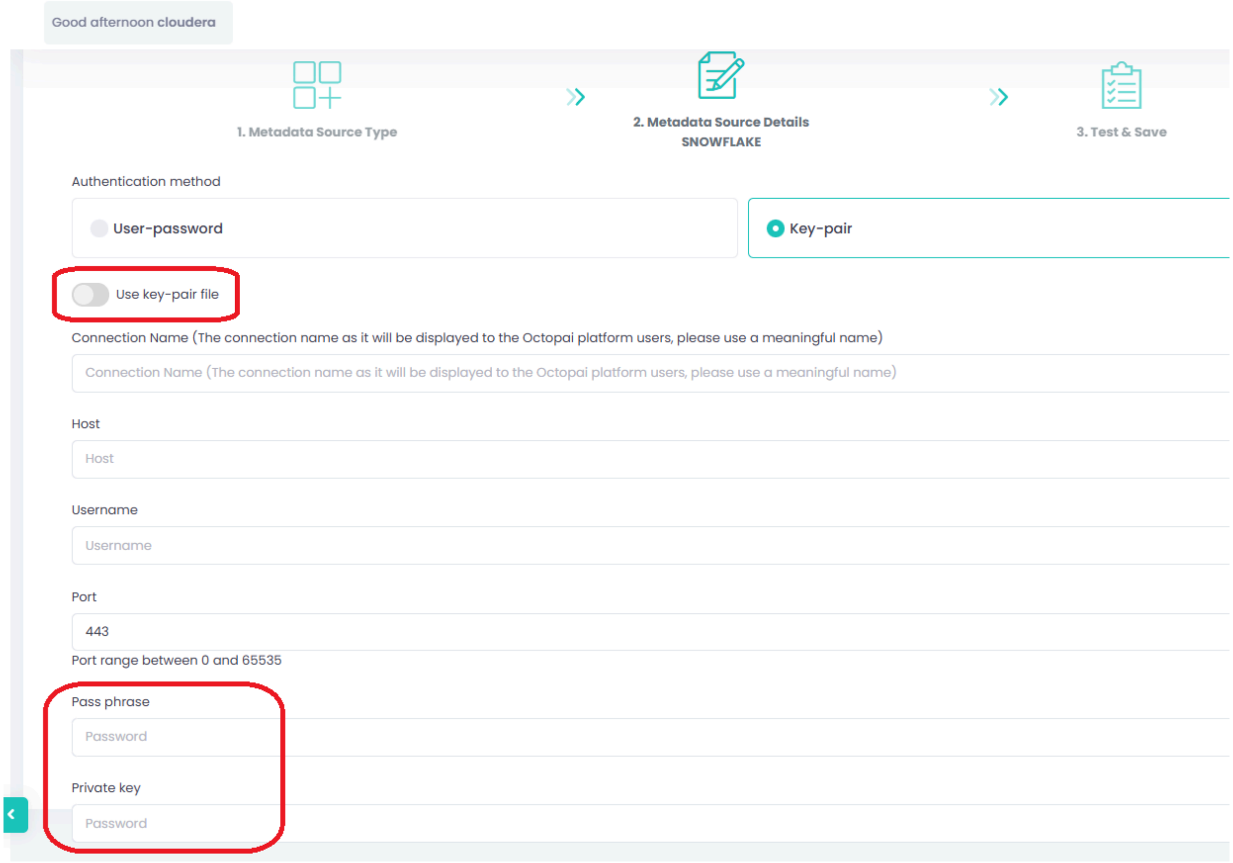
Cloudera Octopai supports **key pair authentication**, a secure alternative to passwords as Snowflake deprecates legacy methods.

Authentication options

You can configure key pair authentication in two ways:

- Paste the encrypted private key and passphrase directly into the Cloudera Octopai Agent configuration.

Figure 28: Pasting an encrypted private key



Good afternoon cloudera

Authentication method

○ User-password ● Key-pair

Use key-pair file

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Host

Host

Username

Username

Port

443

Port range between 0 and 65535

Pass phrase

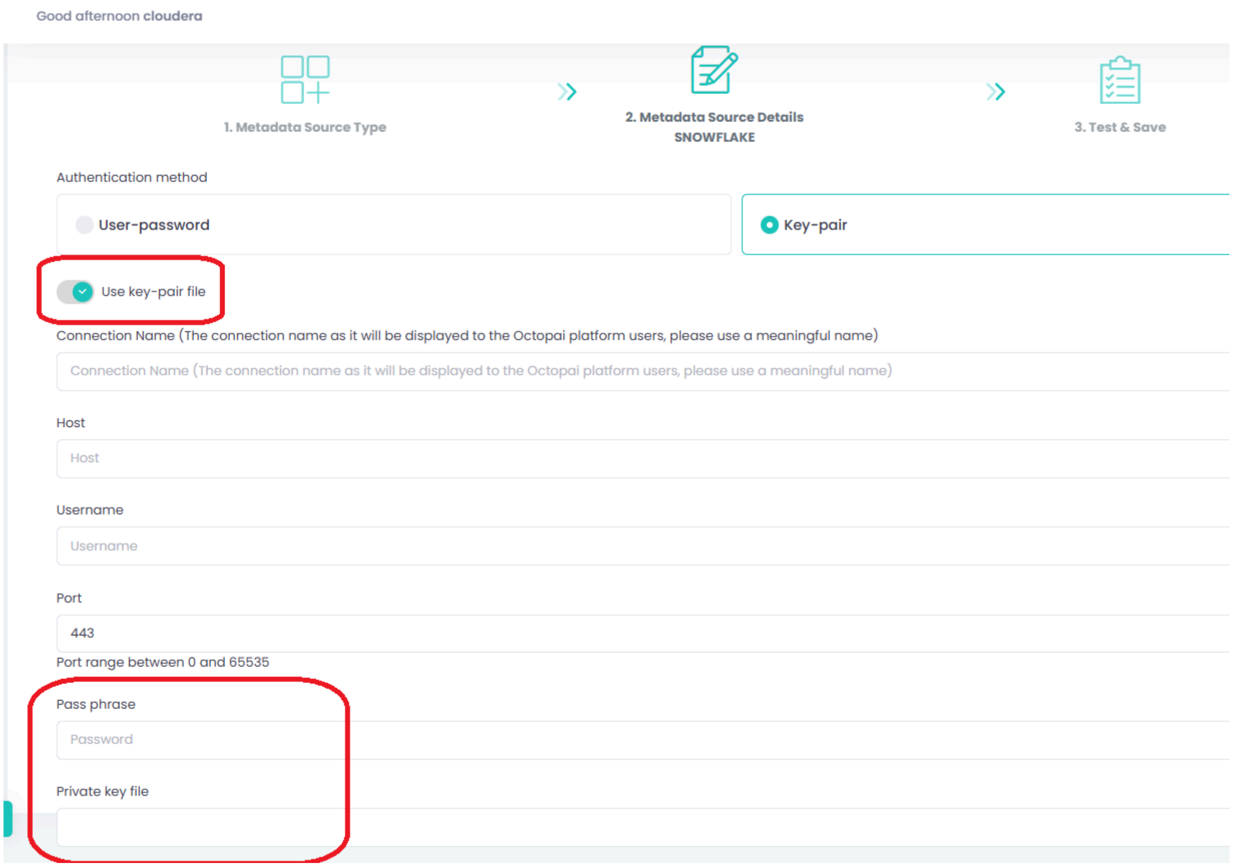
Password

Private key

Password

- Provide the file path to the encrypted private key along with its passphrase.

Figure 29: Referencing a private key file



Good afternoon cloudera

Authentication method

○ User-password

● Key-pair

☑ Use key-pair file

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Host

Host

Username

Username

Port

443

Port range between 0 and 65535

Pass phrase

Password

Private key file

Both methods ensure secure authentication during metadata extraction.

Why migrate from passwords

Snowflake is retiring single-factor password authentication:

- **November 2025:** password-only sign-ins are blocked for service and human users. See Snowflake announcement and community update.
- **March 2026:** programmatic access for password-based legacy service accounts is fully disabled. Refer to Snowflake MFA rollout documentation.

Supported authentication methods will be multi-factor (SAML/OAuth) and key pair authentication.

Recommended actions

- Configure key pair authentication in Cloudera Octopai using one of the available methods.
- Ensure the Snowflake service account is marked as a SERVICE user (not LEGACY_SERVICE).
- Password-based service account access in Cloudera Octopai will no longer be supported after March 2026. Plan your migration to avoid service disruptions.

Review the full setup steps in the Snowflake Key Pair Authentication Guide.

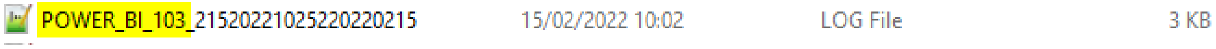
Verify the extracted metadata files

Troubleshoot extraction issues

If extraction fails:

- Confirm Snowflake permissions.
- Send logs (including connector number and name) to Cloudera Support. Logs reside at **C:\Program Files (x86)\Octopai\Service\log**.

  Figure 30: Example log files

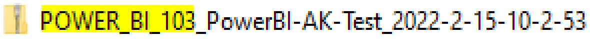
  

  POWER_BI_103_21520221025220220215 15/02/2022 10:02 LOG File 3 KB

Access the Cloudera Octopai target folder

1. On the server hosting the Cloudera Octopai Client, open the TGT folder (default: **C:\Program Files (x86)\Octopai\Service\TGT**).
2. Locate the ZIP file named after the connector and open it.

   Figure 31: Connector ZIP contents

   POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53
3. Review the inner files for completeness and quality.

Vertica

Learn how to set up Vertica metadata sources with Cloudera Octopai.

Note: Version supported: up to v12.0

Tool Permissions Prerequisites

Warning: Missing permissions could end up in broken lineages.

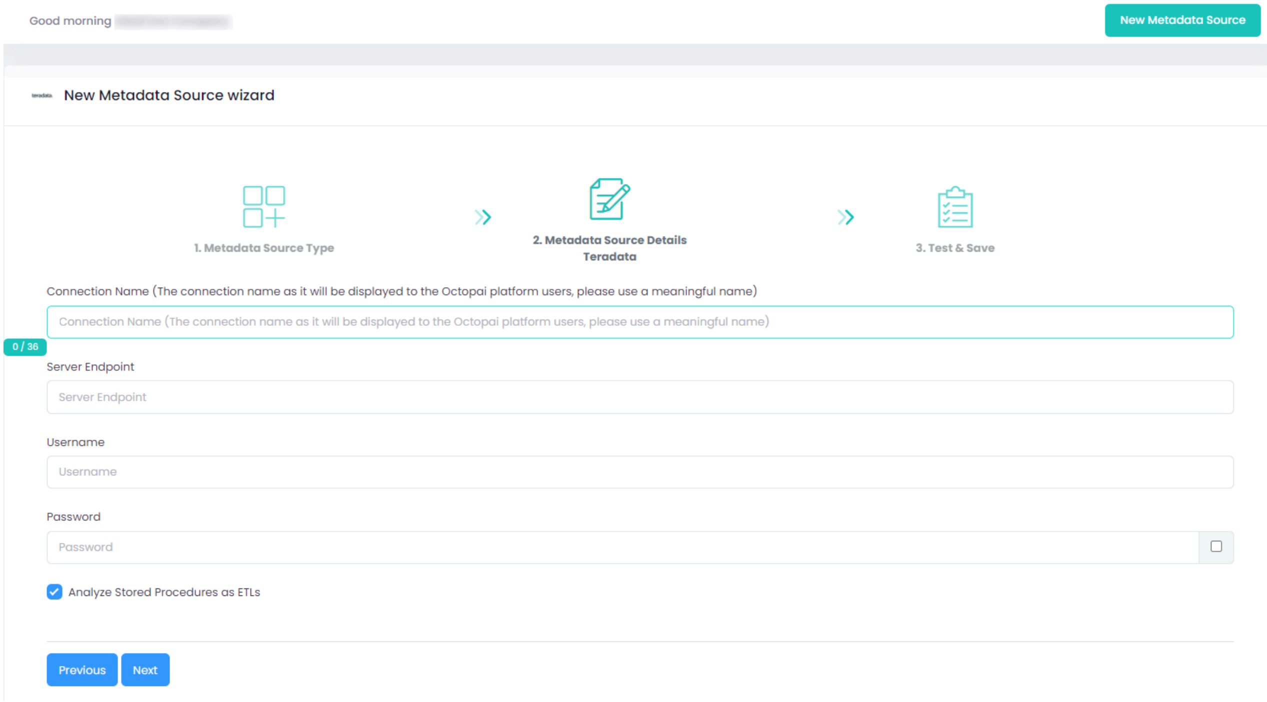
Ensure the following prerequisites are met:

- Open Server Port for each Vertica Database Connection

- Admin Vertica user for each connection with grant permission for the following dictionary tables:
  - PROJECTIONS
  - TEXT_INDICES
  - USER_FUNCTIONS
  - USER_PROCEDURES
  - USER_TRANSFORMS
  - VIEW_COLUMNS
  - VIEWS
  - VIEW_TABLES
  - ALL_TABLES
  - DATABASES
  - COLUMNS
  - COMMENTS
  - HCATALOG_TABLES
  - CONSTRAINT_COLUMNS
  - HCATALOG_COLUMNS
  - HCATALOG_SCHEMATA
  - HCATALOG_TABLE_LIST
  - PROJECTION_COLUMNS

Setting up Vertica Metadata Source

Metadata Sources are set on the Cloudera Octopai Client.



Good morning [redacted]

New Metadata Source wizard

1. Metadata Source Type

2. Metadata Source Details
Teradata

3. Test & Save

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Server Endpoint

Server Endpoint

Username

Username

Password

Password

☑ Analyze Stored Procedures as ETLs

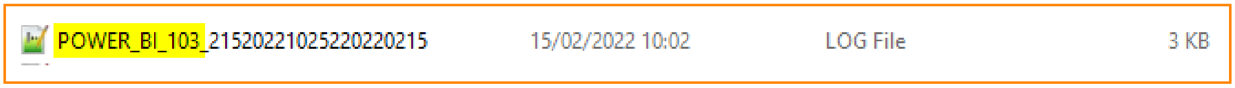
How to verify the extracted Metadata File

Troubleshoot

Error during the extraction:

- Check the permissions

- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service \log



POWER_BI_103_21520221025220220215 15/02/2022 10:02 LOG File 3 KB

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:

POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Netezza

Learn how to set up Netezza metadata sources with the Cloudera Octopai Client.

Note: Version supported: v7.2

Tool Permissions Prerequisites

Warning:

Missing permissions could end up in broken lineages.

Check/Install Netezza ODBC on the server of the Cloudera Octopai Client

Open Server Port for each Netezza Database Connection

Admin Netezza user for each connection with grant permission for the following dictionary tables:

- SYSTEM.ADMIN(ADMIN)=> select * from _v_sys_columns;
- SYSTEM.ADMIN(ADMIN)=> select * from _v_synonym;
- SYSTEM.ADMIN(ADMIN)=> select * from _v_objects;
- SYSTEM.ADMIN(ADMIN)=> select * from _v_database;
- SYSTEM.ADMIN(ADMIN)=> select * from _v_procedure;
- SYSTEM.ADMIN(ADMIN)=> select * from _v_table;
- SYSTEM.ADMIN(ADMIN)=> select * from _v_view;

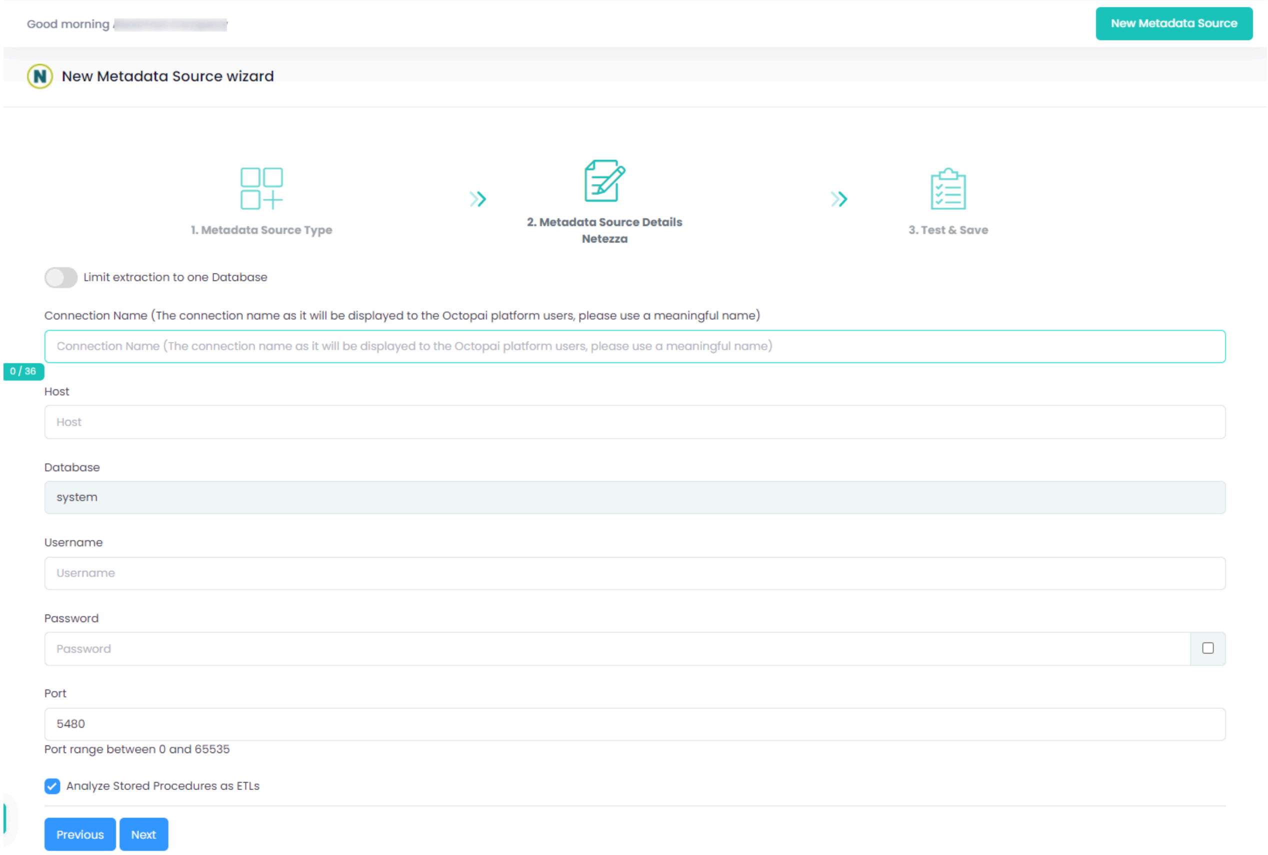
Only if previous grants are not enough:

Netezza Grant Permission Steps:

- SYSTEM.ADMIN(ADMIN)=> CREATE USER <UserName> WITH PASSWORD '<UserPassword>';
- SYSTEM.ADMIN(ADMIN)=> GRANT SELECT ON ALL.ALL.TABLE TO <UserName>;
- SYSTEM.ADMIN(ADMIN)=> GRANT SELECT ON _T_DATABASE TO <UserName>;
- SYSTEM.ADMIN(ADMIN)=> SET CATALOG <Database Name>;
- <Database Name>.ADMIN(ADMIN)=> GRANT LIST ON DATABASE to <UserName>;
- <Database Name>.ADMIN(ADMIN)=> GRANT ALL ON TABLE, FUNCTION, AGGREGATE, MATERIALIZED VIEW, PROCEDURE, SEQUENCE, SYNONYM, VIEW, SYSTEM VIEW TO <UserName>;

Setting up Netezza Metadata Source

Metadata Sources are set on the Cloudera Octopai Client



Good morning [redacted]

○ Limit extraction to one Database

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Host

Host

Database

system

Username

Username

Password

Password

Port

5480

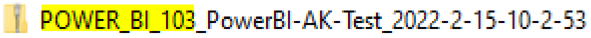
Port range between 0 and 65535

☑ Analyze Stored Procedures as ETLs

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:



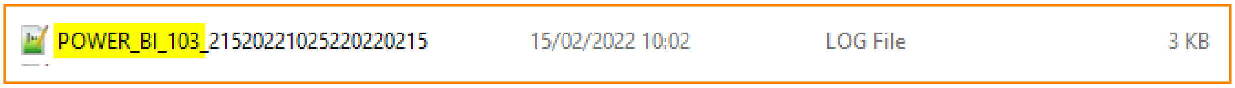
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service\log



POWER_BI_103_21520221025220220215 15/02/2022 10:02 LOG File 3 KB

Teradata

Learn how to set up and verify Teradata metadata sources using Cloudera Octopai.

Note: Version supported: up to v17 / Vantage v17

Tool Permissions Prerequisites

Warning:

Missing permissions could end up in broken lineages.

Open Server Port for each Teradata Database Connection

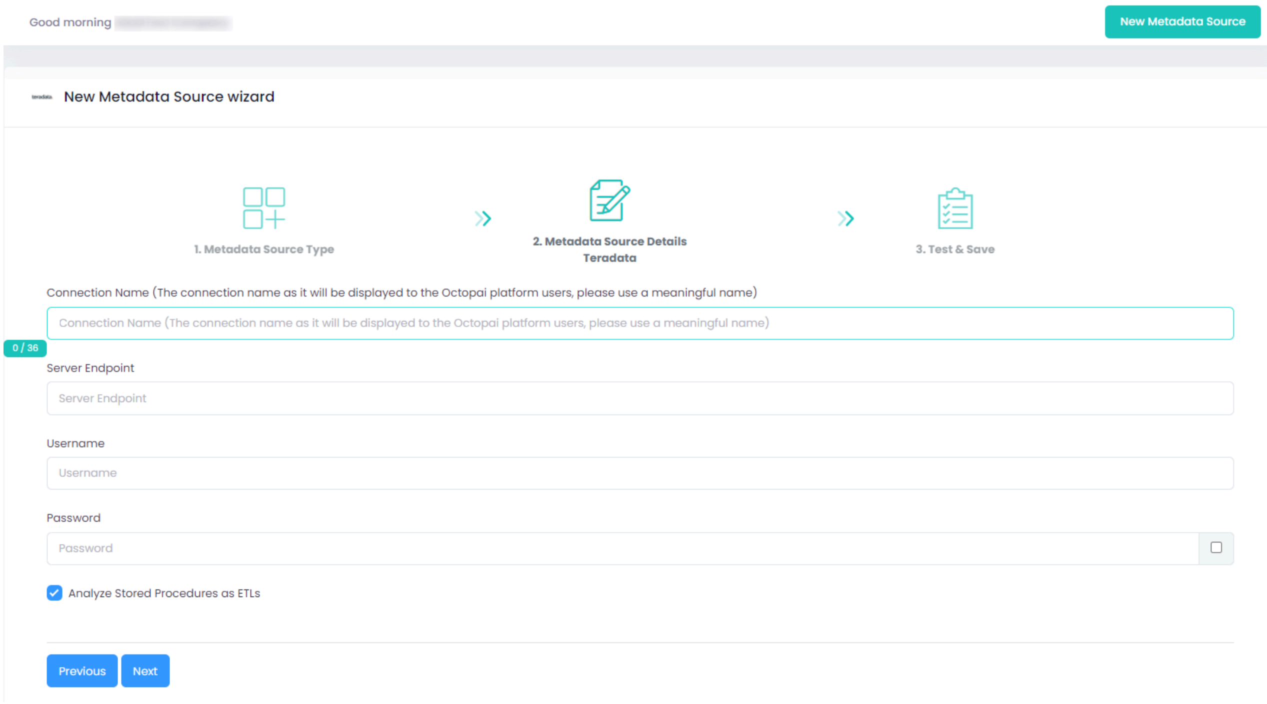
Exiting/New Teradata user (**OCTOPAI_USER***) for each connection with grant read permission for the following dictionary tables:

- dbc.tablesv
- dbc.indicesV
- dbc.tvm
- DBC.TVFields
- DBC.triggersTbl
- dbc.ColumnsV
- dbc.dbase
- dbc.texttbl
- dbc.tablesv

Setting up Teradata Metadata Source

Metadata Sources are set on the Cloudera Octopai Client

Figure 32: New Metadata Source wizard



Good morning [redacted]

New Metadata Source wizard

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Server Endpoint

Server Endpoint

Username

Username

Password

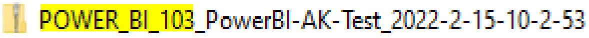
Password

☑ Analyze Stored Procedures as ETLs

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:



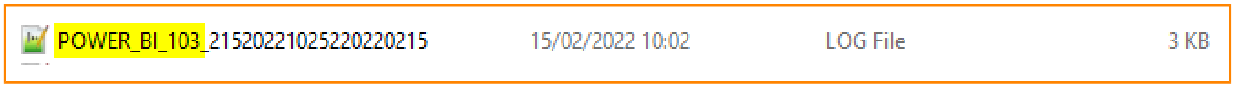
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service\log



POWER_BI_103_21520221025220220215 15/02/2022 10:02 LOG File 3 KB

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:

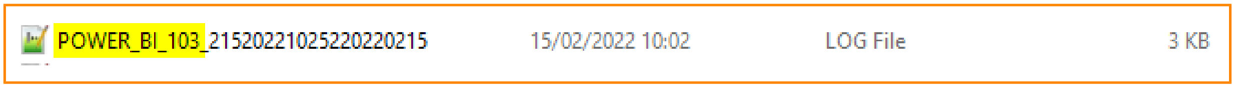
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service \log



POWER_BI_103_2152022102522022021 5 15/02/2022 10:02 LOG File 3 KB

Oracle (DWH)

Learn how to configure and manage Oracle (DWH) metadata sources using Cloudera Octopai.

Note: Version supported: 11g.x, 12c, 12.1

Tool Permissions Prerequisites

Warning:

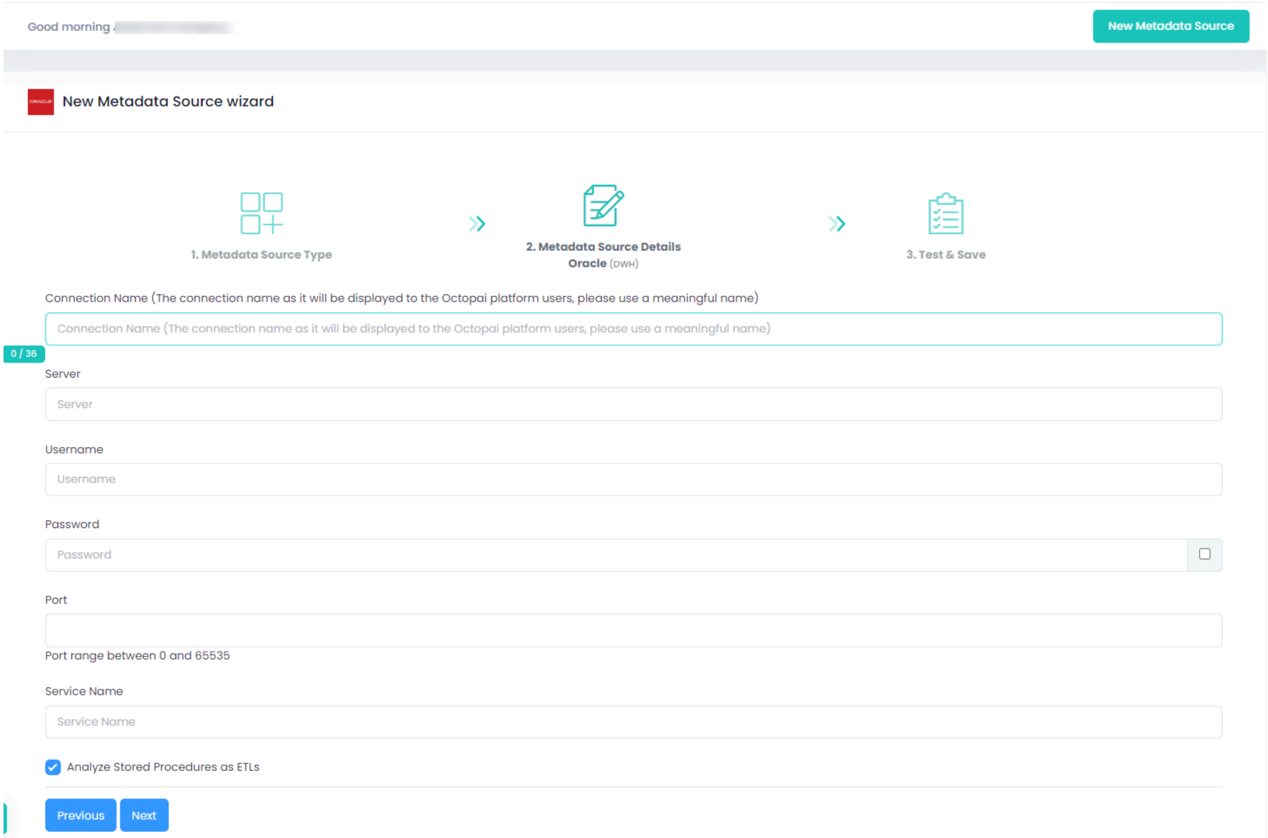
Missing permissions could end up in broken lineages.

- Open Server Port for each Oracle Database Connection
- Existing/New Oracle user for each database with 'READ' permissions for the following dictionary tables:
  - DBA_INDEXES
  - DBA_SOURCE
  - DBA_DEPENDENCIES
  - DBA_CATALOG
  - DBA_VIEWS
  - DBA_TAB_COLUMNS
  - DBA_IND_COLUMNS
  - DBA_SYNONYMS
  - DBA_MVIEWS

Setting up Oracle Metadata Source

Metadata Sources are set on the Cloudera Octopai Client

Figure 33: New Metadata Source wizard



Good morning [redacted]

New Metadata Source

New Metadata Source wizard

1. Metadata Source Type

2. Metadata Source Details
Oracle (DWH)

3. Test & Save

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Server

Server

Username

Username

Password

Password □

Port

Port range between 0 and 65535

Service Name

Service Name

☑ Analyze Stored Procedures as ETLs

Previous Next

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:

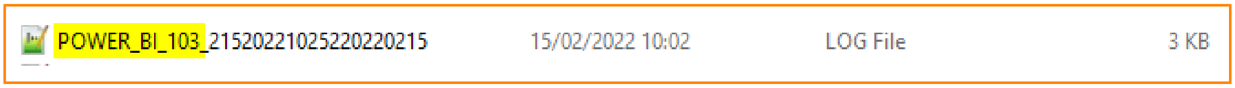
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service\log



POWER_BI_103_215202210252202202215 15/02/2022 10:02 LOG File 3 KB

Azure SQL DWH/DB & Azure Synapse Analytics

Learn how to configure and manage Azure SQL DWH/DB.

Note: Version supported: Managed \ Unmanaged

Tool Permissions Prerequisites

Warning:

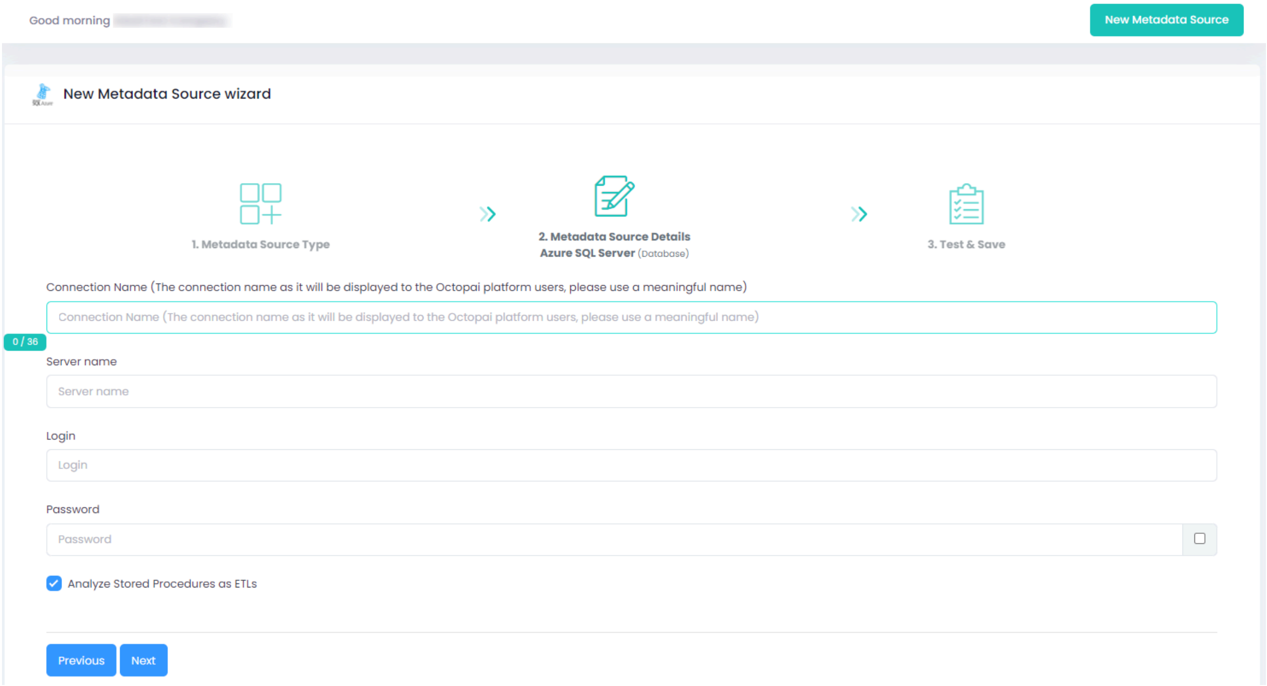
Missing permissions could end up in broken lineages.

- Open Server Port for each Oracle Database Connection
- Existing/New user (SQL Server Authentication only) for each server/database with:
  - Grant 'CONNECT' to MASTER DB
  - Grant 'SELECT' for the following dictionary tables:
    - sys.objects
    - sys.schemas
    - sys.sql_modules
    - sys.columns
    - sys.identity_columns
    - sys.computed_columns
    - sys.check_constraints
    - sys.synonyms
    - sys.indexes
    - sys.index_columns
    - sys.tables
    - sys.foreign_keys
    - sys.foreign_key_columns
- Only if previous grants are not enough:
  - USE MASTER:
  - grant connect any database to "DOMAIN_NAME\OCTOPAI_USER";
  - grant view server state to "DOMAIN_NAME\OCTOPAI_USER";
  - grant view any definition to "DOMAIN_NAME\OCTOPAI_USER";

Setting up SQL Azure DWH/DB Metadata Source

Metadata Sources are set on the Cloudera Octopai Client:

Figure 34: Azure SQL DWH/DB



Good morning [redacted]

New Metadata Source

New Metadata Source wizard

1. Metadata Source Type
2. Metadata Source Details
Azure SQL Server (Database)
3. Test & Save

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Server name

Server name

Login

Login

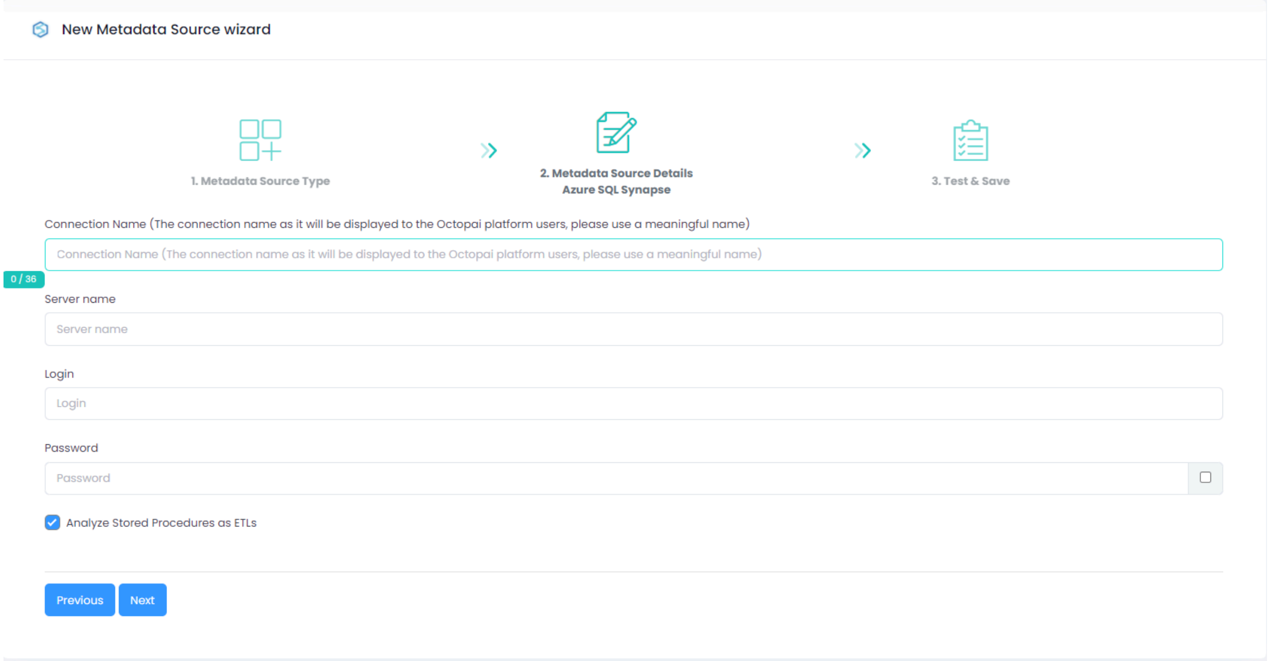
Password

Password

☑ Analyze Stored Procedures as ETLs

Previous Next

Figure 35: Azure SQL Synapse



New Metadata Source wizard

1. Metadata Source Type
2. Metadata Source Details
Azure SQL Synapse
3. Test & Save

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Server name

Server name

Login

Login

Password

Password

☑ Analyze Stored Procedures as ETLs

Previous Next

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:

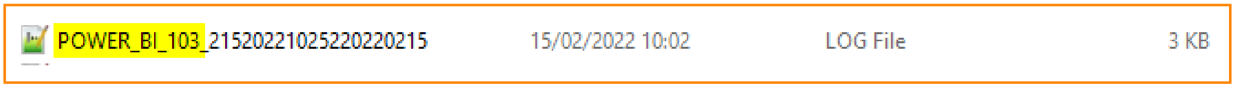
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service \log



```
POWER_BI_103_21520221025220220215    15/02/2022 10:02    LOG File    3 KB
```

SQL Server (SQLS)

Learn how to configure and manage SQL Server connectors, including tool permissions prerequisites, metadata source setup, and troubleshooting steps. This guide also provides instructions for verifying extracted metadata files and accessing the Cloudera Octopai Target Folder.

Note: **Version supported:** up to SQL 2022

Tool Permissions Prerequisites

Warning:

Missing permissions could end up in broken lineages.

- Open Server Port for each SQL Database Connection
- Existing/New user (SQL Server or Windows Authentication) for each server/database - with ‘grant SELECT’ for the following dictionary tables:
  - sys.objects
  - sys.schemas
  - sys.sql_modules
  - sys.columns
  - sys.types
  - sys.identity_columns
  - sys.computed_columns
  - sys.check_constraints
  - sys.synonyms
  - sys.indexes
  - sys.index_columns
  - sys.tables
  - sys.foreign_keys
  - sys.foreign_key_columns
  - sys.sysservers
  - sys.syslogins
  - msdb.sysjobs
  - msdb.sysjobsteps
- Grant execute on sys.sp_linkedservers

- Only if previous grants are not enough:
  - USE MASTER:
  - grant connect any database to "DOMAIN_NAME\OCTOPAI_USER";
  - grant view server state to "DOMAIN_NAME\OCTOPAI_USER";
  - grant view any definition to "DOMAIN_NAME\OCTOPAI_USER";
- USE MSDB:
  - grant select on msdb.dbo.sysjobsteps to "DOMAIN_NAME\OCTOPAI_USER";
  - grant select on msdb.dbo.sysjobs to "DOMAIN_NAME\OCTOPAI_USER";

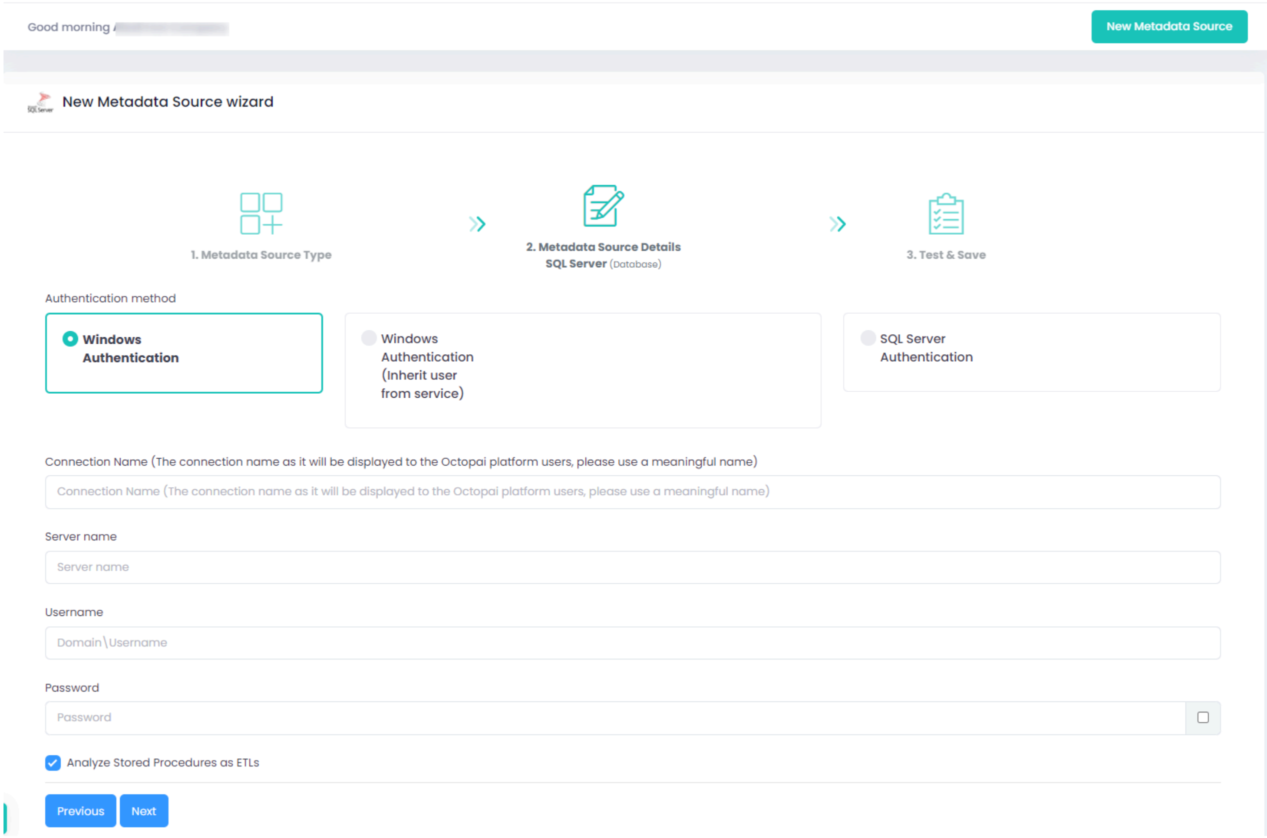
Setting up SQL Server Metadata Source

Windows Authentication: Use the server name, domain, username, and password that are recognized by the SQL Server.

Windows Authentication (inherit user from service): Run the service as another user. Important Clarification: If you choose to run the service as another user, please ensure that this user has the appropriate permissions for ALL tools used by the Cloudera Octopai Client. Otherwise, the other connectors will not be able to authenticate and perform the metadata extraction.

SQL Server Authentication: Log in with an SQL Server user.

Figure 36: New Metadata Source Wizard



Good morning [redacted]

New Metadata Source

New Metadata Source wizard

1. Metadata Source Type

2. Metadata Source Details
SQL Server (Database)

3. Test & Save

Authentication method

● Windows Authentication

○ Windows Authentication (Inherit user from service)

○ SQL Server Authentication

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Server name

Server name

Username

Domain\Username

Password

Password

☑ Analyze Stored Procedures as ETLs

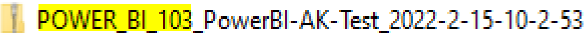
Previous Next

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**

2. Open the zip file having the Connector Name. Example:



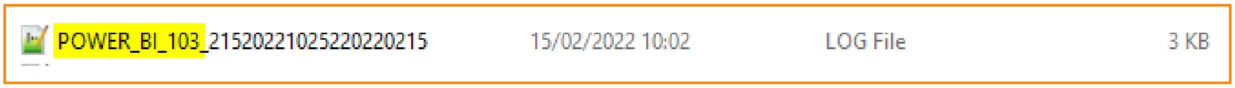
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service \log



POWER_BI_103_21520221025220220215 15/02/2022 10:02 LOG File 3 KB

SSAS (OLAP + Tabular + Azure Analysis Services)

Learn how to set up and manage SSAS (OLAP, Tabular, and Azure Analysis Services).

Note: Version supported:

- Tabular: up to 2022
- OLAP: up to 2022
- Tabular on Azure: up to 2022

Tool Permissions Prerequisites

Warning: Missing permissions could end up in broken lineages.

Missing permissions could end up in broken lineages.

SSAS - OLAP (On Prem or hosted on Azure): NT User with “Read definition” for each DB on the server. (Windows Authentication).

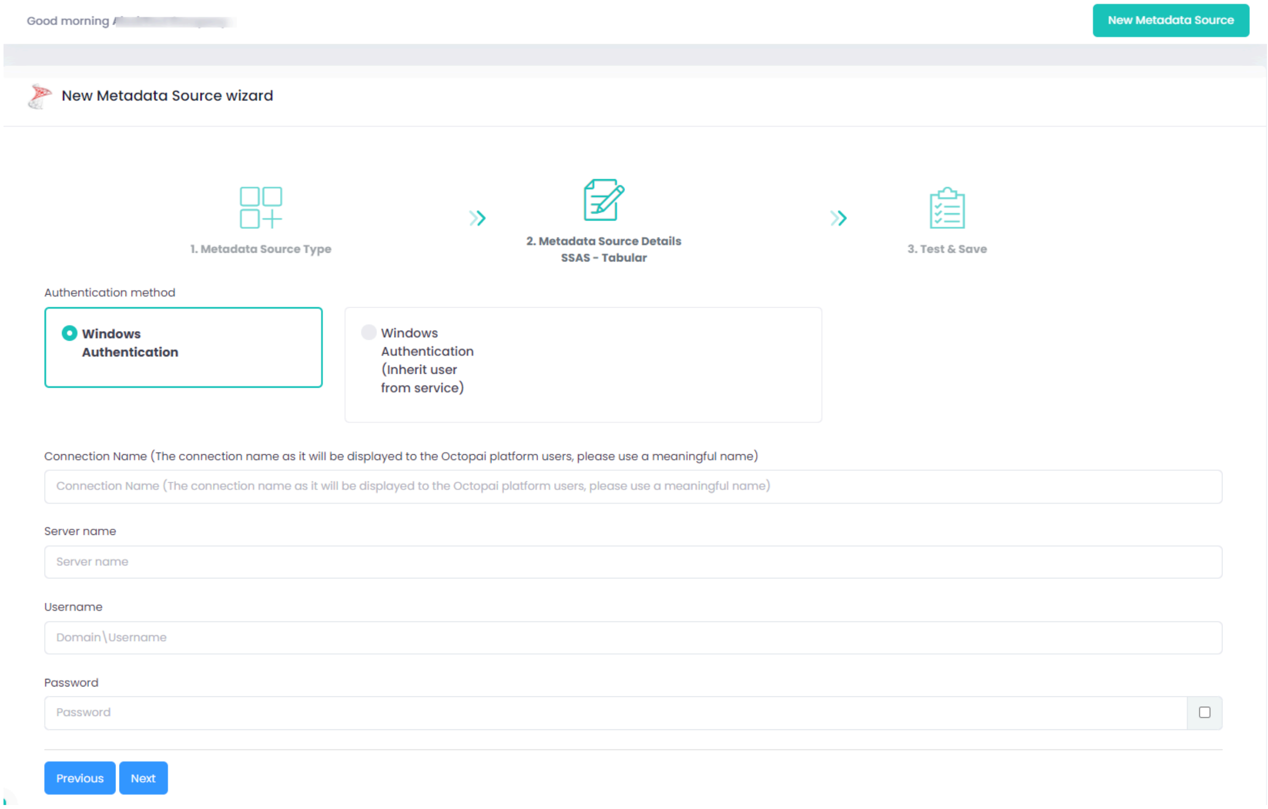
SSAS - Tabular: NT User with “Full control” for each model on the server. (Windows Authentication).

SSAS - Tabular on Azure Analysis Services (managed PaaS): 'Active Directory Password Authentication' User with “Full control” for each model on the server.

Setting up SSAS (OLAP + Tabular + Azure Analysis Services) Metadata Source

Metadata Sources are set on the Cloudera Octopai Client

Tabular:



Good morning [redacted]

New Metadata Source

New Metadata Source wizard

1. Metadata Source Type

2. Metadata Source Details
SSAS - Tabular

3. Test & Save

Authentication method

● Windows Authentication

○ Windows Authentication (Inherit user from service)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Server name

Server name

Username

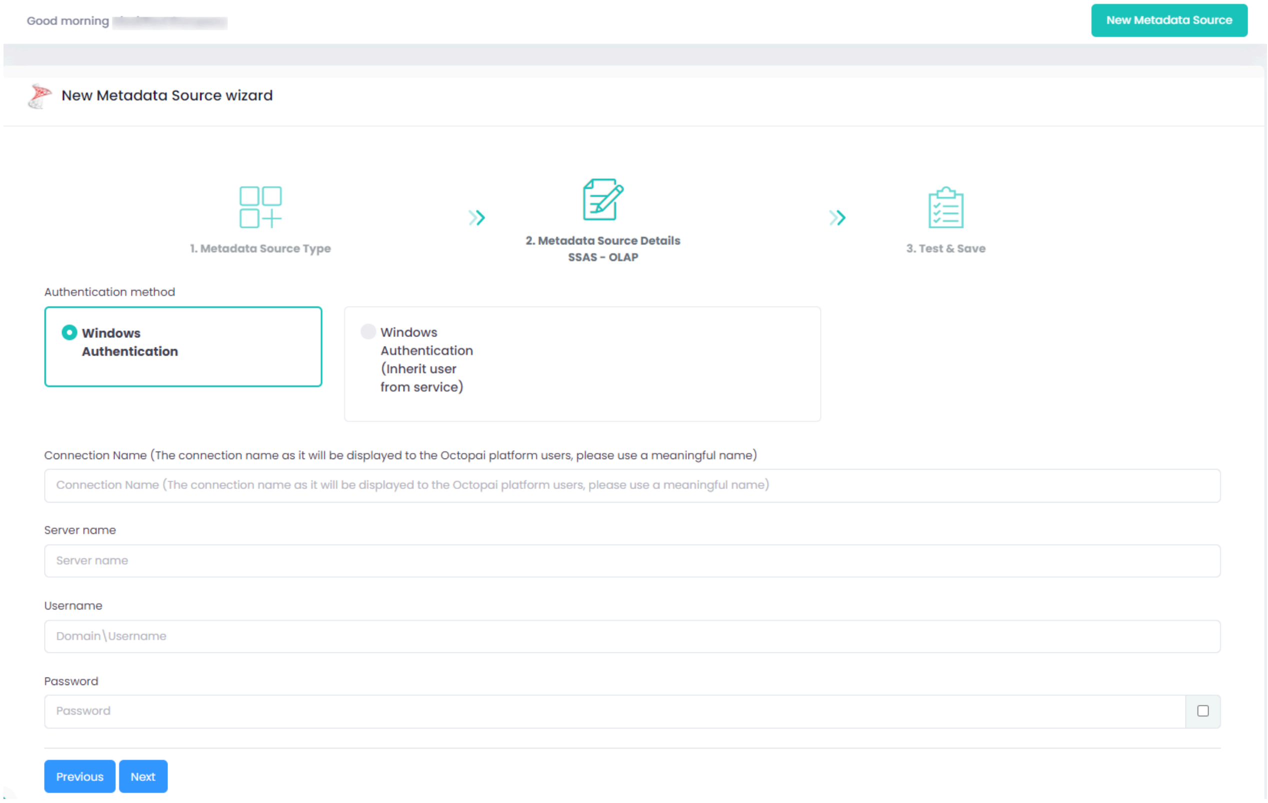
Domain\Username

Password

Password □

Previous Next

OLAP:



Good morning [redacted]

New Metadata Source

New Metadata Source wizard

1. Metadata Source Type

2. Metadata Source Details
SSAS - OLAP

3. Test & Save

Authentication method

● Windows Authentication

○ Windows Authentication (Inherit user from service)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Server name

Server name

Username

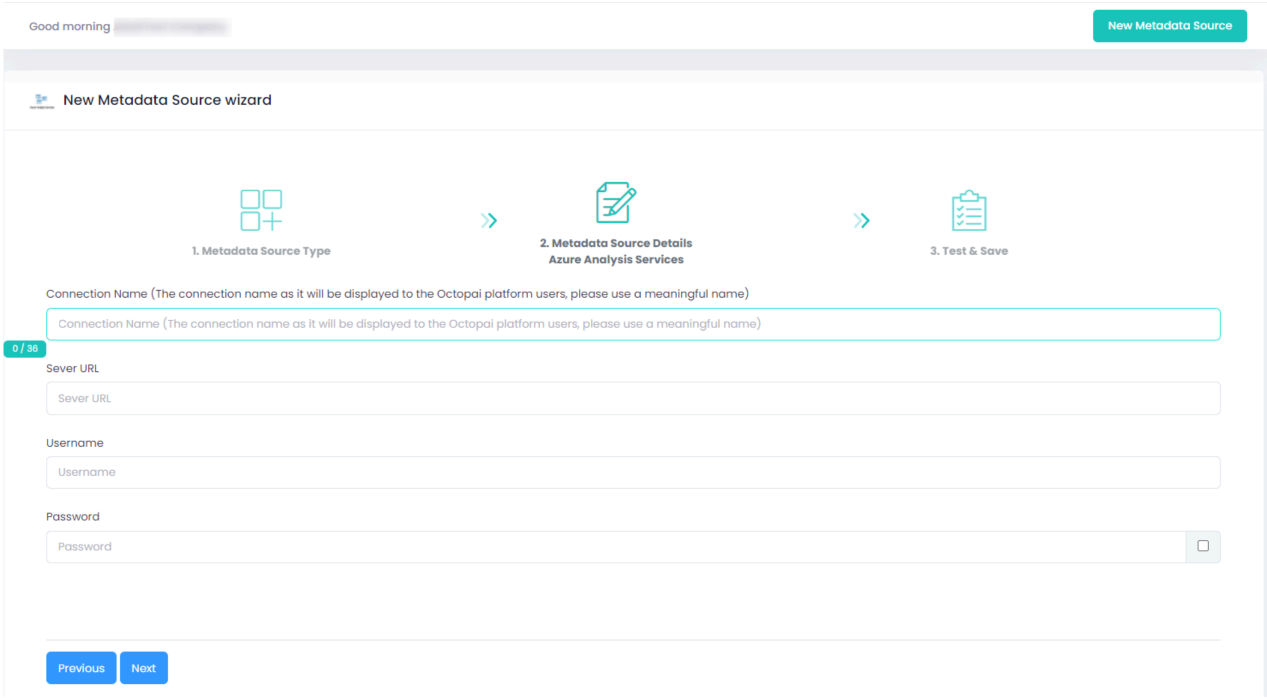
Domain\Username

Password

Password □

Previous Next

Azure Analysis Services:



Good morning [redacted]

New Metadata Source

New Metadata Source wizard

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Sever URL

Sever URL

Username

Username

Password

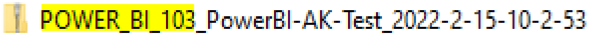
Password

Previous Next

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:

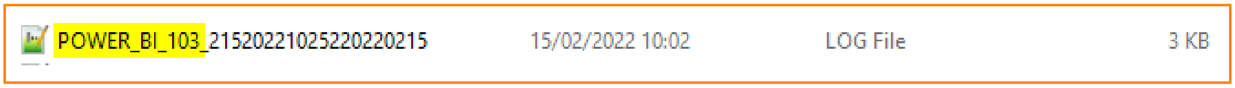
   

   POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53
3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service\log



POWER_BI_103_21520221025220220215 15/02/2022 10:02 LOG File 3 KB

MicroStrategy

Executing the Project Merge executable, projectmerge.exe (Part of MicroStrategy Tools)

Note: Version supported: up to v2023 + Cloud Version

Tool Permissions Prerequisites

⚠ **Warning:** Missing permissions could end up in broken lineages.

Ensure that the Cloudera Octopai Windows NT User has Read Permission to the MicroStrategy *.mmp or backup files folder.

How to Automate Extraction of MMP Files

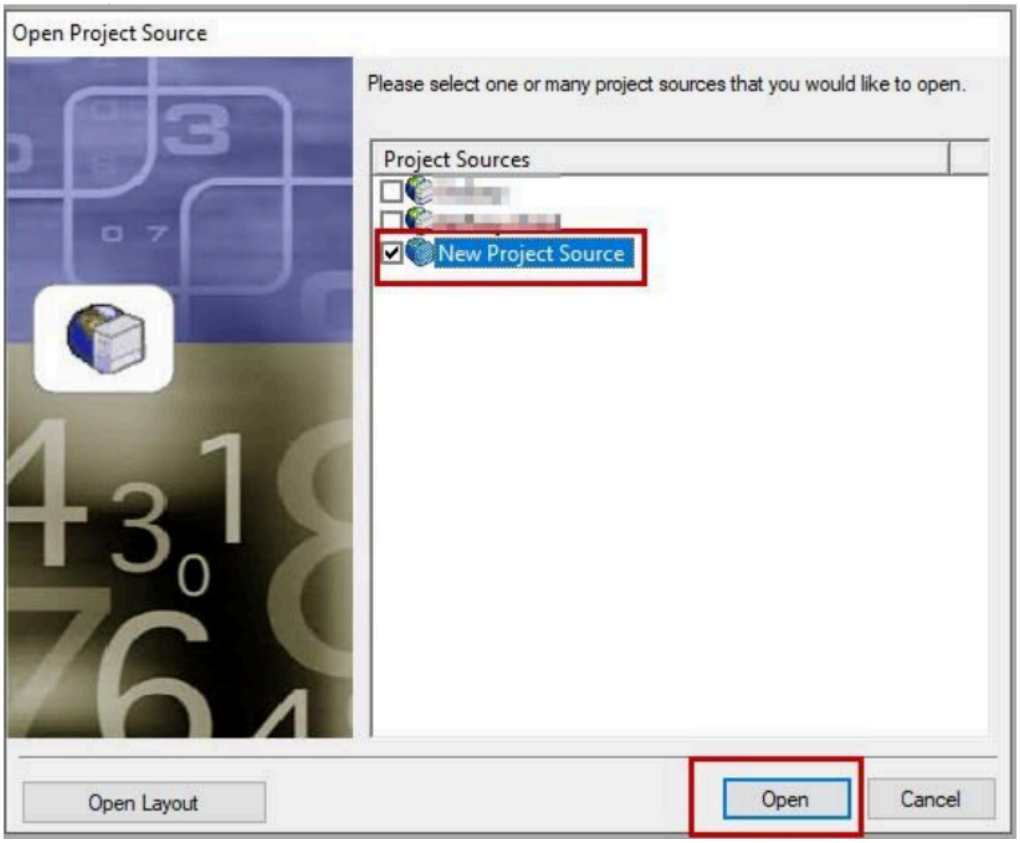
The MMP files contain the necessary metadata for the analysis by Cloudera Octopai .

Note: To automate creating the MMP files you will need the projectmerge.exe installed (part of the MicroStrategy tools)

Follow these steps to create MMP files:

1. Open the MicroStrategy Object Manager

2. Select Project Source and click on Open



Open Project Source

Please select one or many project sources that you would like to open.

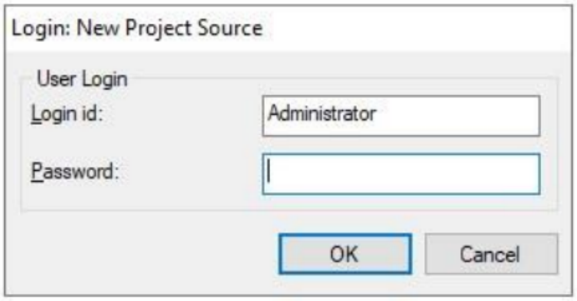
Project Sources

□ [redacted]

□ [redacted]

☑ New Project Source

3. Enter login credentials (with admin user)



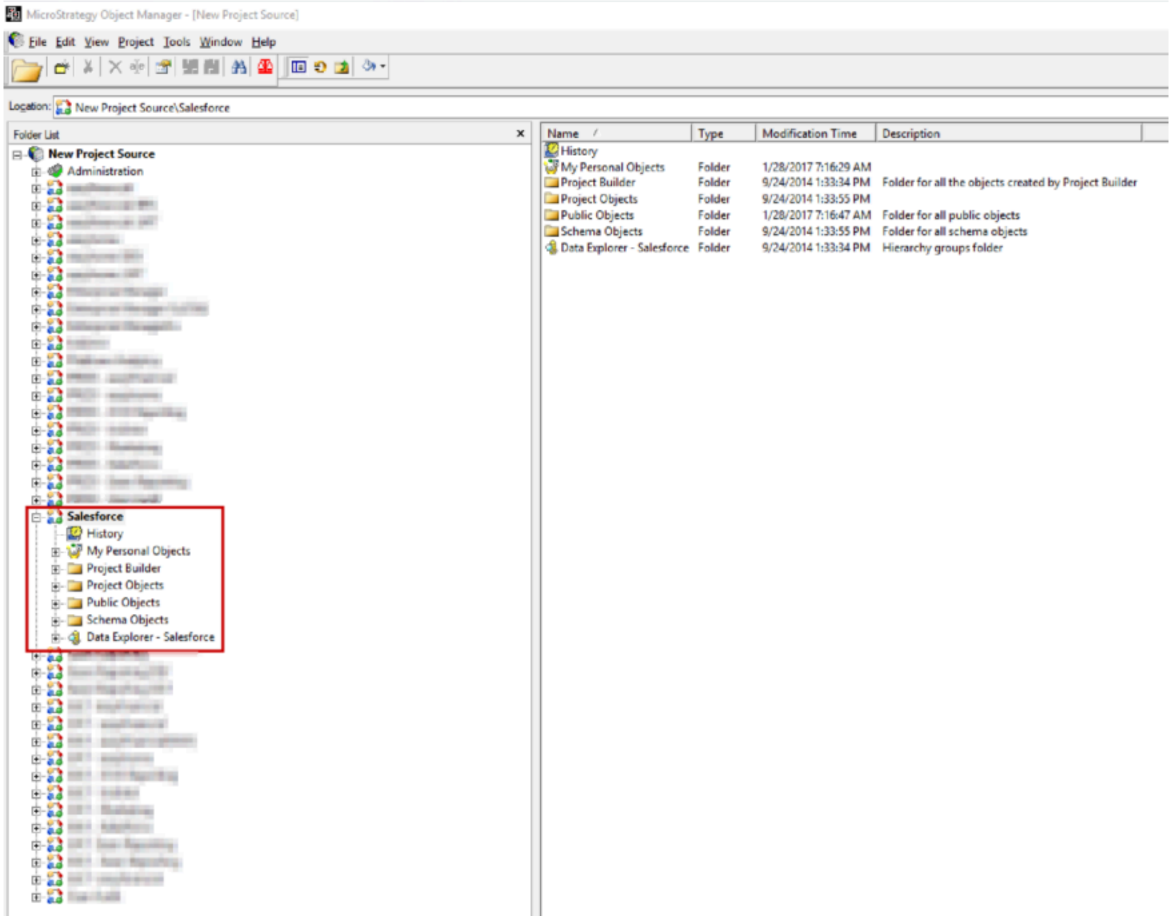
Login: New Project Source

User Login

Login id: Administrator

Password:

4. Select Project from the left side panel



MicroStrategy Object Manager - [New Project Source]
File Edit View Project Tools Window Help

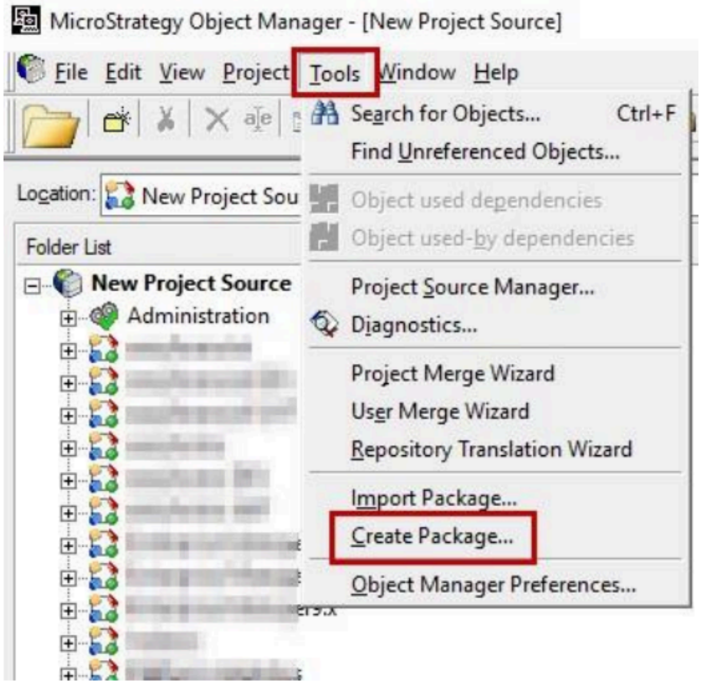
Location: New Project Source\Salesforce

Folder List

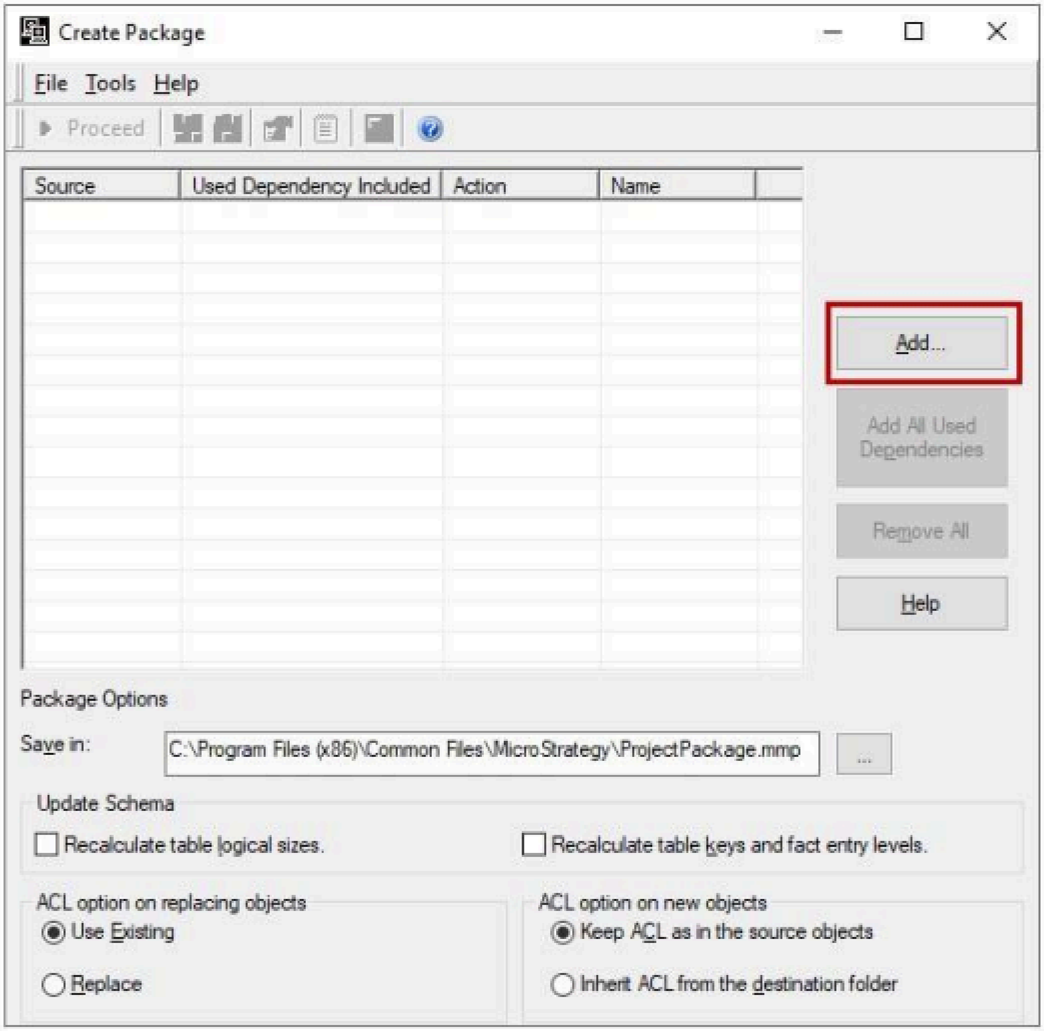
- New Project Source
  - Administration
  - [redacted]
  - Salesforce
    - History
    - My Personal Objects
    - Project Builder
    - Project Objects
    - Public Objects
    - Schema Objects
    - Data Explorer - Salesforce
  - [redacted]

Name	Type	Modification Time	Description
History			
My Personal Objects	Folder	1/28/2017 7:16:29 AM	
Project Builder	Folder	9/24/2014 1:33:34 PM	Folder for all the objects created by Project Builder
Project Objects	Folder	9/24/2014 1:33:55 PM	
Public Objects	Folder	1/28/2017 7:16:47 AM	Folder for all public objects
Schema Objects	Folder	9/24/2014 1:33:55 PM	Folder for all schema objects
Data Explorer - Salesforce	Folder	9/24/2014 1:33:34 PM	Hierarchy groups folder

5. Click on Tools --> Create Package


MicroStrategy Object Manager - [New Project Source]
File
Edit
View
Project
Tools
Window
Help
Search for Objects... Ctrl+F
Find Unreferenced Objects...
Object used dependencies
Object used-by dependencies
Project Source Manager...
Diagnostics...
Project Merge Wizard
User Merge Wizard
Repository Translation Wizard
Import Package...
Create Package...
Object Manager Preferences...
Location:
New Project Sou
Folder List
New Project Source
Administration

6. Click on Add



Package Options

Save in: C:\Program Files (x86)\Common Files\MicroStrategy\ProjectPackage.mmp

Update Schema

□ Recalculate table logical sizes. □ Recalculate table keys and fact entry levels.

ACL option on replacing objects

● Use Existing

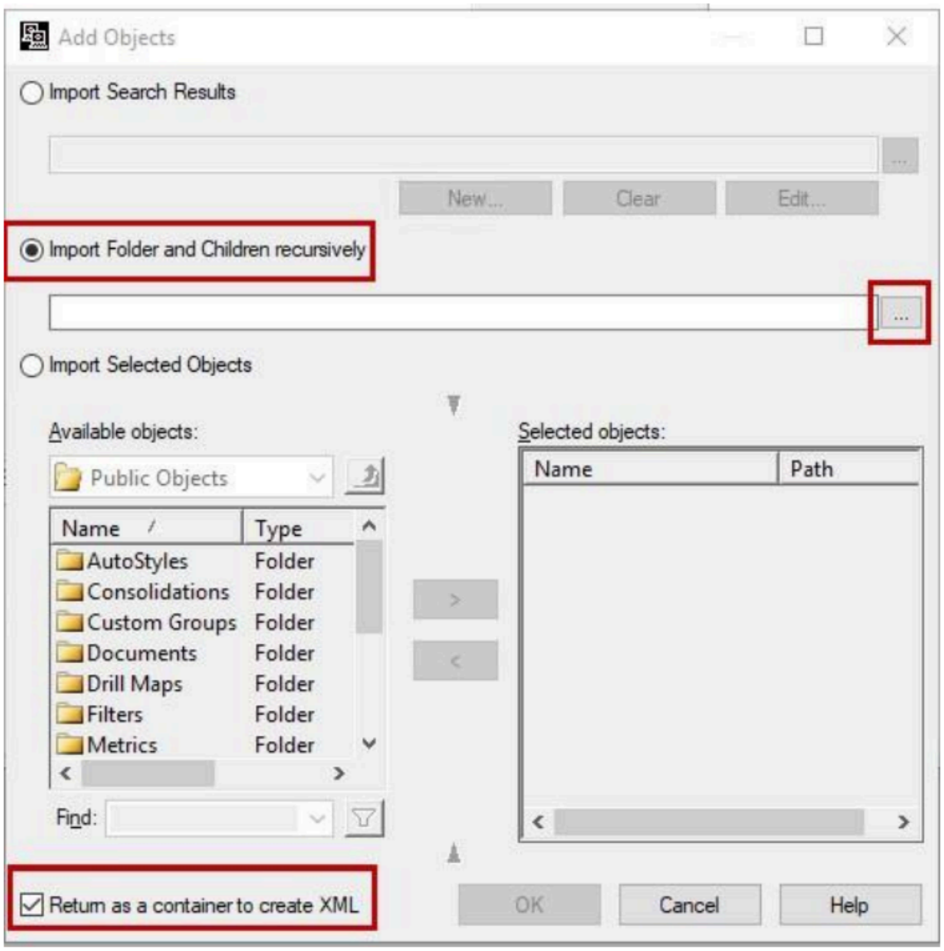
○ Replace

ACL option on new objects

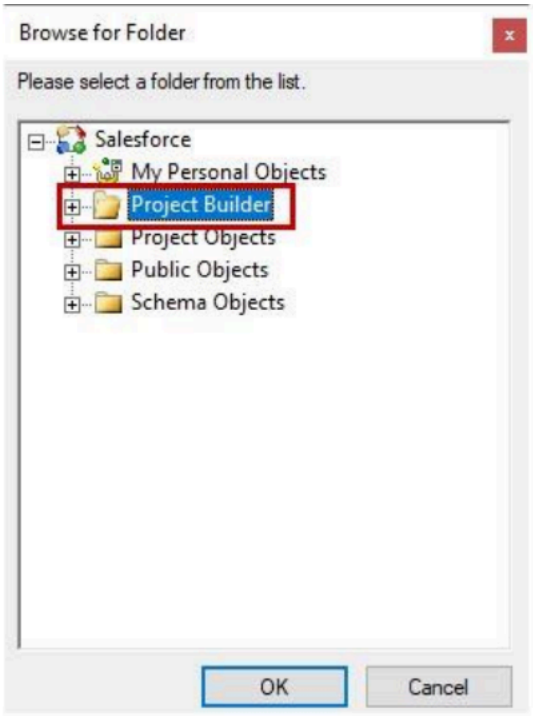
● Keep ACL as in the source objects

○ Inherit ACL from the destination folder

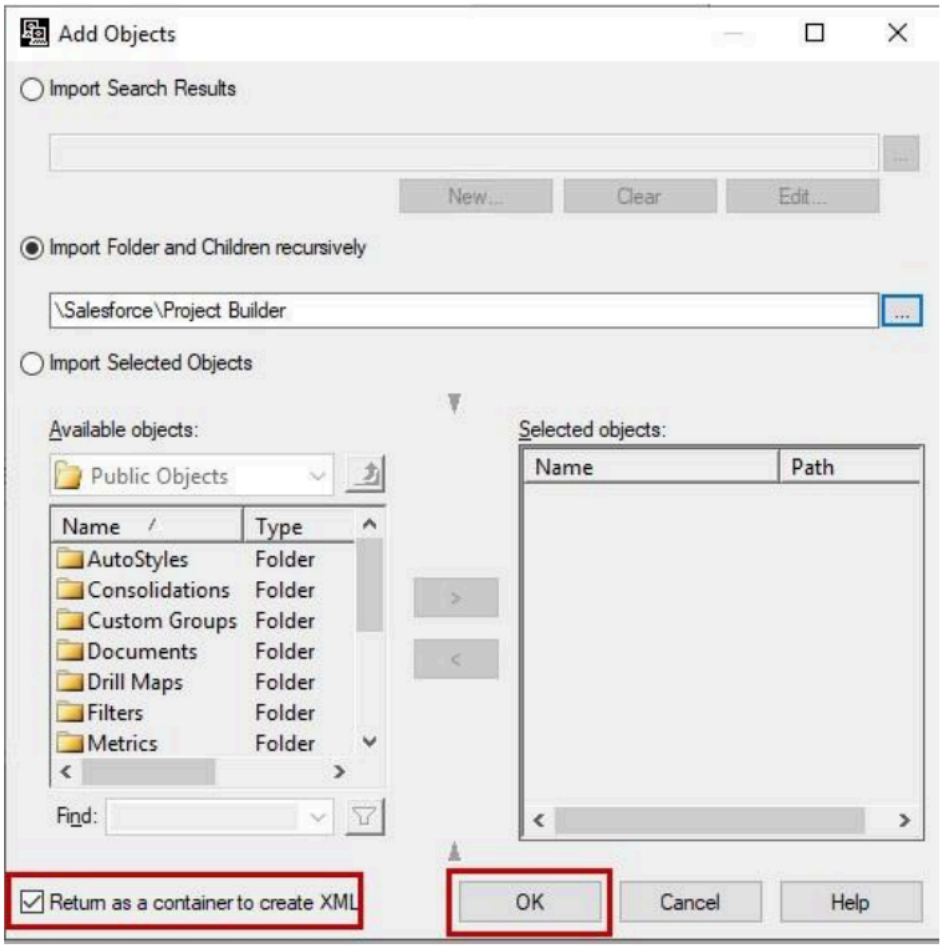
7. Check Import Folder and Children recursively



8. Select one folder and click OK



9. Check Return as a container to create XML and click OK



○ Import Search Results

● Import Folder and Children recursively

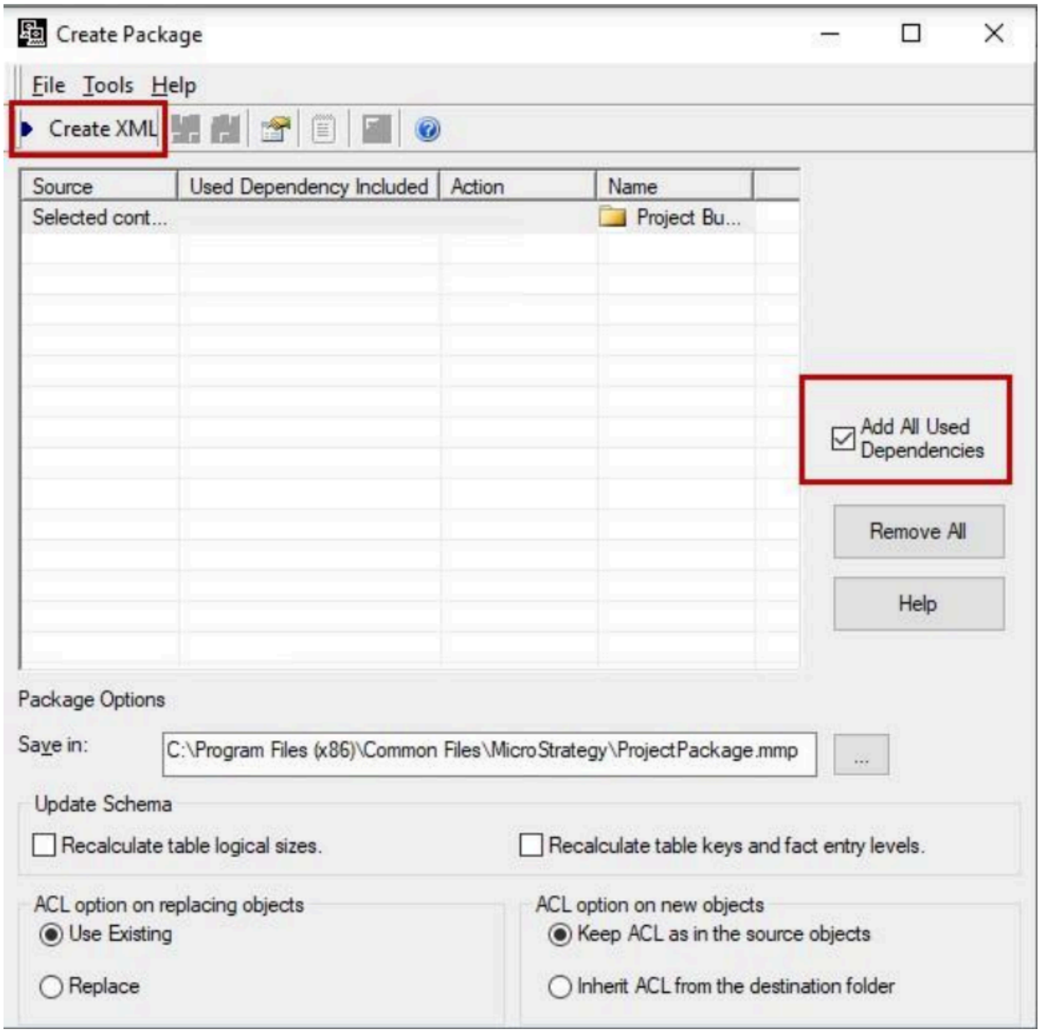
\Salesforce\Project Builder

○ Import Selected Objects

Name	Type
AutoStyles	Folder
Consolidations	Folder
Custom Groups	Folder
Documents	Folder
Drill Maps	Folder
Filters	Folder
Metrics	Folder

☑ Return as a container to create XML

10. Check "Add All Used Dependencies," change the Save in path, rename the target file as necessary, and click on Create XML



Source	Used Dependency Included	Action	Name
Selected cont...			Project Bu...

☑ Add All Used Dependencies

Package Options

Save in: C:\Program Files (x86)\Common Files\MicroStrategy\ProjectPackage.mmp

Update Schema

☐ Recalculate table logical sizes. ☐ Recalculate table keys and fact entry levels.

ACL option on replacing objects

● Use Existing

○ Replace

ACL option on new objects

● Keep ACL as in the source objects

○ Inherit ACL from the destination folder

11. Repeat steps 8-10 for each folder within the project (Normally at least 4 folders per project).

12. Repeat steps 4-11 for each project.

Using the XML file to create an updated package:

Executing the Project Merge executable, projectmerge.exe (Part of MicroStrategy Tools)

Use the user and password that you use in order to log in to the Object Manager.

Enter the following command in a batch file: c:\Users\projectmerge -f\Filename.xml -sp -sup

```
C:\Users\octopai>projectmerge -fC:\temp\T.xml -sp******** -sup
```

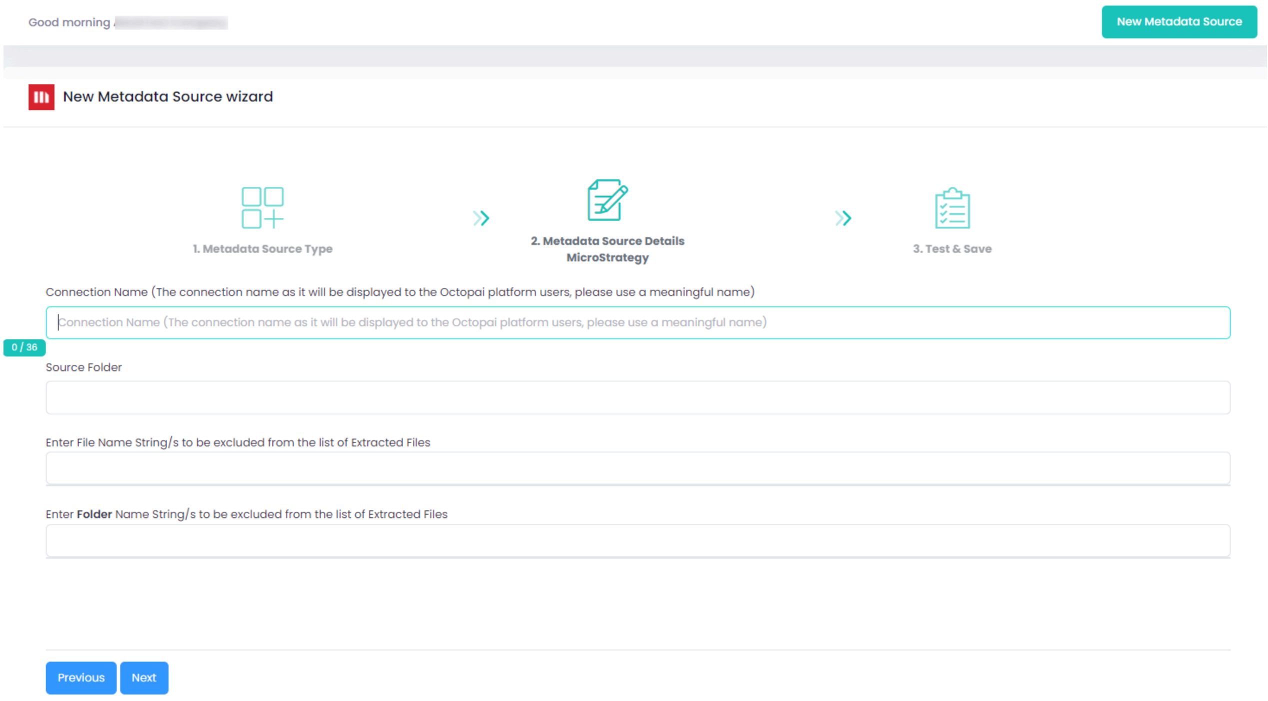
Note: Creating a package from the command line locks the project metadata for the duration of the package creation. Other users cannot make any changes to the project until it becomes unlocked.

Note: If the export gets stuck hanging, split the export to smaller files by creating separate files within each folder by selecting subfolders in step 8.

Note: Schedule the batch files one after the other with 10 minute intervals between each file. The file (*.MMP) will appear in the defined destination folder. Use these files for your Cloudera Octopai metadata source in the Cloudera Octopai Client.

Setting up MicroStrategy Metadata Source

Metadata Sources are set on the Cloudera Octopai Client :



Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Source Folder

Enter File Name String/s to be excluded from the list of Extracted Files

Enter **Folder** Name String/s to be excluded from the list of Extracted Files

Looker

Learn how to configure Looker as a metadata source in Cloudera Octopai Data Lineage.

Note: Version supported: up to 22.x

Tool Permissions Prerequisites

Warning: Missing permissions could end up in broken lineages.

Dedicated API3 Credentials (Client ID and Client Secret)

Looker API authentication | Google Cloud with the following permissions:

- access_data
- see_looks
- explore
- see_sql
- see_lookml
- use_sql_runner
- see_lookml_dashboards
- see_looks
- see_queries
- see_datagroups
- develop

We recommended creating a dedicated Permission Set and Role to associate with the credentials.

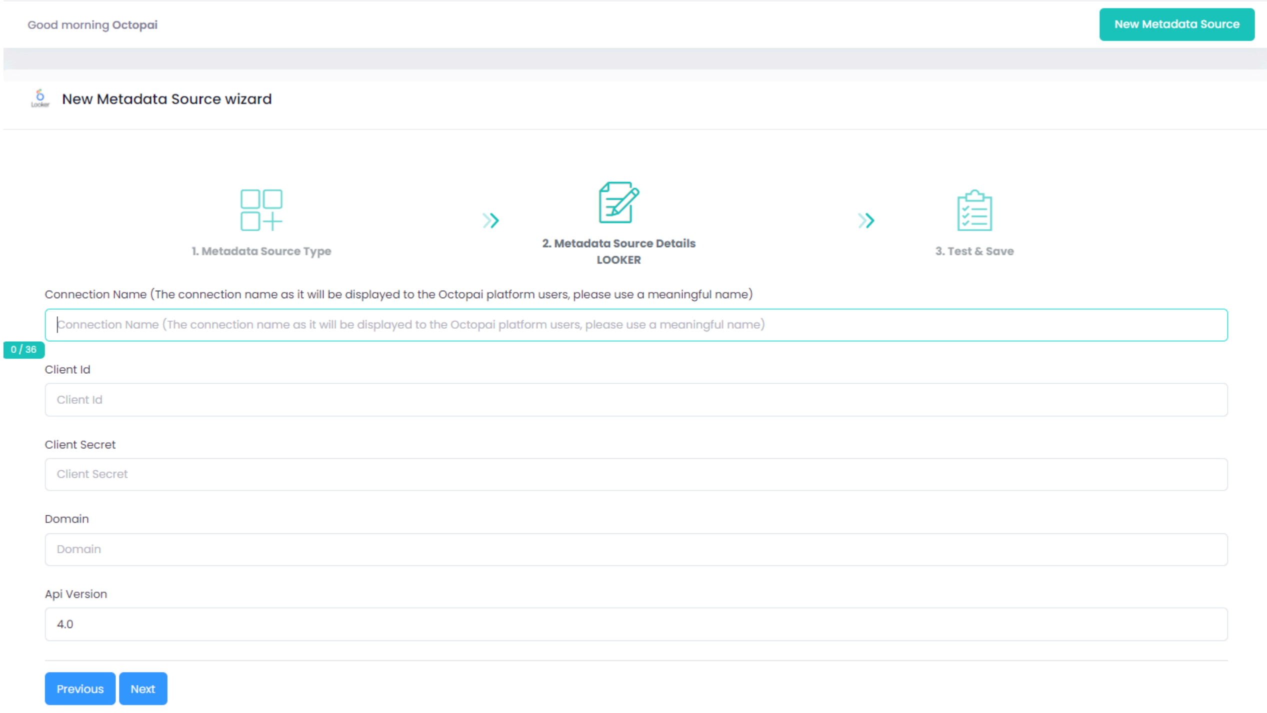
Supported API version 4.0 & up

How to set up the permissions

Looker API authentication | Google Cloud

Setting up Looker Metadata Source

Metadata Sources are set on the Cloudera Octopai Client



Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Client Id

Client Id

Client Secret

Client Secret

Domain

Domain

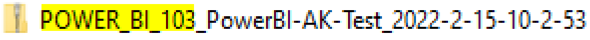
Api Version

4.0

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:



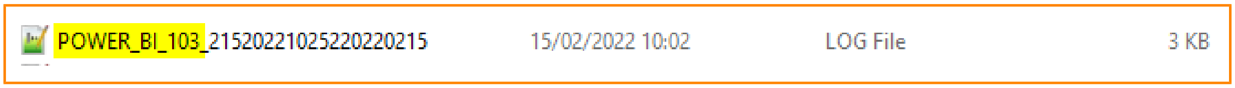
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service\log



POWER_BI_103_2152022102522022015 15/02/2022 10:02 LOG File 3 KB

SSRS

Learn how to set up and manage SSRS metadata sources using Cloudera Octopai Data Lineage.

Note: Version supported: up to SQL 2022

Tool Permissions Prerequisites

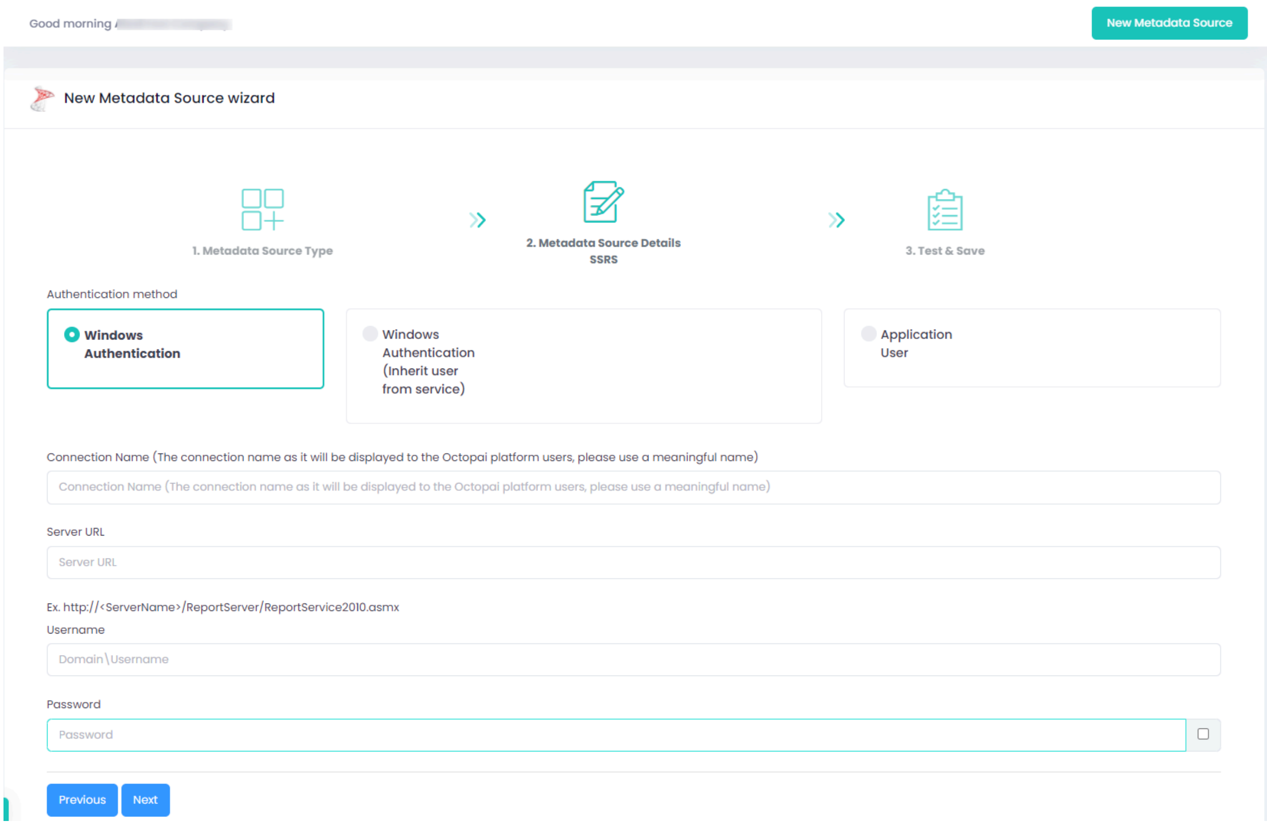
Warning: Missing permissions could end up in broken lineages.

User and password with permissions to the SSRS API (Azure Active Directory authentication not supported)

- The user needs to have “Content Manager” permissions on the report’s folder or the report itself.
    - http://<ServerName>/Reportserver/ReportService2010.asmx * Native Mode Min Permission: Content Manager
    - http://<sp server name>/_vti_bin/ReportServer/ReportService2010.asmx * Sharepoint Integrated Mode Min Permission: Owners

Setting up SSRS Metadata Source

Metadata Sources are set on the Cloudera Octopai Client



Good morning [redacted]

New Metadata Source wizard

Authentication method

● Windows Authentication ○ Windows Authentication (Inherit user from service) ○ Application User

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Server URL

Server URL

Ex. http://<ServerName>/ReportServer/ReportService2010.asmx

Username

Domain\Username

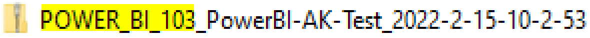
Password

Password □

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:



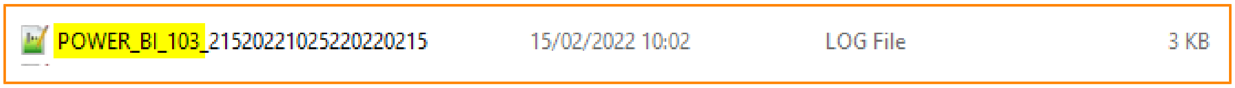
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service\log



POWER_BI_103_2152022102522022015 15/02/2022 10:02 LOG File 3 KB

Tableau Server

Learn how to configure Tableau Server for metadata extraction.

Note: Version supported: up to 2023

Tool Permissions Prerequisites

Warning: Missing permissions could end up in broken lineages.

Tableau Server/Online – Supported (Tableau Desktop - Not Supported by automation)

- Tableau server from version 10.3 and up
- API of Tableau Server is Enabled (by Default)
- Login permissions to the Tableau server using one of the following options:
  - Tableau Personal Access Token (PAT) associated with a user that has Server/Site Admin permissions to the Tableau API & the relevant Sites (Recommended). More about PAT can be found here: Personal Access Tokens
  - Tableau User and password with Server/Site Admin permissions to the Tableau API & the Relevant Sites.

The error message displayed when API is disabled:

```
System.Net.Http.HttpRequestException: Method: POST, RequestUri: 'http://<Ser
ver>/api/<api_version>/auth/signin', Version: 1.1, Content…
```

Check and run Tableau API:

Tableau server version 10.3 or lower:

- Check if the default Rest API of Tableau is enabled --> tabadmin get api.server.enabled
- If it isn't enabled (false): tabadmin set api.server.enabled

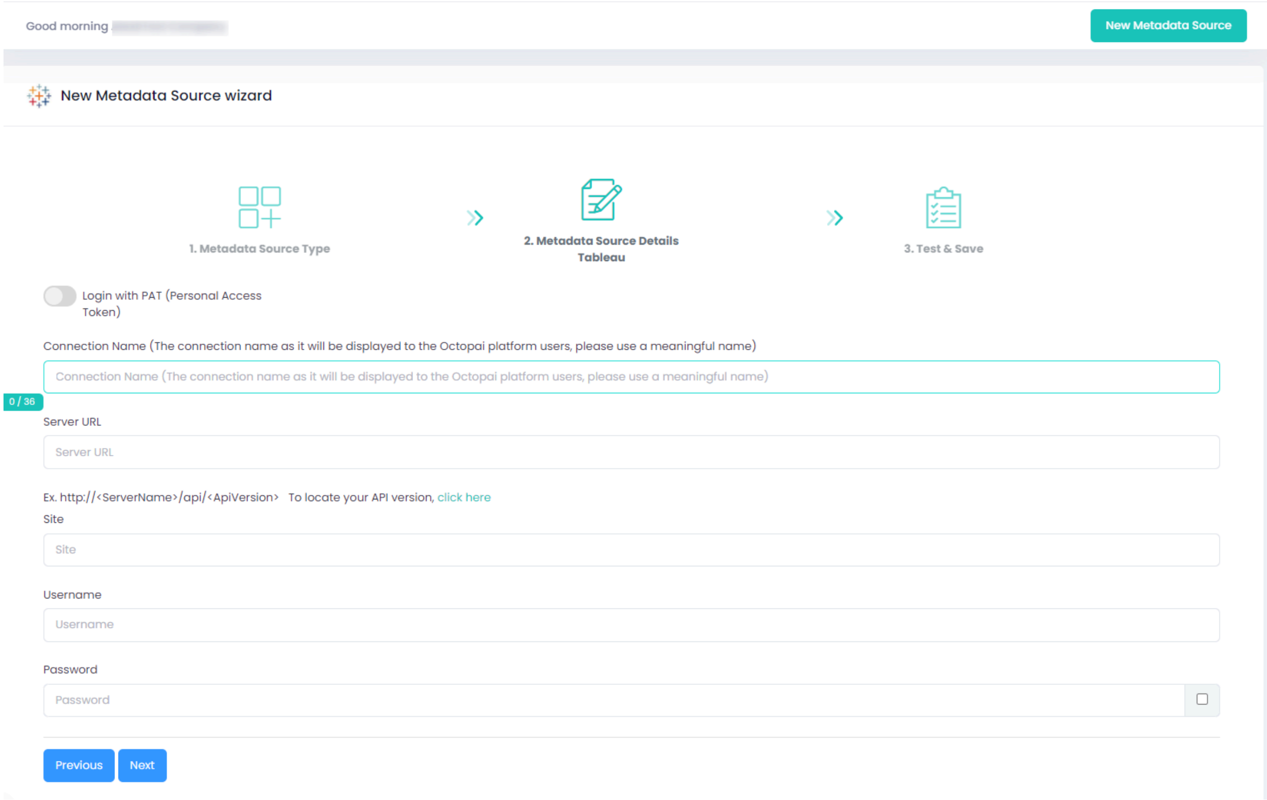
- For this, one of the following pre-requisites is required:
  - tabadmin stop
  - tabadmin start
  - tabadmin config

Tableau server versions later than 10.3:

- Check if the default Rest API of Tableau is enabled: tsm configuration get --key api.server.enabled
- If it isn’t enabled (false): tsm configuration set --key api.server.enabled true
- For this, one of the following pre-requisites is required:
  - tsm stop
  - tsm start

Setting up Tableau Metadata Source

Metadata Sources are set on the Cloudera Octopai Client



Good morning [redacted]

New Metadata Source wizard

1. Metadata Source Type
2. Metadata Source Details Tableau
3. Test & Save

Login with PAT (Personal Access Token)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Server URL

Server URL

Ex. http://<ServerName>/api/<ApiVersion> To locate your API version, click here

Site

Site

Username

Username

Password

Password

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed. By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector Name. Example:

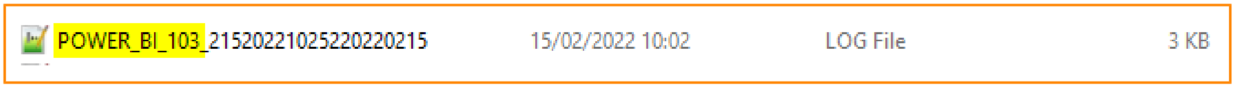
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its content: Quantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service \log



```
POWER_BI_103_21520221025220220215    15/02/2022 10:02    LOG File    3 KB
```

QlikView

How to configure the Cloudera Octopai QlikView connector.

Before you begin

- **Note:** Supported version: 2022
- Read Permission for Cloudera Octopai Windows NT User to the QlikView Log files folder.

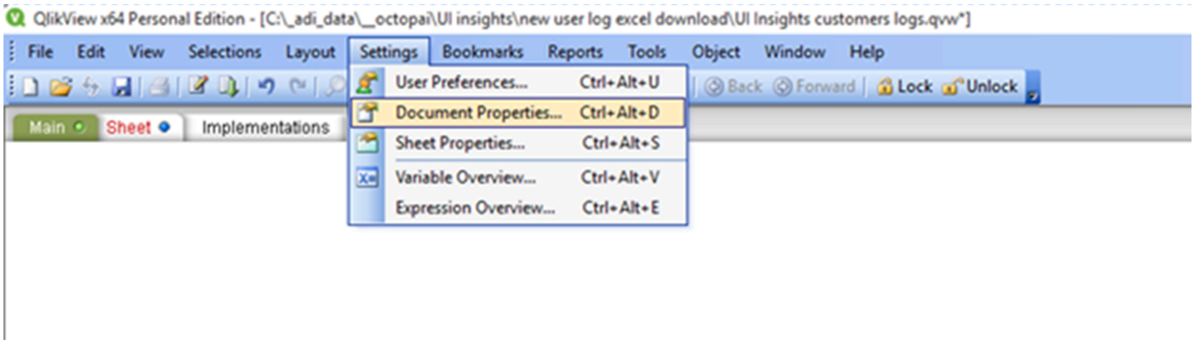
  Warning: Missing permissions could end up in broken lineages.

Procedure

1. Generate QlikView log files

a) Open the QlikView Model that needs to be uploaded to Cloudera Octopai.

b) Choose "Settings" > "Document Properties".

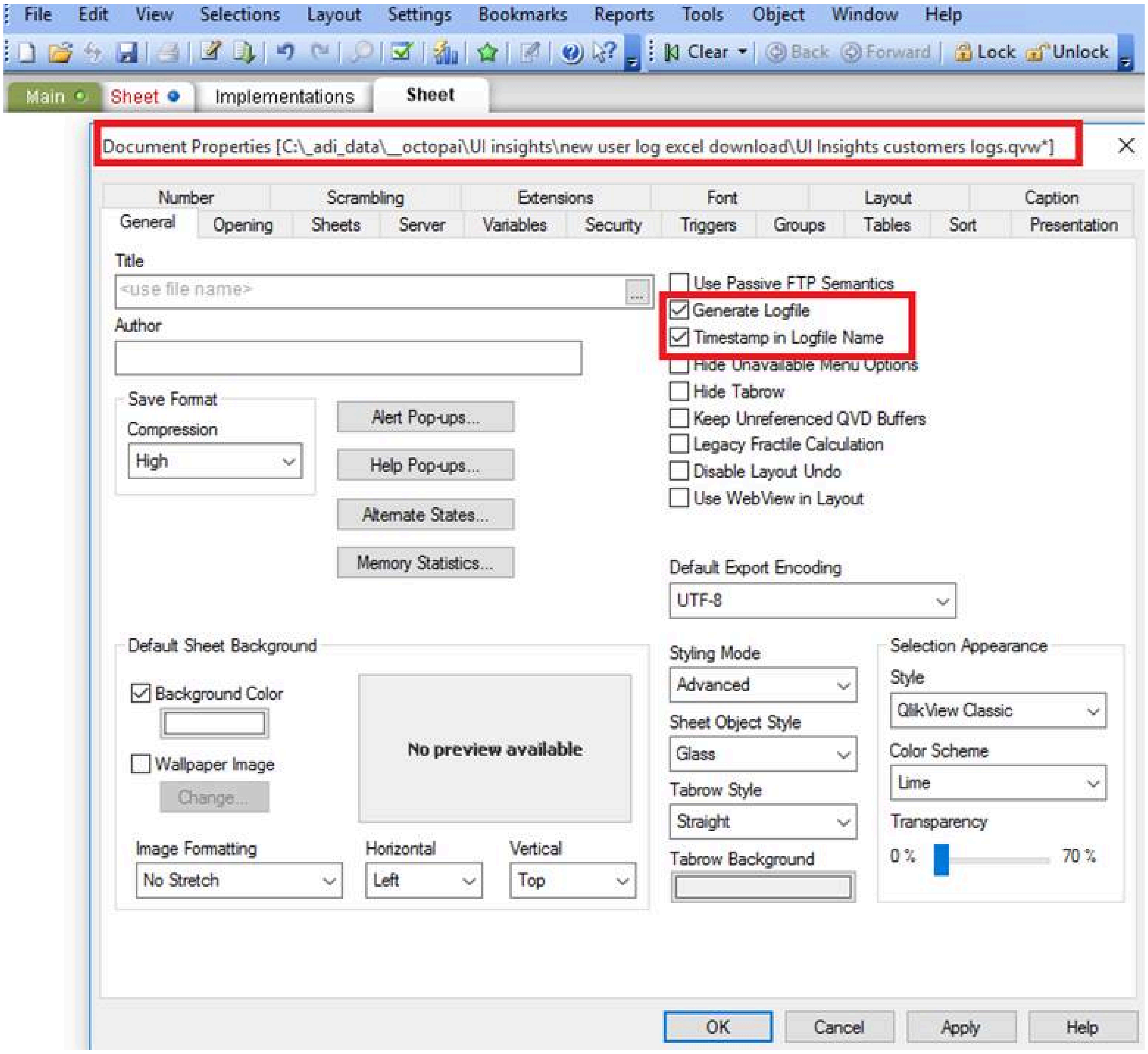


QlikView x64 Personal Edition - [C:_adi_data__octopai\UI insights\new user log excel download\UI Insights customers logs.qvw*]

File
Edit
View
Selections
Layout
Settings
Bookmarks
Reports
Tools
Object
Window
Help
Back
Forward
Lock
Unlock
Main
Sheet
Implementations
User Preferences... Ctrl+Alt+U
Document Properties... Ctrl+Alt+D
Sheet Properties... Ctrl+Alt+S
Variable Overview... Ctrl+Alt+V
Expression Overview... Ctrl+Alt+E

c) Check the boxes "Generate Logfile" and "Timestamp in Logfile Name".

d) Copy and save the Header – This is the location of the log files that will be used for the Cloudera Octopai Application.



Document Properties [C:_adi_data__octopai\UI insights\new user log excel download\UI Insights customers logs.qvw*]

Title: <use file name>

Author:

Save Format – Compression: High

☐ Use Passive FTP Semantics
☑ Generate Logfile
☑ Timestamp in Logfile Name
☐ Hide Unavailable Menu Options
☐ Hide Tabrow
☐ Keep Unreferenced QVD Buffers
☐ Legacy Fractile Calculation
☐ Disable Layout Undo
☐ Use WebView in Layout

Default Export Encoding: UTF-8

Default Sheet Background: ☑ Background Color ☐ Wallpaper Image

Image Formatting: No Stretch; Horizontal: Left; Vertical: Top

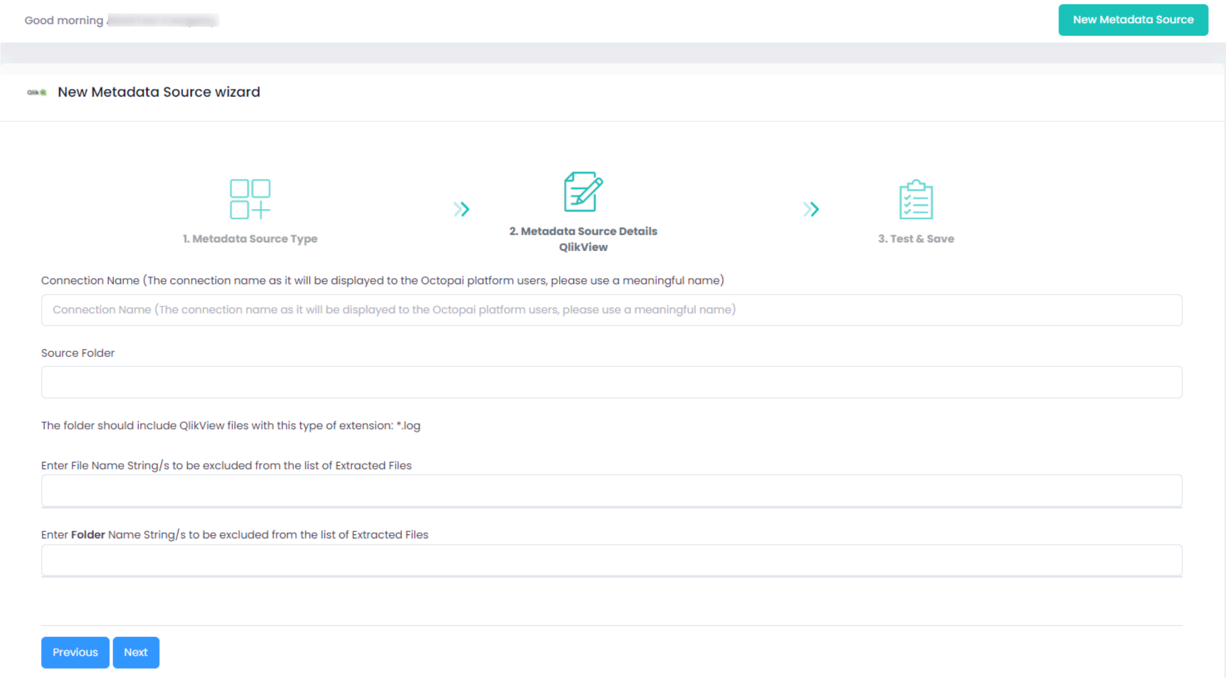
Styling Mode: Advanced; Sheet Object Style: Glass; Tabrow Style: Straight

Selection Appearance – Style: QlikView Classic; Color Scheme: Lime; Transparency: 0 % – 70 %

e) Click on Apply and OK.

f) Use the generated Path (header) for the log file location in the Cloudera Octopai client.

2. Set up QlikView Metadata Source



Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Source Folder

The folder should include QlikView files with this type of extension: *.log

Enter File Name String/s to be excluded from the list of Extracted Files

Enter **Folder** Name String/s to be excluded from the list of Extracted Files

If an error occurred during the setup, perform the following troubleshooting steps:

1. Check the permissions.
2. Send the log with the connector number and name to Cloudera Octopai Support.

Qlik Sense

Configure the Cloudera Octopai Qlik Sense connector, including metadata setup, user allocation, and required permissions.

Note: Version supported: 2023

Tool Permissions Prerequisites

Warning: Missing permissions could end up in broken lineages.

- Login permissions from the server running the Cloudera Octopai Client to the Qlik Sense Server (User and password with permissions to the Qlik Sense REST API, which needs to be enabled, and permissions to all apps you wish to extract).
- Read Permission for Cloudera Octopai Windows NT User on Qlik Sense Log files folder (Default location for log files:C:\ProgramData\Qlik\Sense\Log\Script).

- Allocated licensed user with permissions to streams you would like analyzed by Cloudera Octopai, use the following steps if needed:
  - Create Custom property Creating a custom property # Qlik Sense for administrators
    - Name ‘Octopai_Group'
    - Resource type - ‘Users’, ‘Streams’
    - Value 'Octopai’
  - Assign Property ‘Octopai_ Group' and value 'Octopai’ to each stream you would like Cloudera Octopai to analyze.
  - Assign Property ‘Octopai_ Group' and value 'Octopai’ to the user that will be used for extraction
  - Add Security rule Creating security rules # Qlik Sense for administrators
    - Name - Octopai_Stream_Rule
    - Resource filter - Stream_*
    - Action - Read
    - Condition - ((user.@Octopai_Group=resource.@Octopai_Group))
  - Proxy port values most contain default values Editing proxies # Qlik Sense for administrators
    - Service listen port HTTPS (default) - 443
    - Authentication listen port - 4244
    - REST API listen port - 4243

How to allocate a user in Qlik Sense

Note: The allocated user is necessary for a successful extraction from Qlik Sense.

Prerequisites:

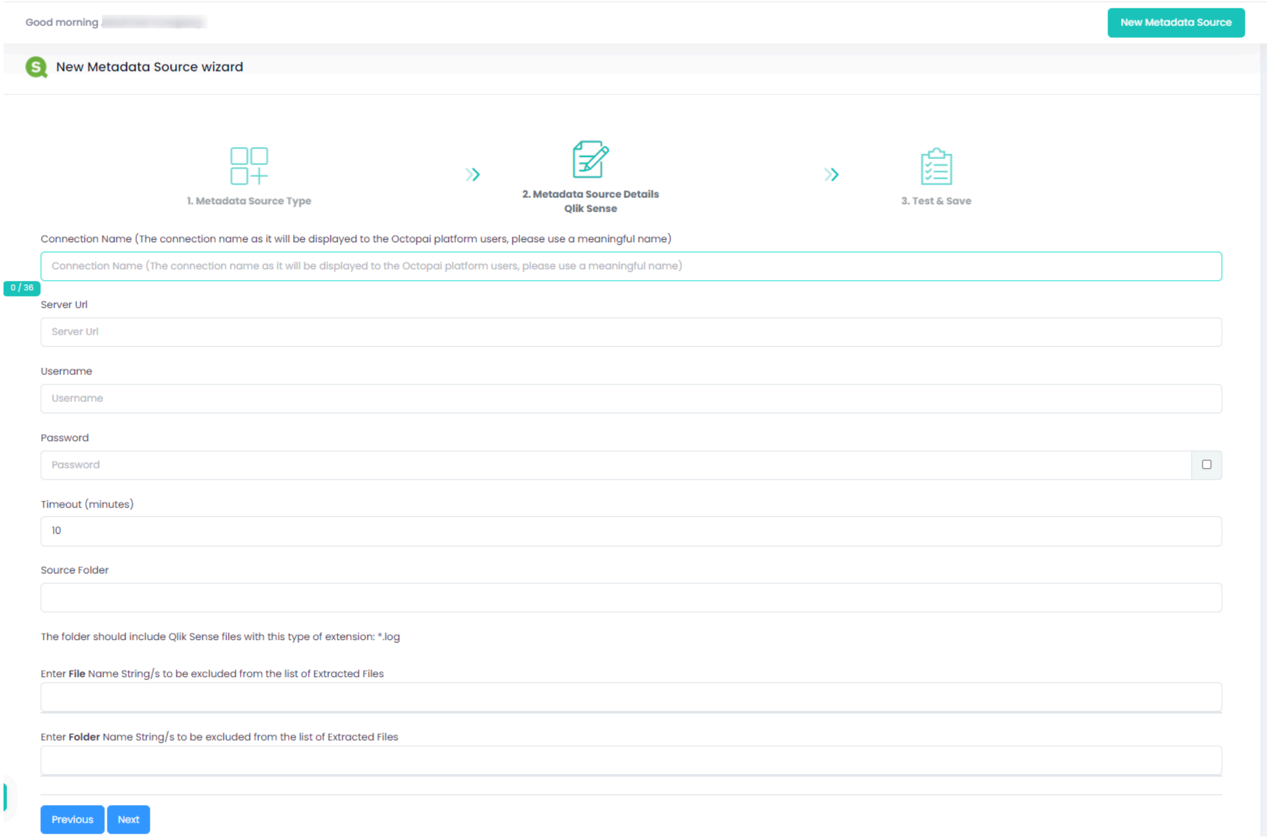
- Login permissions from the server running the Octopai Client to the Qlik Sense Server (User and password with permissions to the Qlik Sense REST API, which needs to be enabled and permissions to all apps you would like to extract).
- Read Permission for Cloudera Octopai Windows NT User on Qlik Sense Log files folder.
- Default location for log files: C:\ProgramData\Qlik\Sense\Log\Script.
- Dedicated Header Authentication: Configuring header authentication.

Grant permission to all relevant streams to an allocated user:

- Create Custom property: Creating a custom property.
  - Name ‘Octopai_Group'
  - Resource type - ‘Users’, ‘Streams’
  - Value 'Octopai’
- Assign Property ‘Octopai_ Group' and value 'Octopai’ to each stream you would like Cloudera Octopai to analyze.
- Assign Property ‘Octopai_ Group' and value 'Octopai’ to the user that will be used for extraction
- Add Security rule: Creating security rules.
- Name - Octopai_Stream_Rule
- Action - Read
- Condition - User@Octopai_Group = #Stream@Octopai_Group

Setting up Qlik Sense Metadata Source

Metadata Sources are set on the Cloudera Octopai Client



Good morning [redacted]

New Metadata Source

New Metadata Source wizard

1. Metadata Source Type

2. Metadata Source Details
Qlik Sense

3. Test & Save

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Server Url

Server Url

Username

Username

Password

Password □

Timeout (minutes)

10

Source Folder

The folder should include Qlik Sense files with this type of extension: *.log

Enter **File** Name String/s to be excluded from the list of Extracted Files

Enter **Folder** Name String/s to be excluded from the list of Extracted Files

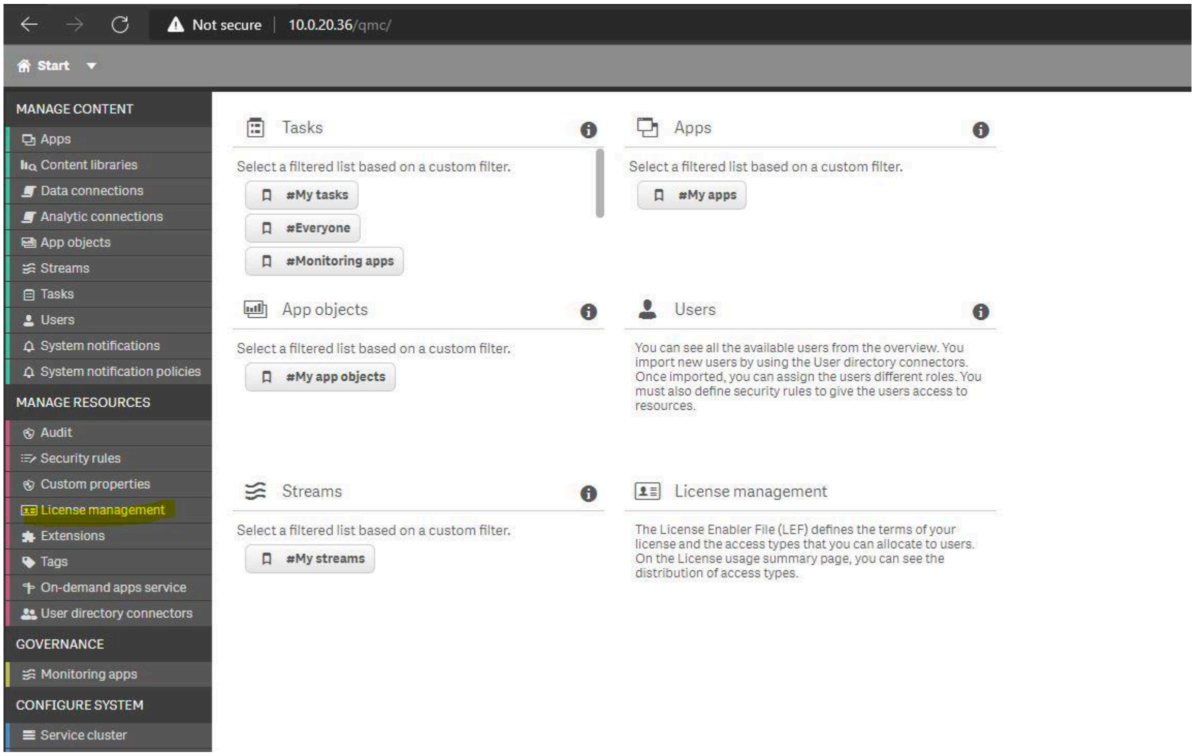
Previous Next

Configure Qlik Sense for Data Extraction

Learn how to configure Qlik Sense for data extraction with Cloudera Octopai Data Lineage.

Procedure

1. Assign a license to the Qlik user used for the data extraction.
   a) Open the Qlik Management Console at https://qlik-server-address:/qmc.
   b) Navigate to the License management menu.



Select a filtered list based on a custom filter.

Select a filtered list based on a custom filter.

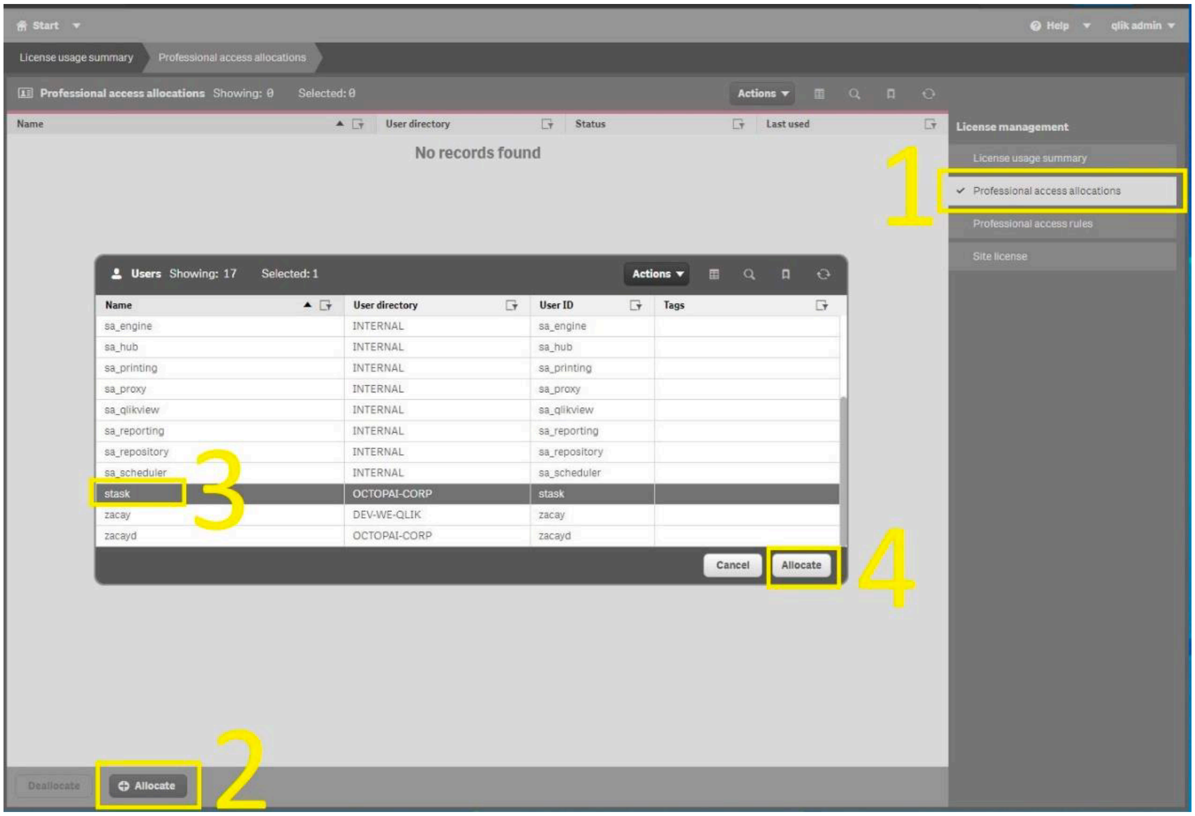
Select a filtered list based on a custom filter.

You can see all the available users from the overview. You import new users by using the User directory connectors. Once imported, you can assign the users different roles. You must also define security rules to give the users access to resources.

Select a filtered list based on a custom filter.

The License Enabler File (LEF) defines the terms of your license and the access types that you can allocate to users. On the License usage summary page, you can see the distribution of access types.

   c) Click Professional license allocation, select the target user, and click Allocate.



Professional access allocations Showing: 0 Selected: 0

Name	User directory	Status	Last used

No records found

License management

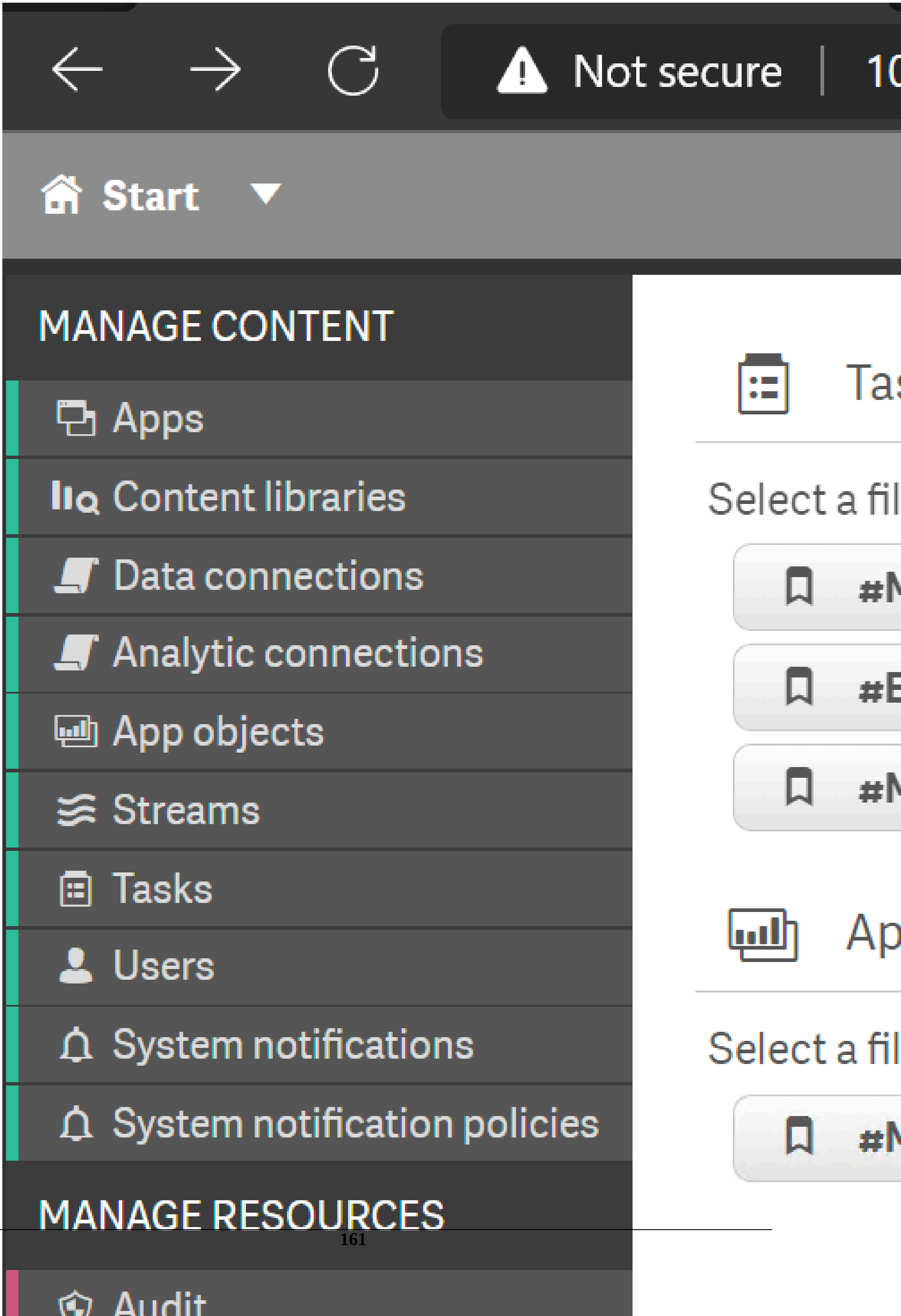
- License usage summary
- ✓ Professional access allocations
- Professional access rules
- Site license

Users Showing: 17 Selected: 1

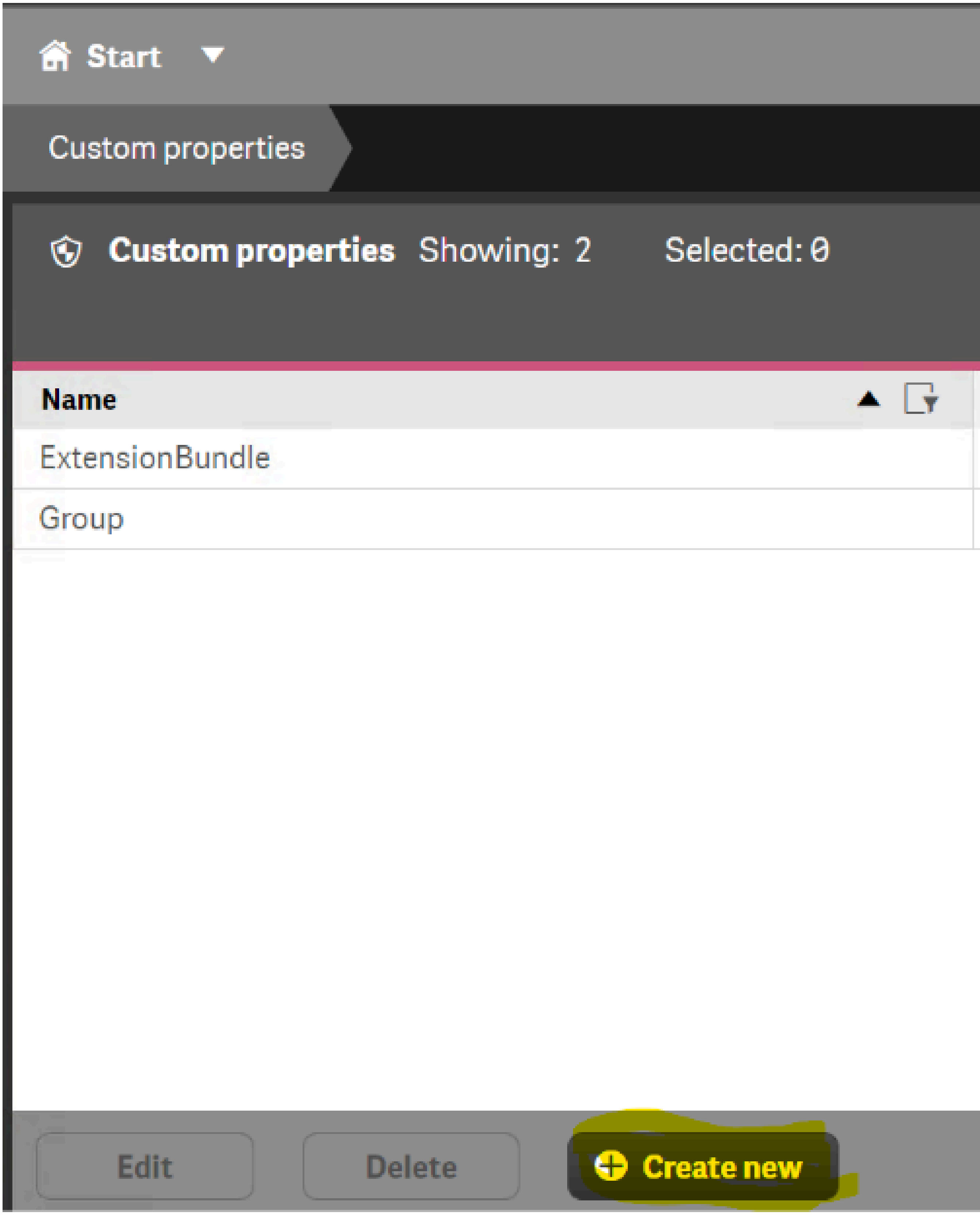
Name	User directory	User ID	Tags
sa_engine	INTERNAL	sa_engine	
sa_hub	INTERNAL	sa_hub	
sa_printing	INTERNAL	sa_printing	
sa_proxy	INTERNAL	sa_proxy	
sa_qlikview	INTERNAL	sa_qlikview	
sa_reporting	INTERNAL	sa_reporting	
sa_repository	INTERNAL	sa_repository	
sa_scheduler	INTERNAL	sa_scheduler	
stask	OCTOPAI-CORP	stask	
zacay	DEV-WE-QLIK	zacay	
zacayd	OCTOPAI-CORP	zacayd	

2. Create a custom property to tag specific users and streams.

a) Go to the Custom properties menu.


Not secure
10
Start
MANAGE CONTENT
Apps
Content libraries
Data connections
Analytic connections
App objects
Streams
Tasks
Users
System notifications
System notification policies
MANAGE RESOURCES
Audit
Ta
Select a fil
#M
#E
#M
Ap
Select a fil
#M

b) Click Create new and add the following properties:



Name

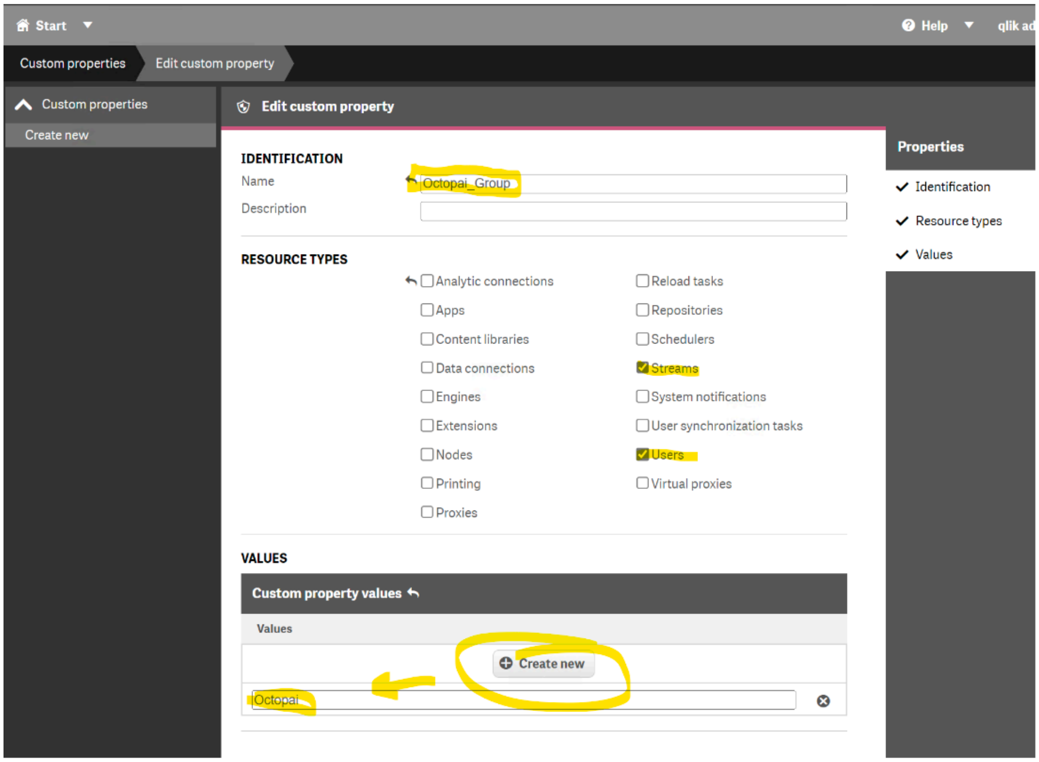
Octopai_Group

Resource types

Streams, Users

Value

Octopai



IDENTIFICATION

Name: Octopai_Group

Description:

RESOURCE TYPES

□ Analytic connections	□ Reload tasks
□ Apps	□ Repositories
□ Content libraries	□ Schedulers
□ Data connections	☑ Streams
□ Engines	□ System notifications
□ Extensions	□ User synchronization tasks
□ Nodes	☑ Users
□ Printing	□ Virtual proxies
□ Proxies	

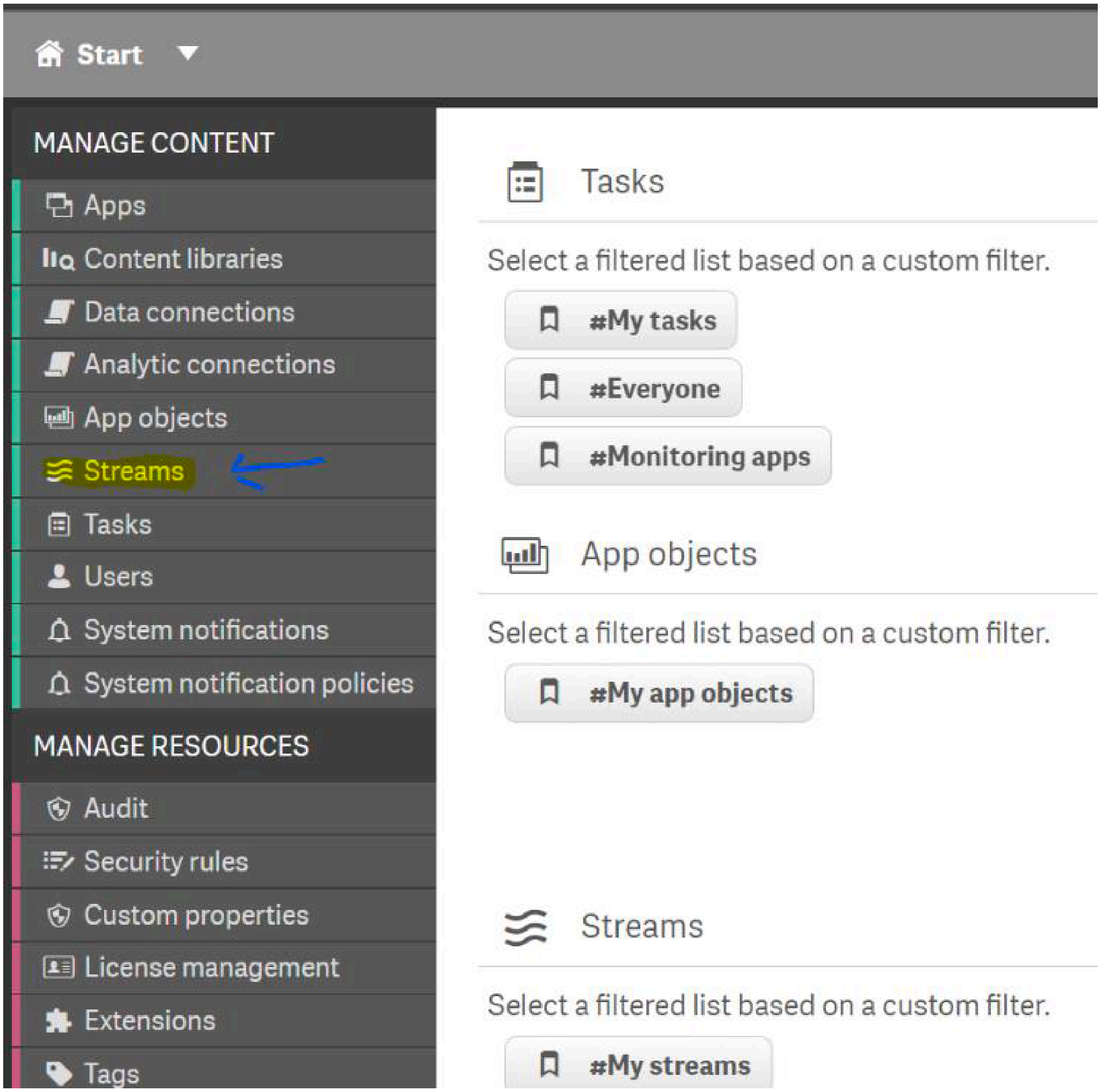
VALUES

Custom property values

Values

Octopai

3. Assign the custom property to the targeted streams.
   a) Go to the Streams menu.

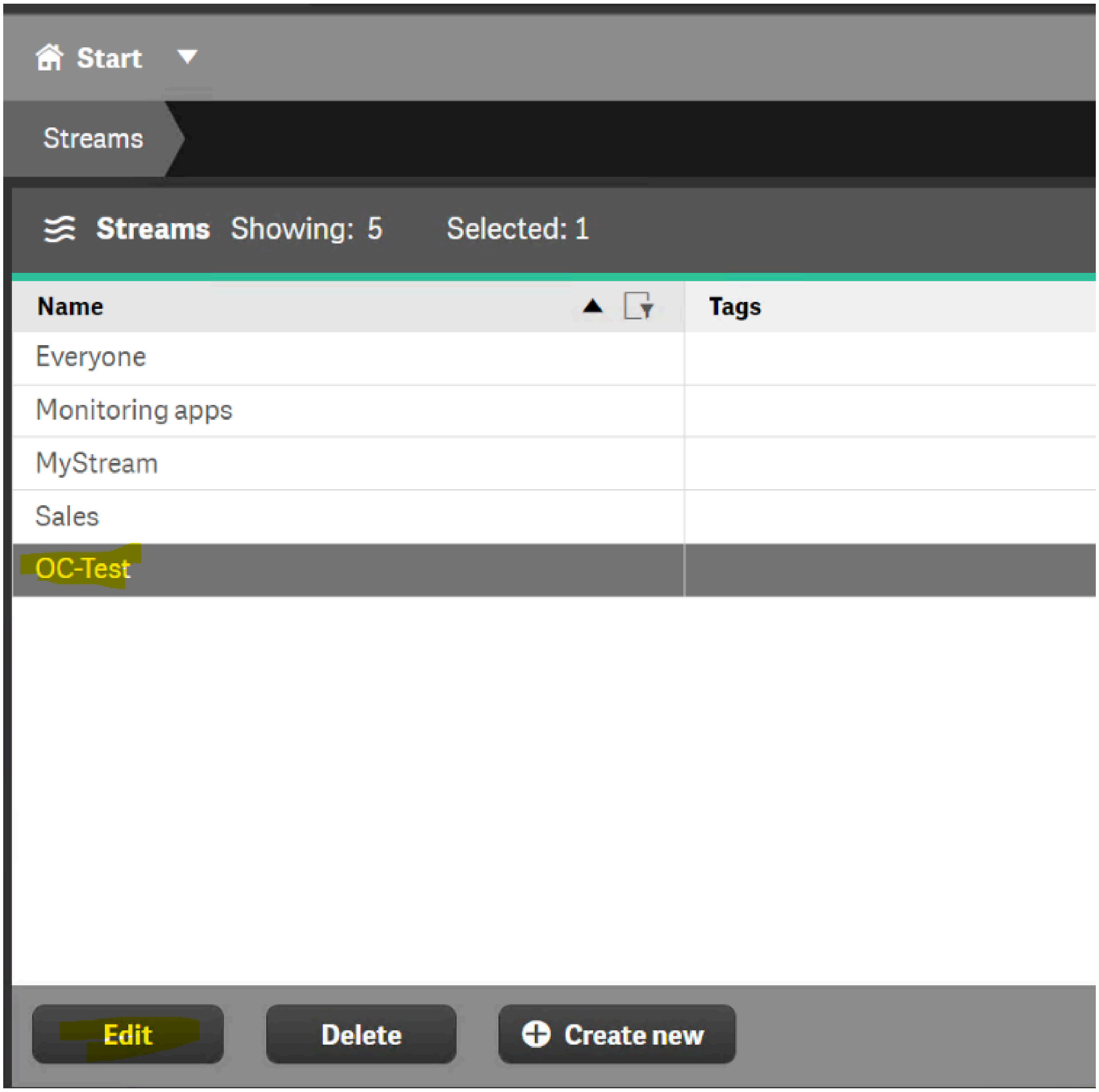


Select a filtered list based on a custom filter.

Select a filtered list based on a custom filter.

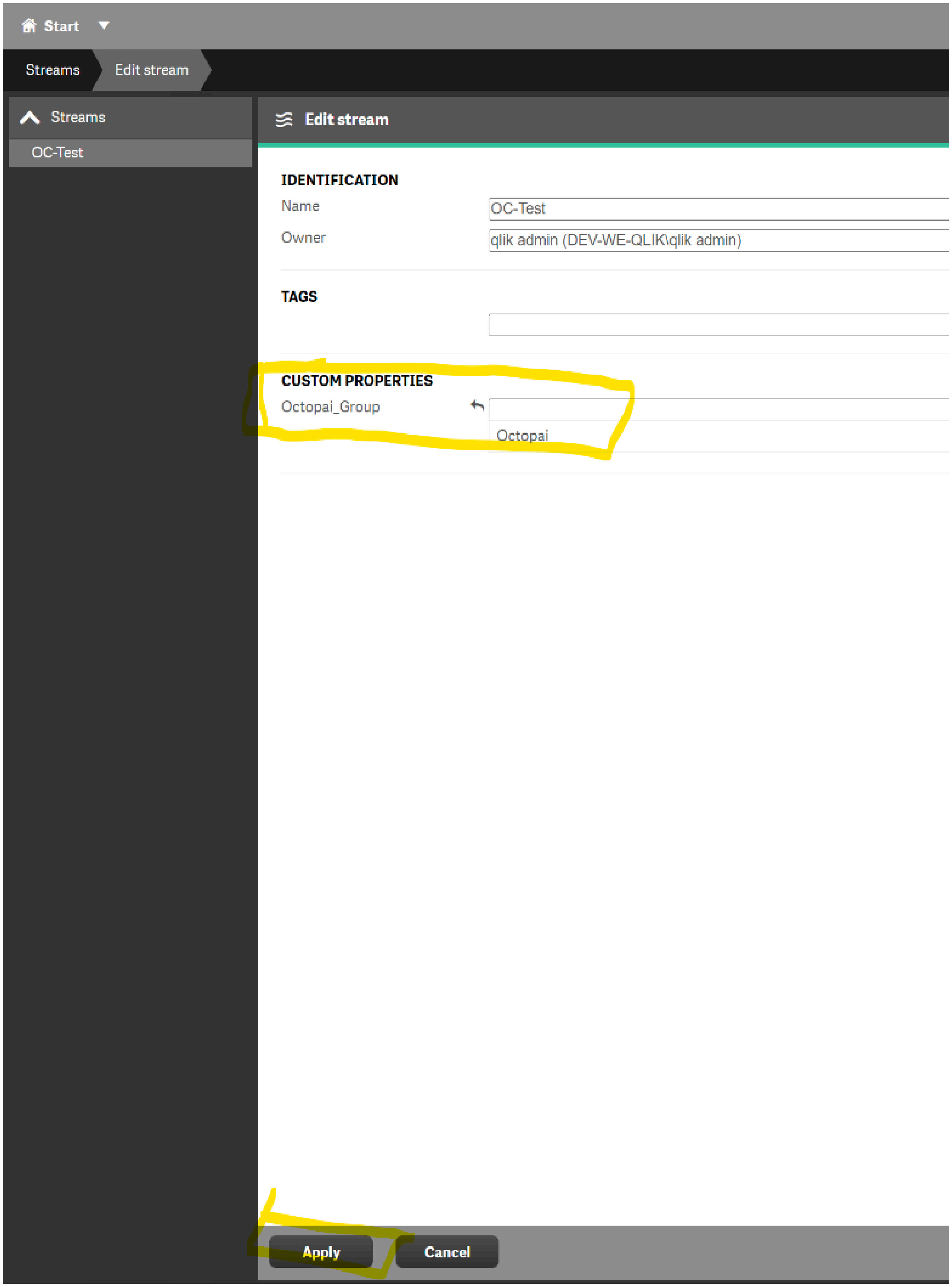
Select a filtered list based on a custom filter.

   b) Select the desired stream and click Edit.



Name	Tags
Everyone	
Monitoring apps	
MyStream	
Sales	
OC-Test	

c) Under Custom properties, find Octopai_Group and select Octopai from the dropdown menu.



IDENTIFICATION

Name: OC-Test

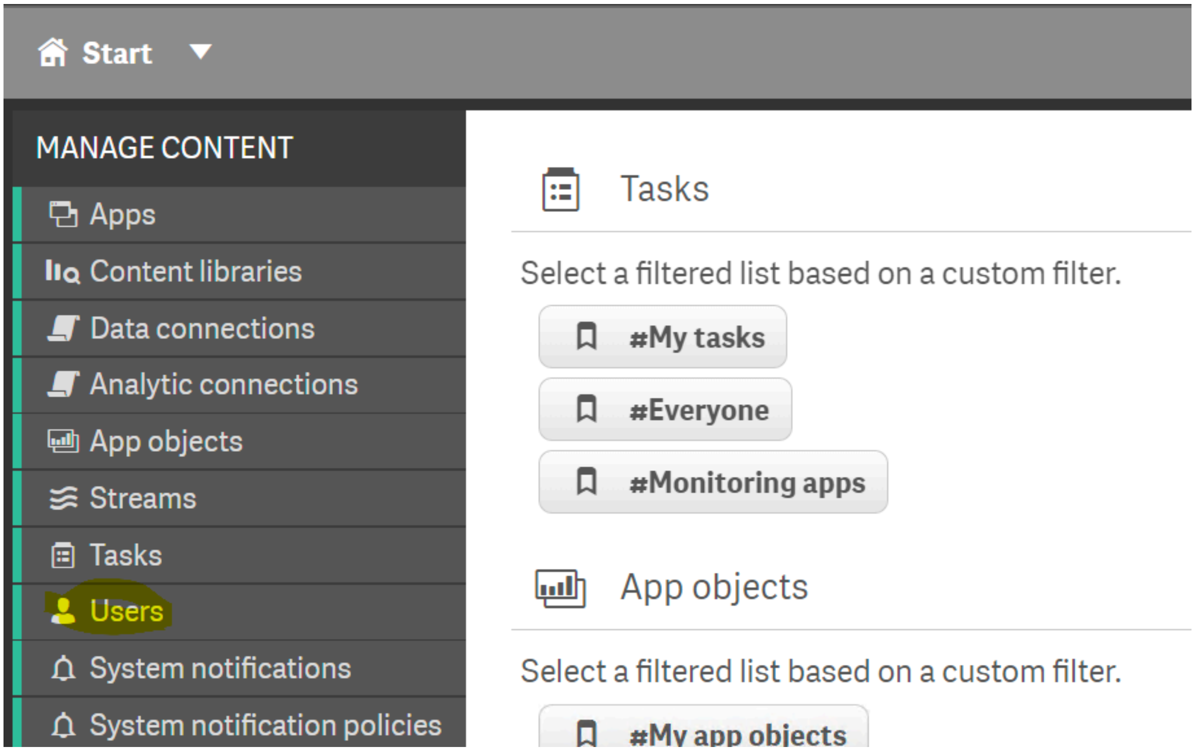
Owner: qlik admin (DEV-WE-QLIK\qlik admin)

TAGS

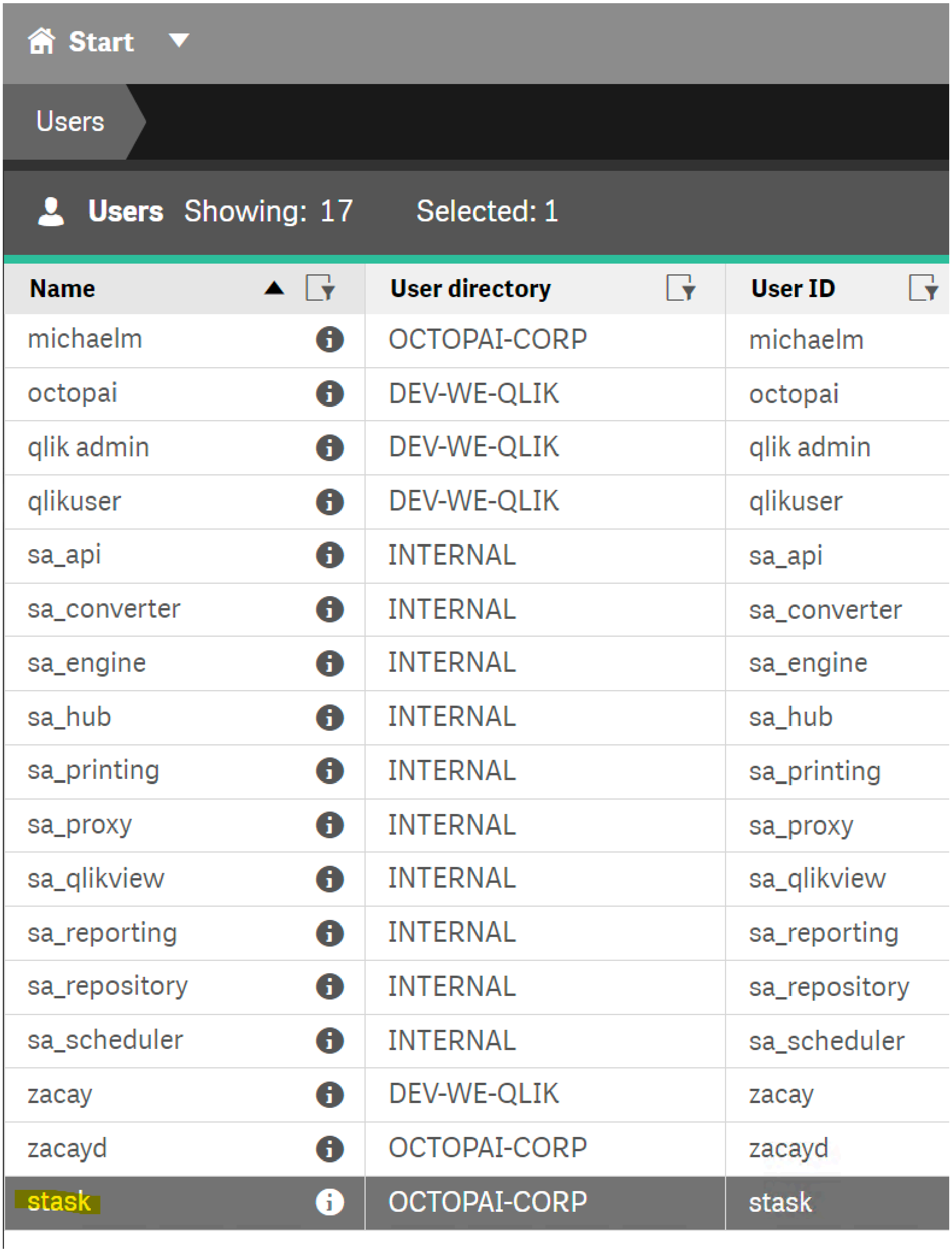
CUSTOM PROPERTIES

Octopai_Group: Octopai

4. Assign the custom property to the user used for extraction.
   a) Go to the Users menu.

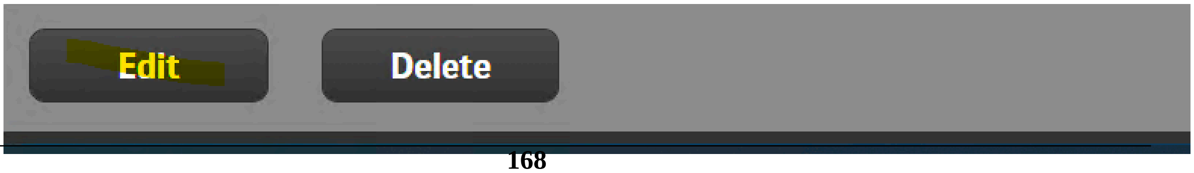


   b) Select the intended user and click Edit.



Users Showing: 17 Selected: 1

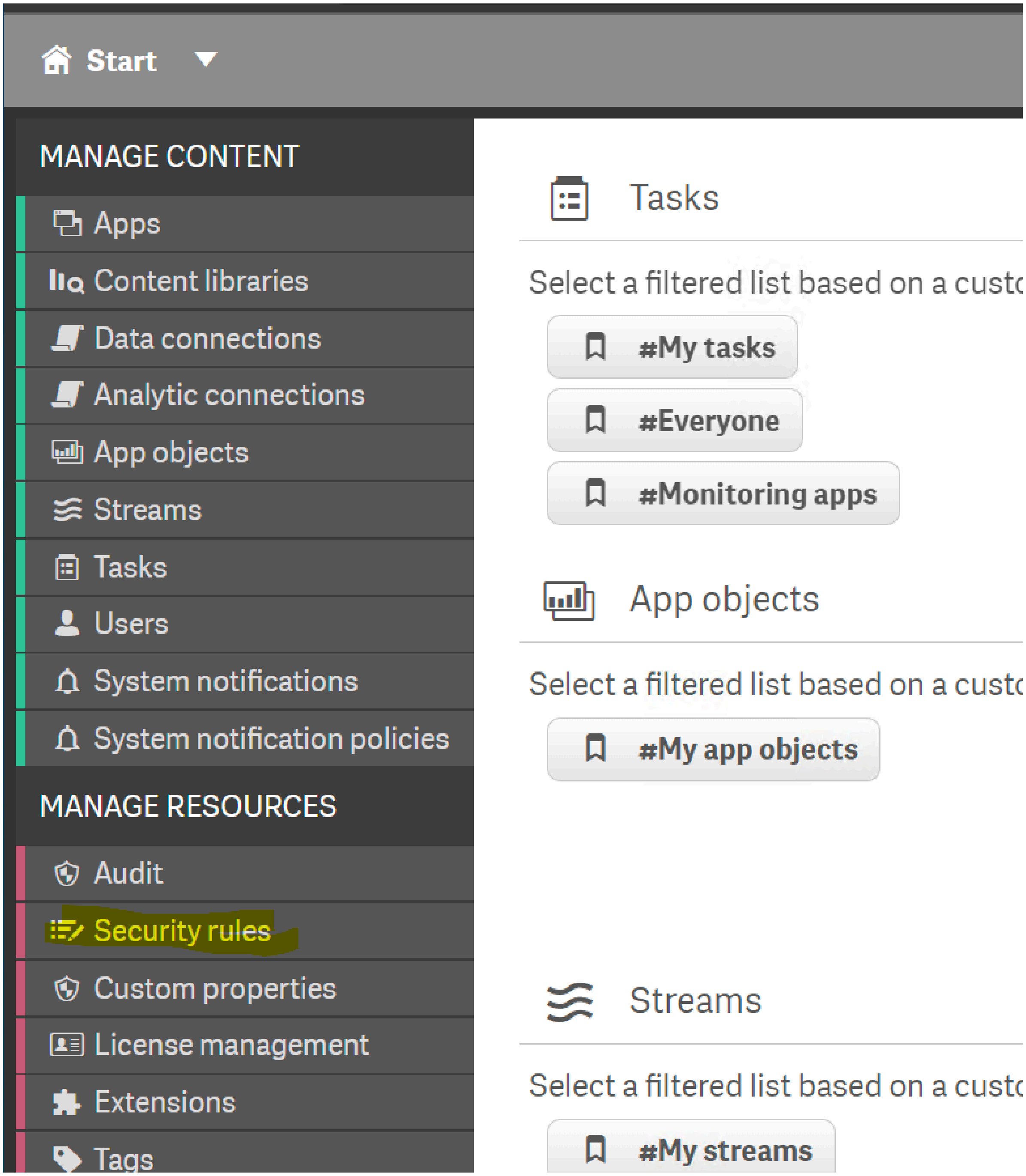
Name	User directory	User ID
michaelm	OCTOPAI-CORP	michaelm
octopai	DEV-WE-QLIK	octopai
qlik admin	DEV-WE-QLIK	qlik admin
qlikuser	DEV-WE-QLIK	qlikuser
sa_api	INTERNAL	sa_api
sa_converter	INTERNAL	sa_converter
sa_engine	INTERNAL	sa_engine
sa_hub	INTERNAL	sa_hub
sa_printing	INTERNAL	sa_printing
sa_proxy	INTERNAL	sa_proxy
sa_qlikview	INTERNAL	sa_qlikview
sa_reporting	INTERNAL	sa_reporting
sa_repository	INTERNAL	sa_repository
sa_scheduler	INTERNAL	sa_scheduler
zacay	DEV-WE-QLIK	zacay
zacayd	OCTOPAI-CORP	zacayd
stask	OCTOPAI-CORP	stask



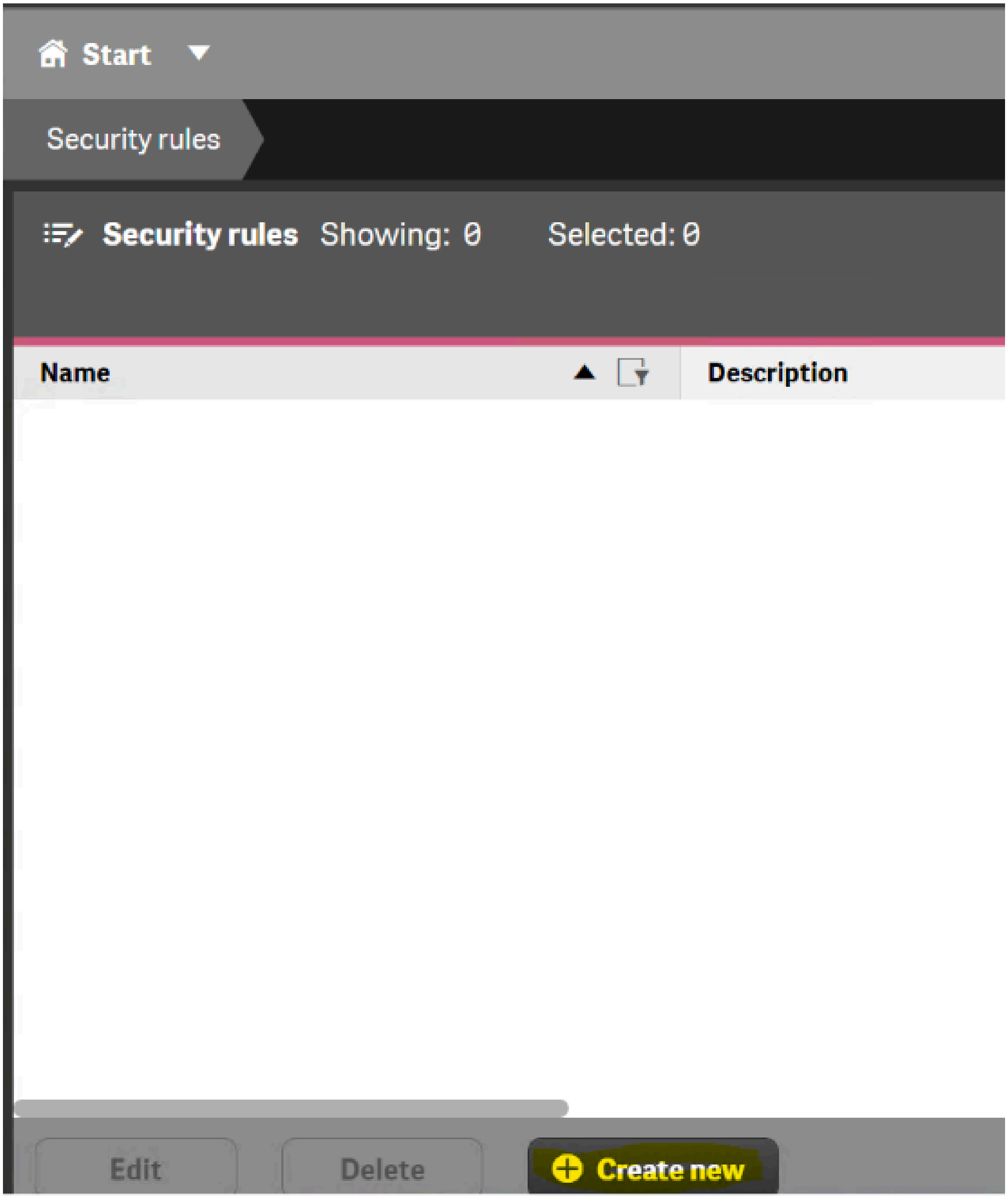
c) Under Custom properties, find Octopai_Group and select Octopai from the dropdown menu.

5. Create a security rule to connect the user and the streams.

a) Go to the Security rules menu.



b) Click Create new and use the following values:



Name

Octopai_Stream_Rule

Resource filter

Stream_*

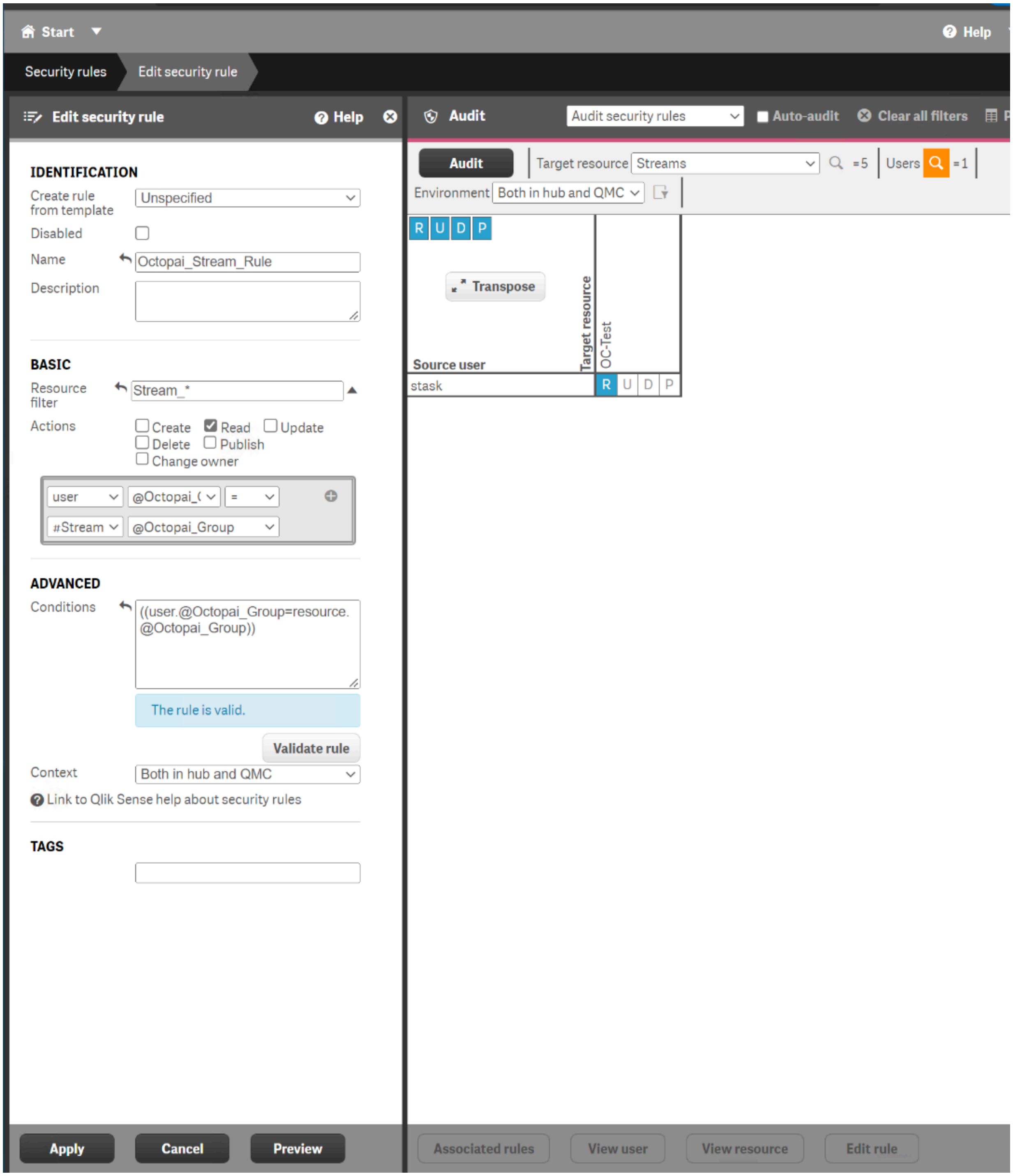
Actions

Read

Conditions

((user.@Octopai_Group=resource.@Octopai_Group))

c) (Optional) Click Validate rule to verify the configuration.



IDENTIFICATION

Create rule from template: Unspecified

Disabled: □

Name: Octopai_Stream_Rule

Description:

BASIC

Resource filter: Stream_*

Actions: □ Create ☑ Read □ Update □ Delete □ Publish □ Change owner

user @Octopai_(=

#Stream @Octopai_Group

ADVANCED

Conditions: ((user.@Octopai_Group=resource.@Octopai_Group))

The rule is valid.

Validate rule

Context: Both in hub and QMC

Link to Qlik Sense help about security rules

TAGS

Oracle OBIEE / Oracle Analytics Server (OAS)

Learn how to extract and configure Oracle OBIEE metadata files for analysis by Cloudera Octopai Data Lineage.

Before you begin

- Supported version: up to 12c / 2023

- Read the Permission for Cloudera Octopai Windows NT User to the OBIEE RDP and Catalog files folder.

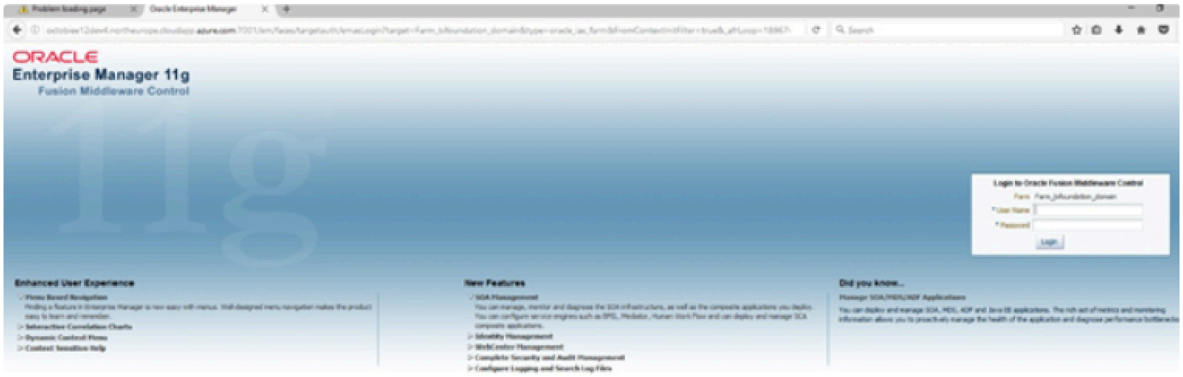
  ⚠ **Warning:** Missing permissions could result in broken lineages.

Procedure

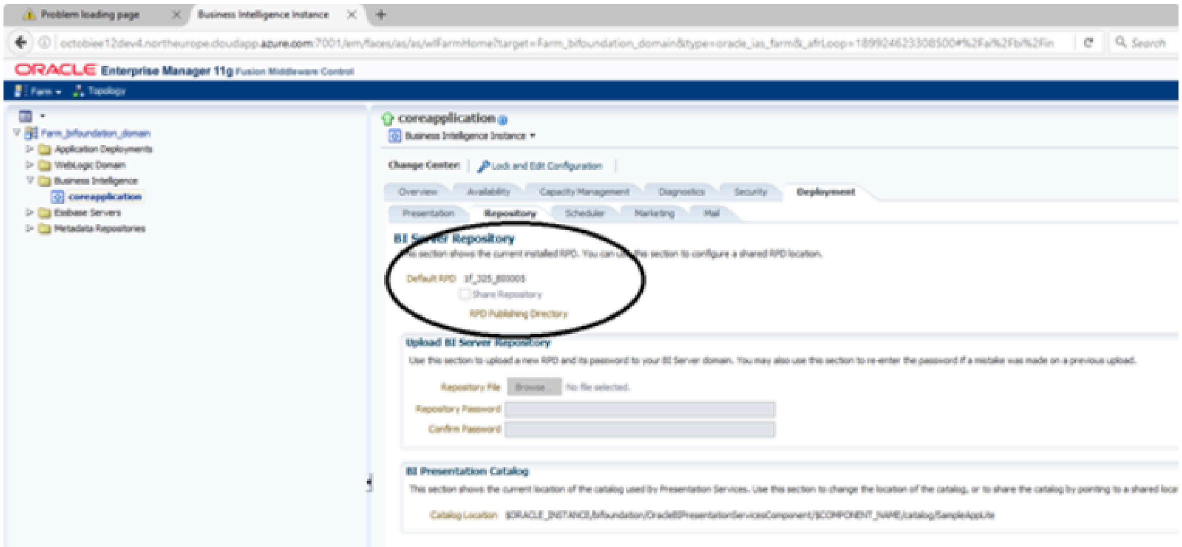
1. **Note:** The rpd and catalog files contain the necessary metadata for the analysis by Cloudera Octopai.

Retrieve the repository rpd file

a) Log in to Oracle OBIEE enterprise manager (em) at http://obiee12c:9500/em/.



b) Go to Business Intelligence coreapplication Repository .



Default RPD 1f_325_B00005

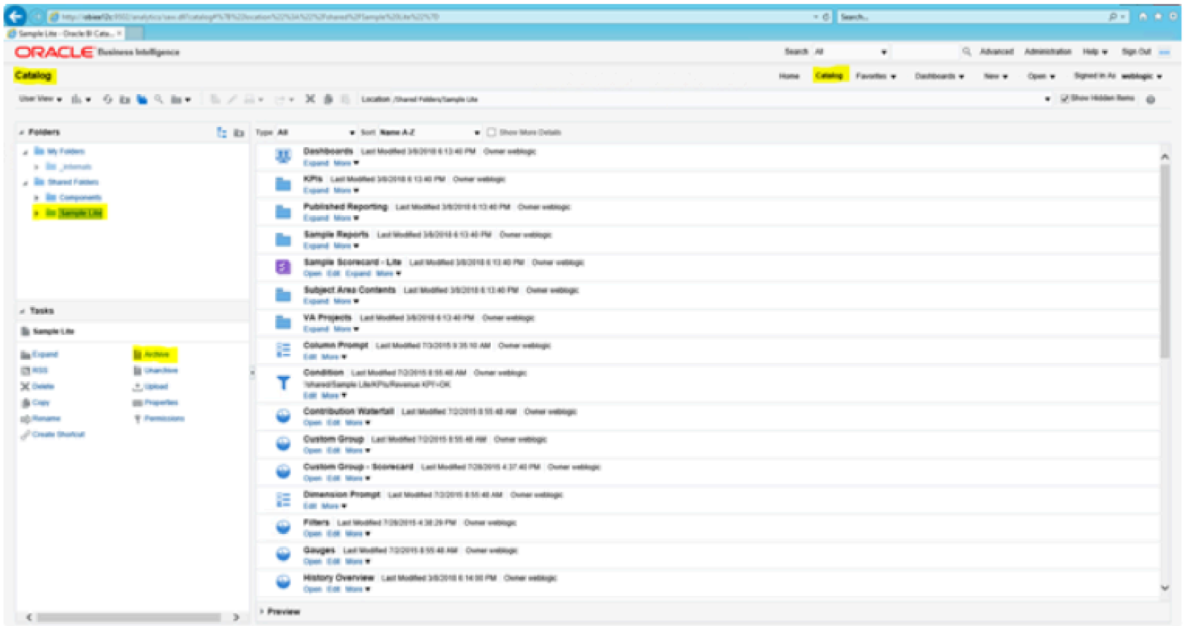
Use this section to upload a new RPD and its password to your BI Server domain. You may also use this section to re-enter the password if a mistake was made on a previous upload.

Catalog Location $ORACLE_INSTANCE/bifoundation/OracleBIPresentationServicesComponent/$COMPONENT_NAME/catalog/SampleAppLite

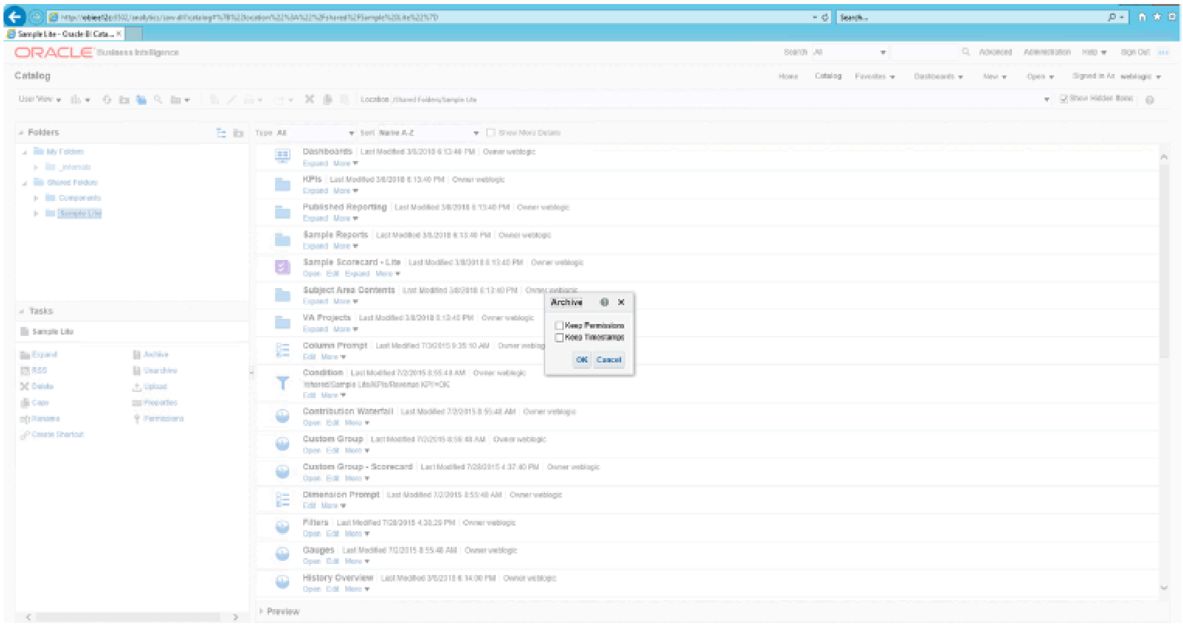
The rpd file can be found at D:\Oracle\Middleware\ user_projects\domains\bi\bidata\service_instances\ssi \metadata\datamodel\customizations.

c) Extract the OBIEE reports that is the catalog file.

1. Log in to OBIEE analytics at http://<Server Name>:9502/analytics/saw.dll?bieehome.
2. Click on Catalog at the top right corner.
3. Select the folder you want to archive, and then click on the Archive button.



d) In the task box select archive to zip the files.



e) Deselect the Keep Permissions and Keep Timestamps checkboxes.

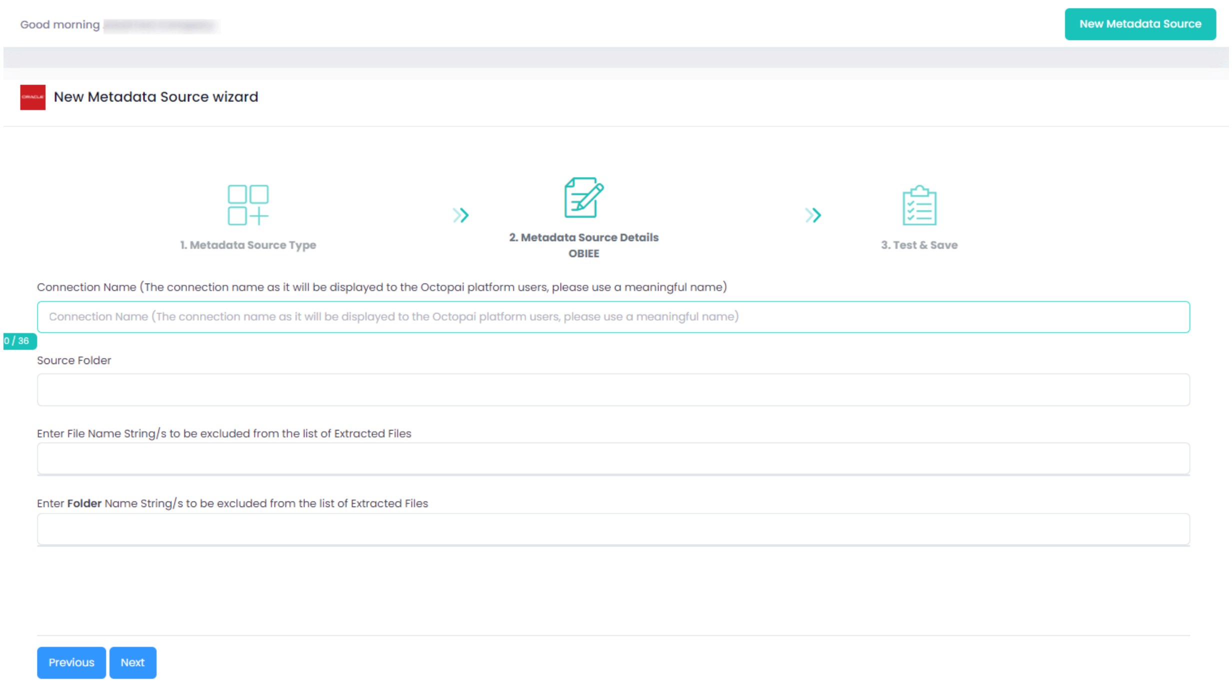
f) Click OK.

g) Save the file.

Note:

- The files will appear in the defined destination folder. Use this file as source for your Cloudera Octopai metadata OBIEE connection in the Cloudera Octopai Client.
- Ensure that you have appropriate permissions to the path so that the Cloudera Octopai Client can access the file with the user running it.

2. Set up the Oracle OBIEE metadata source on the Cloudera Octopai Client.



Good morning [redacted]

New Metadata Source wizard

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Source Folder

Enter File Name String/s to be excluded from the list of Extracted Files

Enter **Folder** Name String/s to be excluded from the list of Extracted Files

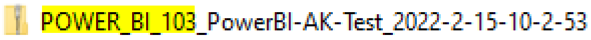
3. Verify the extracted metadata file by accessing the Cloudera Octopai Target Folder (TGT).

a) Go to the TGT Folder located on the server where the Cloudera Octopai Client is installed.

The default location is C:\Program Files (x86)\Octopai\Service\TGT.

b) Open the zip file with the connector name.

Example



POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

c) Verify the file content by checking the quantity and quality of the included files.

If an error occurred during the extraction, perform the following troubleshooting steps:

1. Check the permissions.

2. Send the log with the connector number and name to Cloudera Support.

You can find the log files at C:\Program Files (x86)\Octopai\Service\log.

POWER_BI_103_21520221025220220215 15/02/2022 10:02 LOG File 3 KB

IBM Cognos Report Studio

Learn how to export, schedule, and manage IBM Cognos reports and packages.

Note: Version supported: up to v11.2.x

Before you begin

- Read the Permission for Cloudera Octopai Windows NT User to the Cognos ZIP files folder.

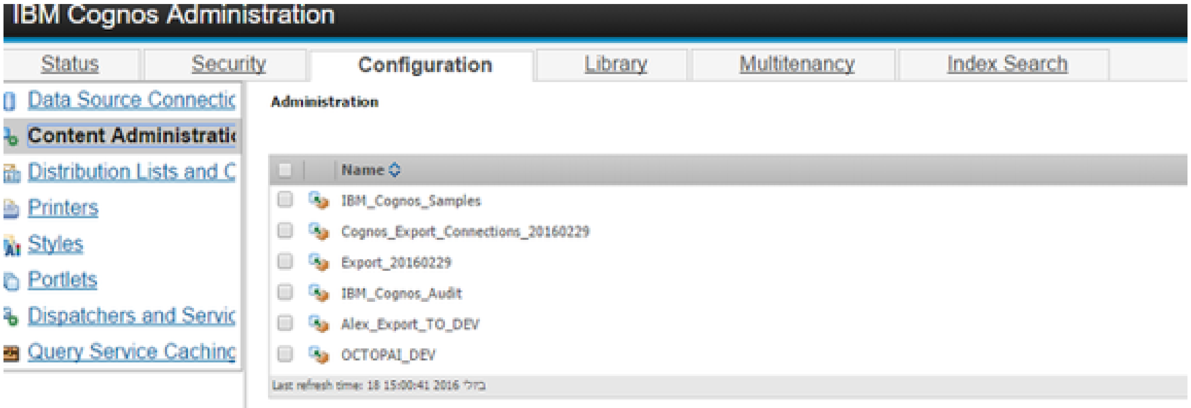
  Warning: Missing permissions could result in broken lineages.

About this task

Procedure

1. Export and Schedule Export of IBM Cognos Reports and Packages

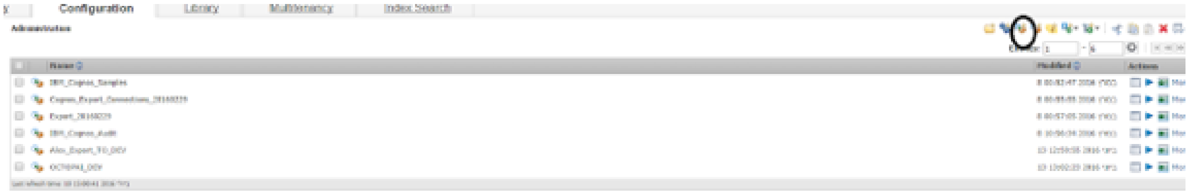
   a) Open IBM Cognos Administration and go to the Configuration tab.



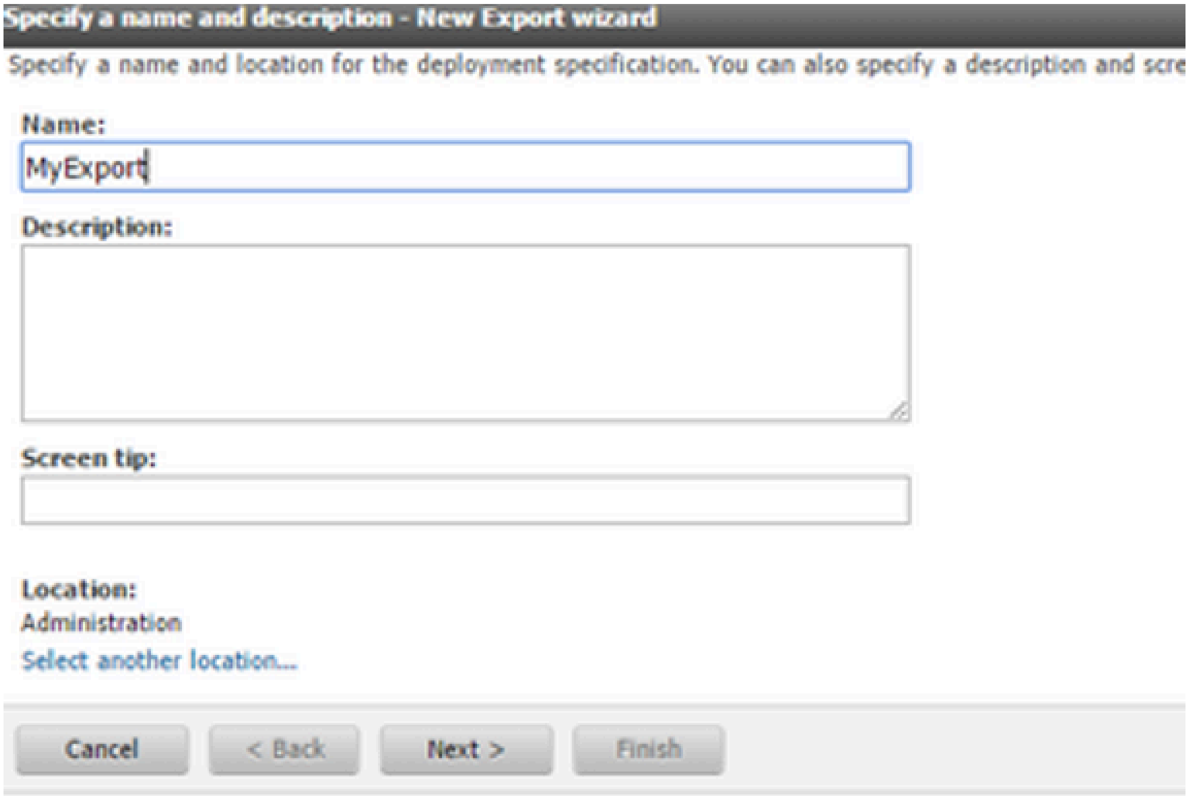
Name
IBM_Cognos_Samples
Cognos_Export_Connections_20160229
Export_20160229
IBM_Cognos_Audit
Alex_Export_TO_DEV
OCTOPAI_DEV

Last refresh time: 18 15:00:41 2016 יול

   b) Click the New export button.



   c) Name the export.


Specify a name and description - New Export wizard

Specify a name and location for the deployment specification. You can also specify a description and scre

Name: MyExport

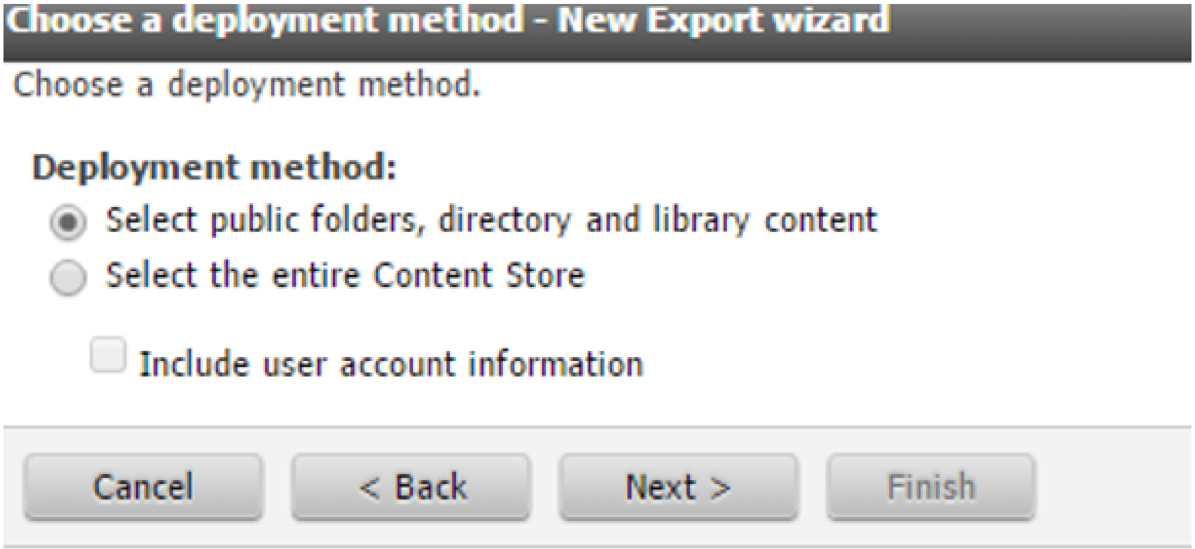
Description:

Screen tip:

Location: Administration

Select another location...
Cancel
< Back
Next >
Finish

   d) Select the deployment method by choosing the first option.



Choose a deployment method - New Export wizard

Choose a deployment method.

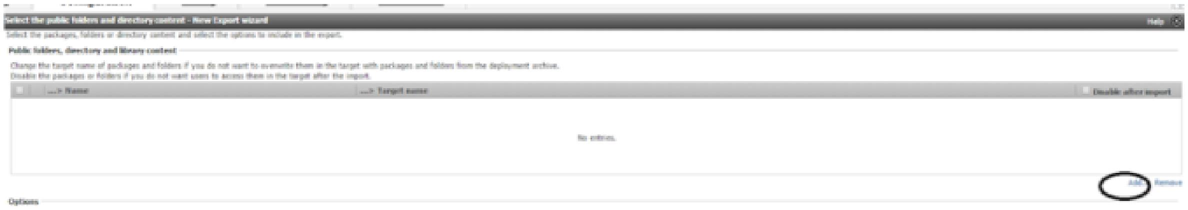
Deployment method:

● Select public folders, directory and library content

○ Select the entire Content Store

□ Include user account information

e) Click Add.



Select the public folders and directory content - New Export wizard

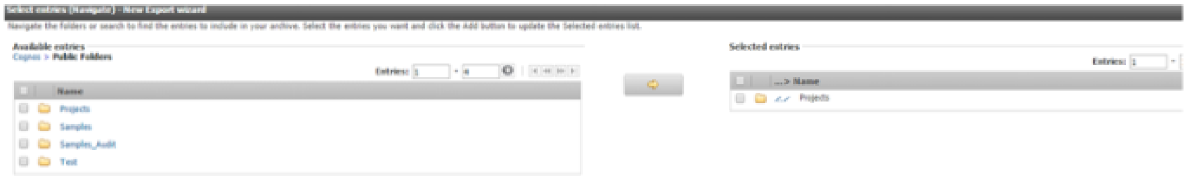
Select the packages, folders or directory content and select the options to include in the export.

Public folders, directory and library content

Change the target name of packages and folders if you do not want to overwrite them in the target with packages and folders from the deployment archive.
Disable the packages or folders if you do not want users to access them in the target after the import.

□	...> Name	...> Target name	□ Disable after import
	No entries.		

f) Add packages and reports you want by dragging them from left to right and then click OK.



Select entries (Navigate) - New Export wizard

Navigate the folders or search to find the entries to include in your archive. Select the entries you want and click the Add button to update the Selected entries list.

Available entries
Cognos > Public Folders

Entries: 1 - 4

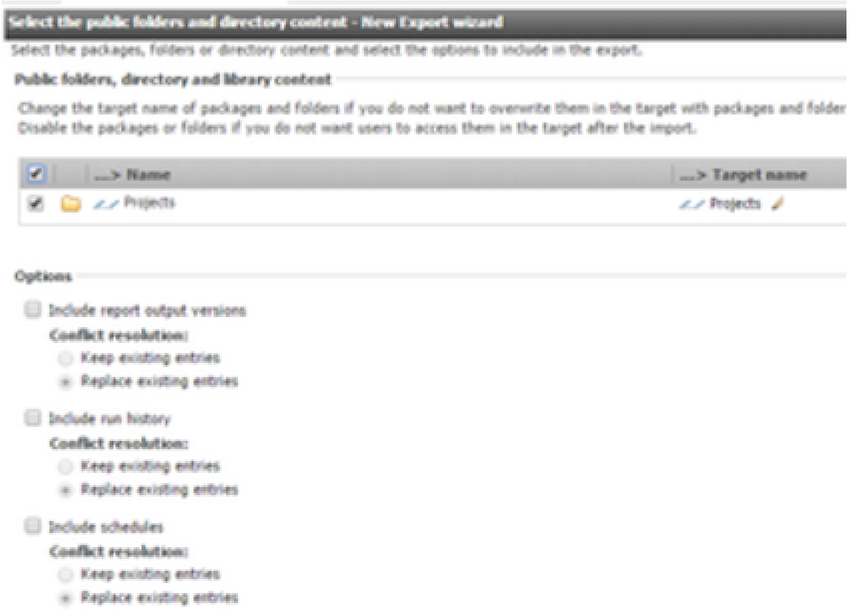
□	Name
□	Projects
□	Samples
□	Samples_Audit
□	Test

Selected entries

Entries: 1 -

□	...> Name
□	Projects

g) Select all the checkbox objects and then click Next.



Select the public folders and directory content - New Export wizard

Select the packages, folders or directory content and select the options to include in the export.

Public folders, directory and library content

Change the target name of packages and folders if you do not want to overwrite them in the target with packages and folder
Disable the packages or folders if you do not want users to access them in the target after the import.

☑	...> Name	...> Target name
☑	Projects	Projects

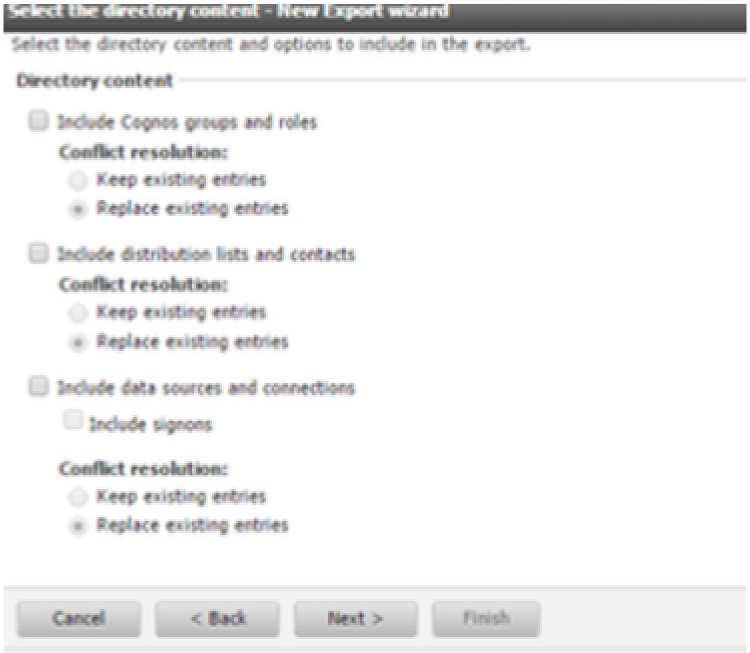
Options

□ Include report output versions
Conflict resolution:
○ Keep existing entries
● Replace existing entries

□ Include run history
Conflict resolution:
○ Keep existing entries
● Replace existing entries

□ Include schedules
Conflict resolution:
○ Keep existing entries
● Replace existing entries

h) Click Next.



Select the directory content - New Export wizard

Select the directory content and options to include in the export.

Directory content

□ Include Cognos groups and roles

Conflict resolution:

○ Keep existing entries

● Replace existing entries

□ Include distribution lists and contacts

Conflict resolution:

○ Keep existing entries

● Replace existing entries

□ Include data sources and connections

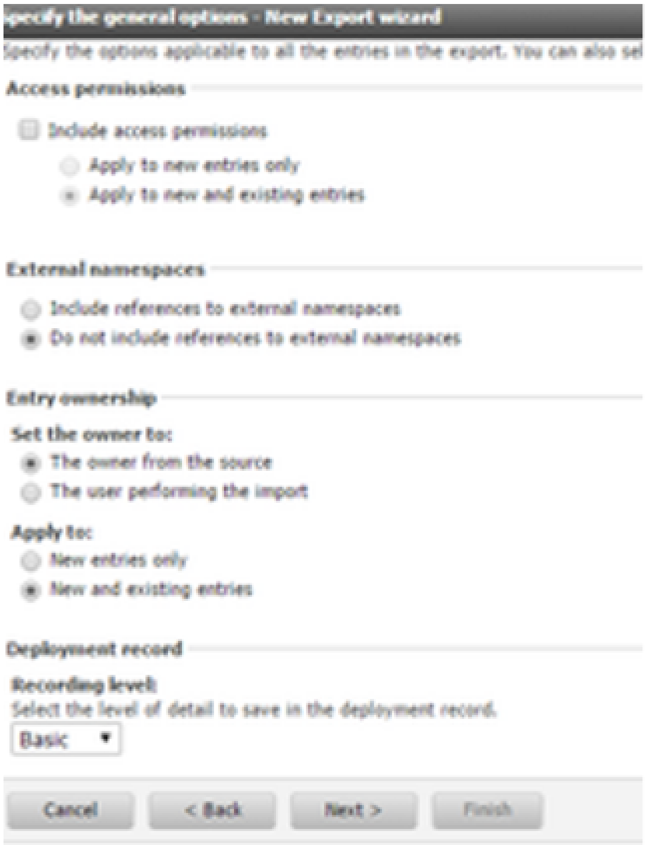
□ Include signons

Conflict resolution:

○ Keep existing entries

● Replace existing entries

i) Click Next.



pecify the general options - New Export wizard

pecify the options applicable to all the entries in the export. You can also sel

Access permissions

□ Include access permissions

○ Apply to new entries only

● Apply to new and existing entries

External namespaces

○ Include references to external namespaces

● Do not include references to external namespaces

Entry ownership

Set the owner to:

● The owner from the source

○ The user performing the import

Apply to:

○ New entries only

● New and existing entries

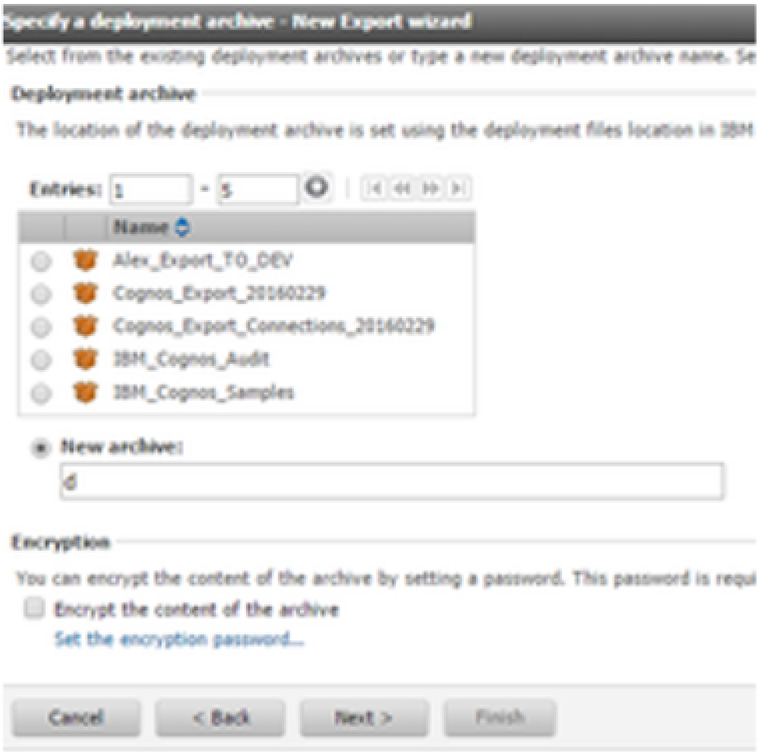
Deployment record

Recording level:

Select the level of detail to save in the deployment record.

Basic ▾

j) Click Next.



Specify a deployment archive - New Export wizard

Select from the existing deployment archives or type a new deployment archive name. Se

Deployment archive

The location of the deployment archive is set using the deployment files location in IBM

Entries: 1 - 5

	Name
○	Alex_Export_TO_DEV
○	Cognos_Export_20160229
○	Cognos_Export_Connections_20160229
○	IBM_Cognos_Audit
○	IBM_Cognos_Samples

● New archive:

d

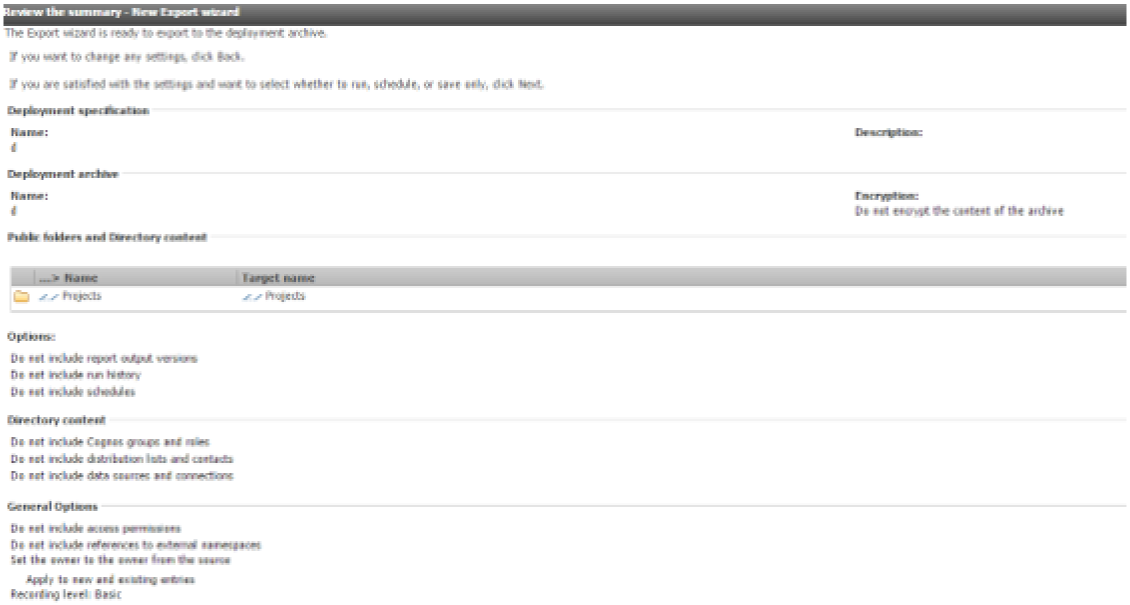
Encryption

You can encrypt the content of the archive by setting a password. This password is requi

□ Encrypt the content of the archive

Set the encryption password...

k) Click Next.



Review the summary - New Export wizard

The Export wizard is ready to export to the deployment archive.

If you want to change any settings, click Back.

If you are satisfied with the settings and want to select whether to run, schedule, or save only, click Next.

Deployment specification

Name:
d

Description:

Deployment archive

Name:
d

Encryption:
Do not encrypt the content of the archive

Public folders and Directory content

...> Name	Target name
Projects	Projects

Options:

Do not include report output versions
Do not include run history
Do not include schedules

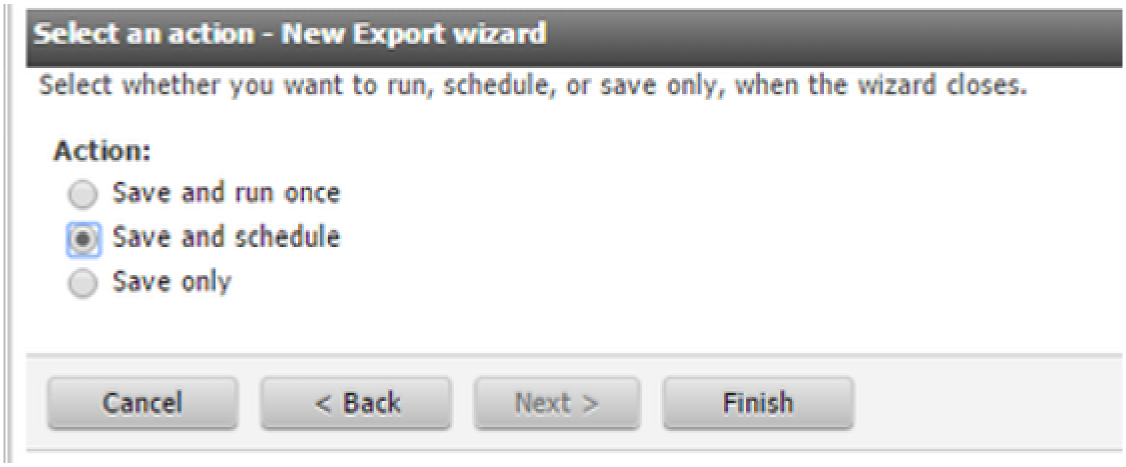
Directory content

Do not include Cognos groups and roles
Do not include distribution lists and contacts
Do not include data sources and connections

General Options

Do not include access permissions
Do not include references to external namespaces
Set the owner to the owner from the source
Apply to new and existing entries
Recording level: Basic

l) Choose the Save and run once or the Save and schedule checkbox and click Finish.



Select an action - New Export wizard

Select whether you want to run, schedule, or save only, when the wizard closes.

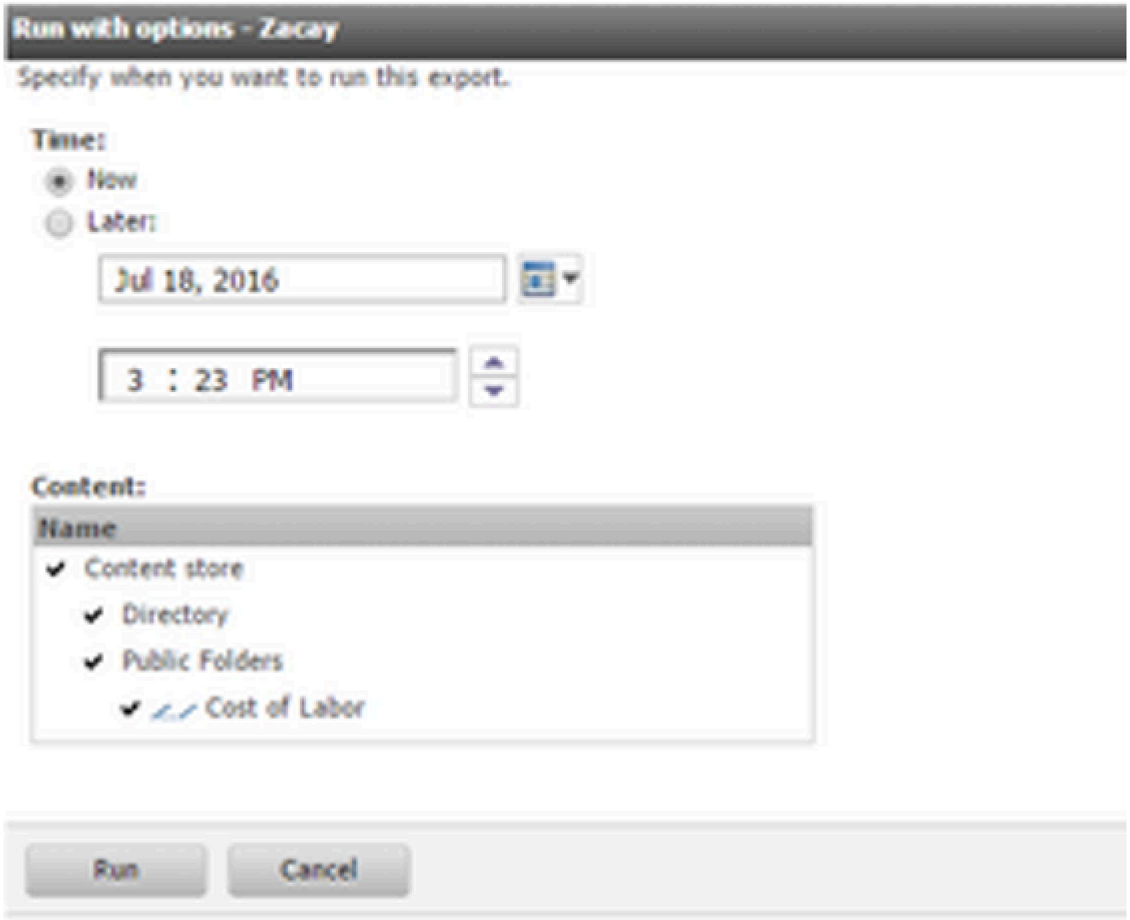
Action:

○ Save and run once

● Save and schedule

○ Save only

m) If you chose Save and schedule, choose when to schedule and then click Run.



Run with options - Zacay

Specify when you want to run this export.

Time:

● Now

○ Later:

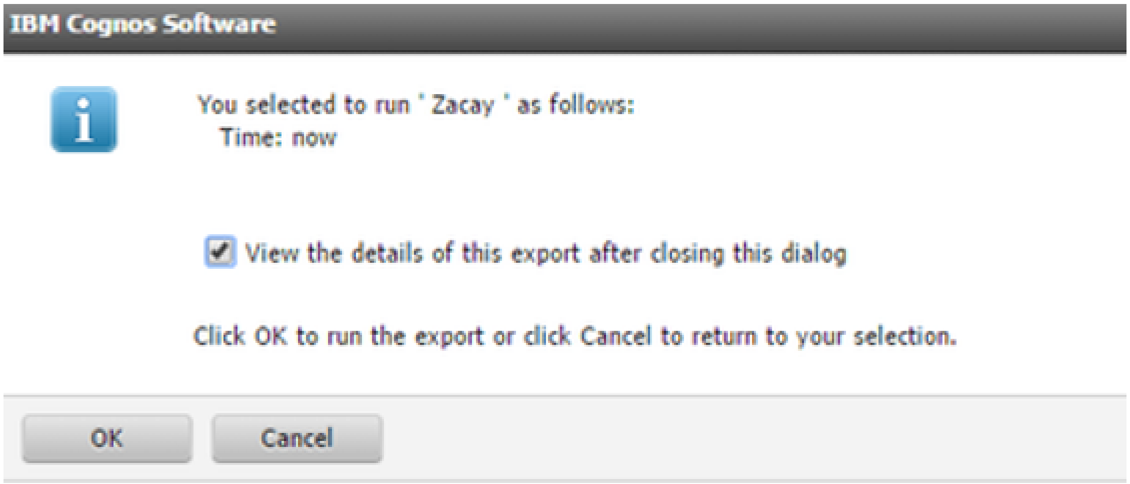
Jul 18, 2016

3 : 23 PM

Content:

Name
✓ Content store
✓ Directory
✓ Public Folders
✓ Cost of Labor

n) Click OK.



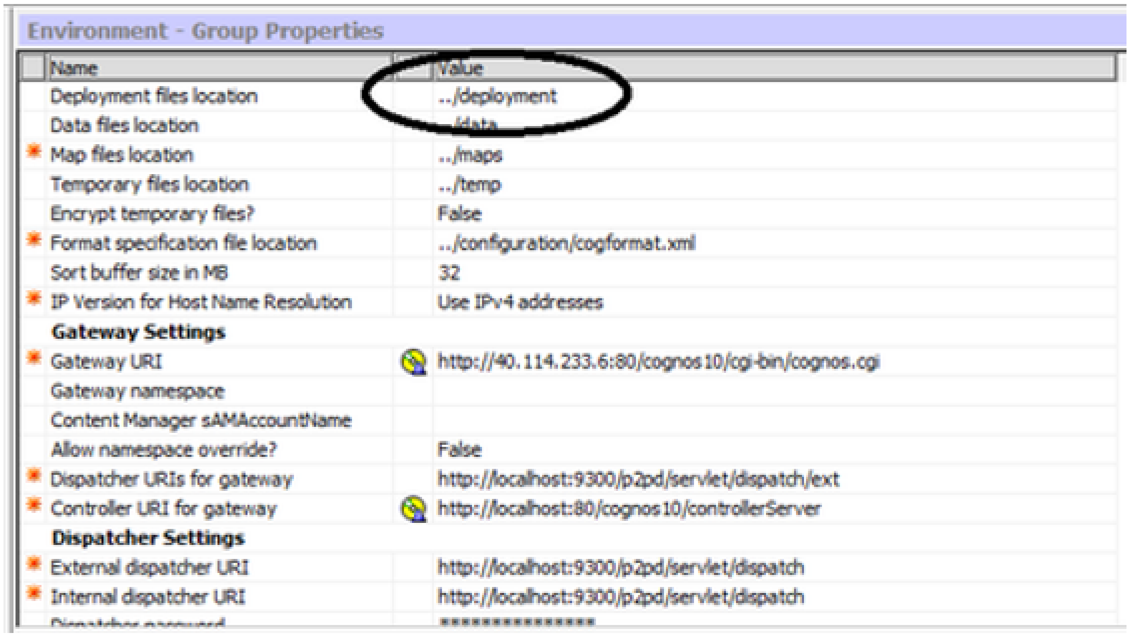
IBM Cognos Software

You selected to run ' Zacay ' as follows:
Time: now

☑ View the details of this export after closing this dialog

Click OK to run the export or click Cancel to return to your selection.

The file is saved under the deployment folder on the Cognos server. Usually the installation folder is at C:\Program Files\ibm\cognos\c10_64\deployment. The extracted file will appear in the destination defined folder.



Environment - Group Properties

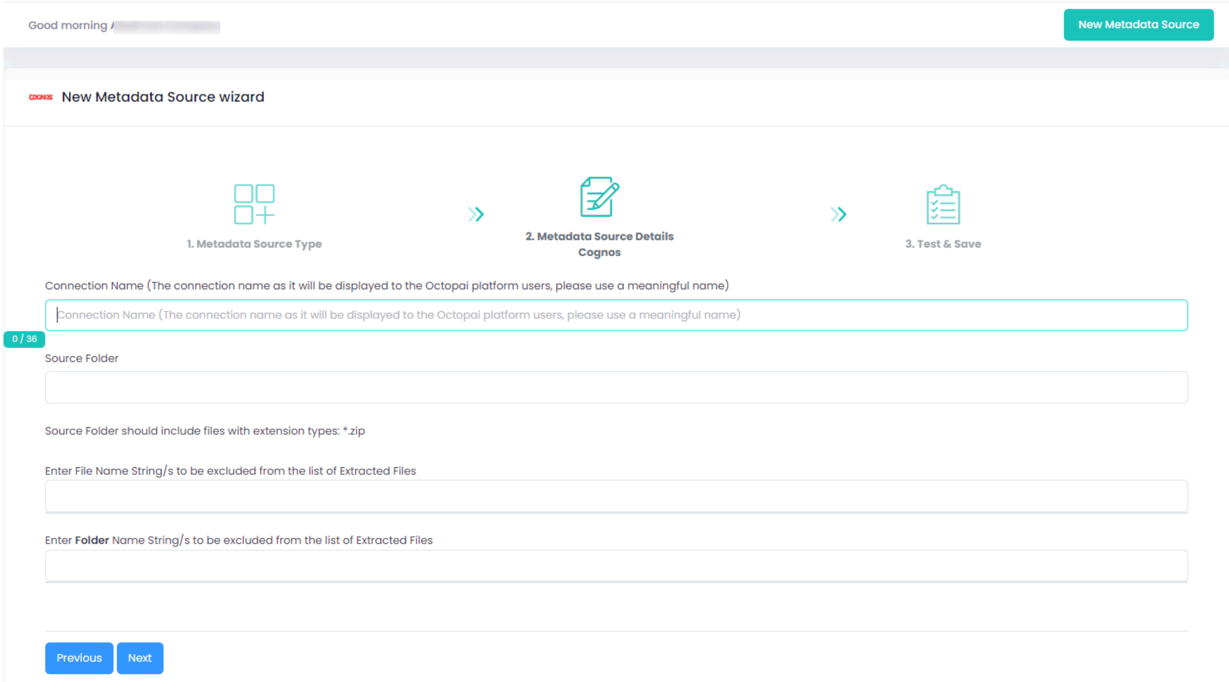
Name	Value
Deployment files location	../deployment
Data files location	../data
* Map files location	../maps
Temporary files location	../temp
Encrypt temporary files?	False
* Format specification file location	../configuration/cogformat.xml
Sort buffer size in MB	32
* IP Version for Host Name Resolution	Use IPv4 addresses
Gateway Settings	
* Gateway URI	http://40.114.233.6:80/cognos10/cgi-bin/cognos.cgi
Gateway namespace	
Content Manager sAMAccountName	
Allow namespace override?	False
* Dispatcher URIs for gateway	http://localhost:9300/p2pd/servlet/dispatch/ext
* Controller URI for gateway	http://localhost:80/cognos10/controllerServer
Dispatcher Settings	
* External dispatcher URI	http://localhost:9300/p2pd/servlet/dispatch
* Internal dispatcher URI	http://localhost:9300/p2pd/servlet/dispatch
Dispatcher password	***************

2. Zip the metadata result file and use it as the source for your Cloudera Octopai metadata Cognos connection in the Cloudera Octopai Client.

   Note: Ensure that you have appropriate permissions to the path so that the Cloudera Octopai Client can access the file with the user running it.

3. Upload the zip file to the prospect folder in the portal.
4. Share the password of the Cognos package or framework file that defined in the guideline.
5. Share the prospect IBM Cognos version.

6. Set up the IBM Cognos Metadata Source.



Good morning [redacted]

New Metadata Source wizard

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Source Folder

Source Folder should include files with extension types: *.zip

Enter File Name String/s to be excluded from the list of Extracted Files

Enter **Folder** Name String/s to be excluded from the list of Extracted Files

7. Verify the extracted metadata file by accessing the Cloudera Octopai Target Folder (TGT).
   a) Go to the TGT Folder located on the server where the Cloudera Octopai Client is installed.

      The default location is C:\Program Files (x86)\Octopai\Service\TGT.
   b) Open the zip file with the connector name.

      Example

      POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53
   c) Verify the file content by checking the quantity and quality of the included files.

If an error occurred during the extraction, perform the following troubleshooting steps:

1. Check the permissions.
2. Send the log with the connector number and name to Cloudera Support.

   You can find the log files at C:\Program Files (x86)\Octopai\Service\log.

POWER_BI_103_21520221025220220215 15/02/2022 10:02 LOG File 3 KB

Cognos Operational Metadata Intelligence Harvester

Learn how to integrate Cognos metadata into Cloudera Octopai by configuring permissions, setting up the metadata source, and verifying the extracted files for enhanced analysis.

Tool Permissions Prerequisites

Ensure the following prerequisites are met before proceeding with the configuration:

1. **API Enablement**:
   - Confirm that the **Cognos Operational Metadata API** is enabled through your Cognos license.

2. **Database Login Permissions**:
   - Secure login credentials with **SELECT** permissions to the relevant Cognos databases and tables.

   From the Audit Database:

   - COGIPF_RUNREPORT
   - COGIPF_USERLOGON

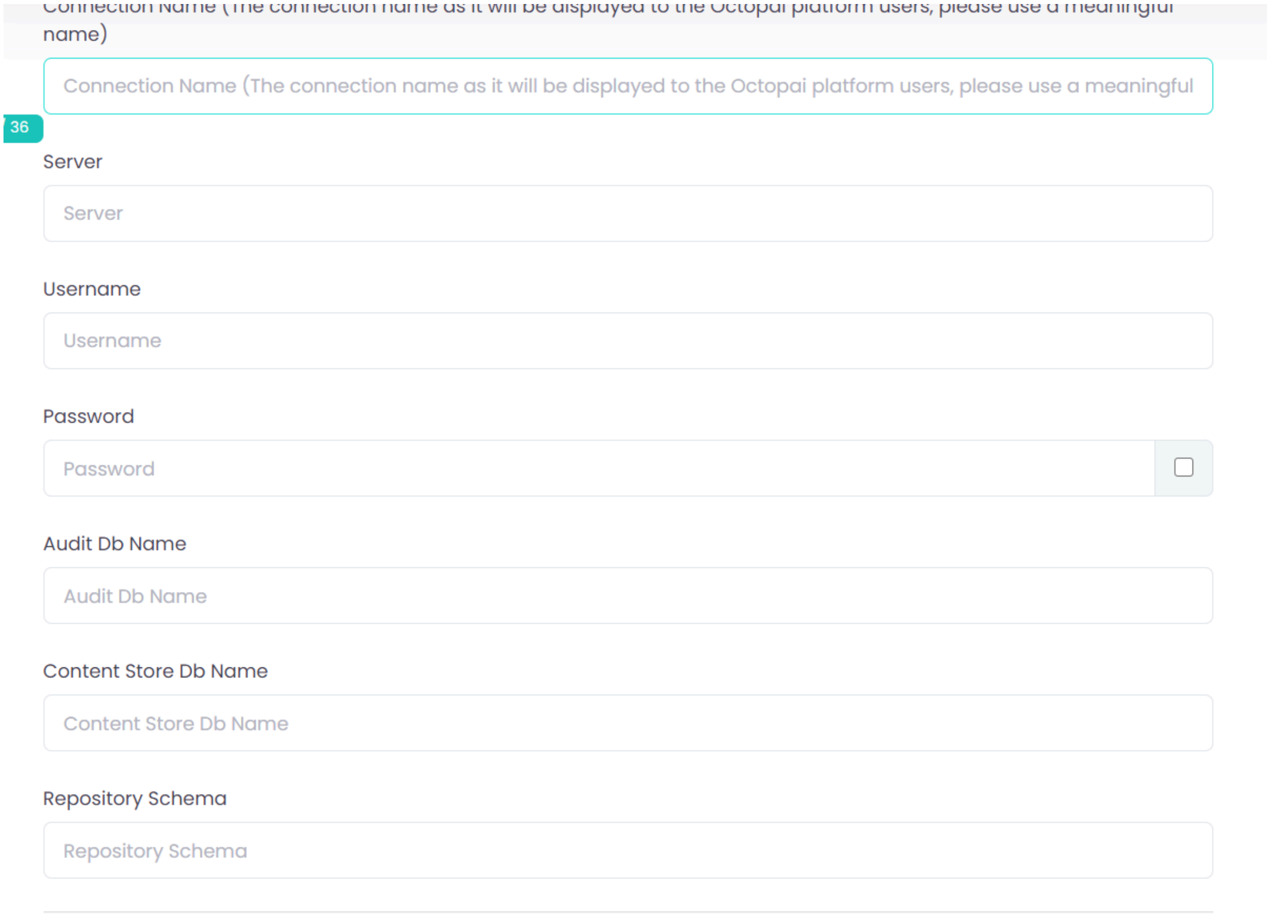
   From the Content Share Database:

   - CMOBJECTS
   - CMCLASSES
   - CMOBJNAMES
   - CMREFNOORD1
   - CMREFNOORD2
   - CMOBJPROPS33
3. **Database Details**:
   - Relevant SQL Server instance details.
   - Username and password with appropriate permissions.

Setting Up Cognos Metadata Source in Cloudera Octopai

Open Cloudera Octopai Client:

- Access the **Octopai Client** installed on your server.



Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful

36

Server

Server

Username

Username

Password

Password □

Audit Db Name

Audit Db Name

Content Store Db Name

Content Store Db Name

Repository Schema

Repository Schema

Metadata Source Configuration:

- Set the Cognos Metadata Source within the client interface.

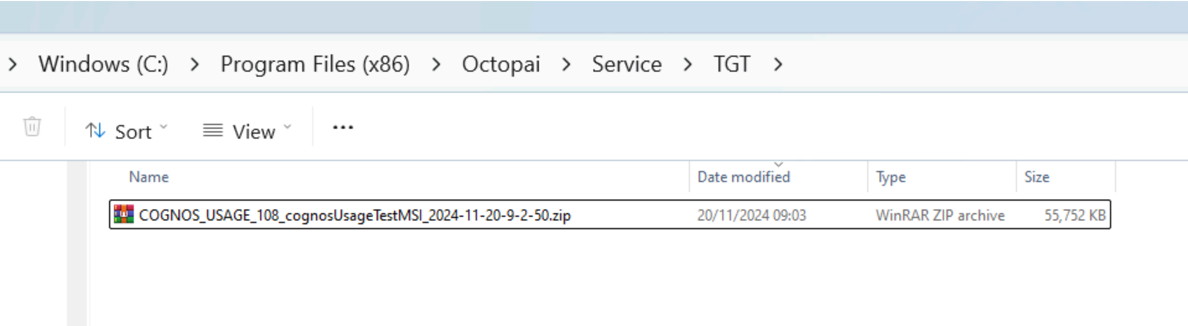
  The **Repository Schema** field is optional. If not specified, it defaults to dbo.

Verifying the Extracted Metadata File

Once the metadata extraction process is complete, verify the extracted file as follows:

1. **Locate the Target Folder**:
   - Navigate to the **Octopai Target Folder (TGT)** on the server where the Cloudera Octopai Client is installed. Default path:

     C:\Program Files (x86)\Octopai\Service\TGT



Windows (C:) > Program Files (x86) > Octopai > Service > TGT >
Sort
View
...

Name	Date modified	Type	Size
COGNOS_USAGE_108_cognosUsageTestMSI_2024-11-20-9-2-50.zip	20/11/2024 09:03	WinRAR ZIP archive	55,752 KB

2. **Open the Extracted File**:
   - Find the ZIP file corresponding to the connector name (e.g., Cognos_Metadata.zip).
3. **Validate File Content**:
   - Check the quantity and quality of the inner files to ensure all expected metadata has been successfully extracted.

By completing the above steps, Cognos Operational Metadata will be integrated into Cloudera Octopai for enhanced metadata intelligence and analysis.

SAP Business Objects (BO)

Extract LCMBIAR files from SAP Business Objects (BO) for Cloudera Octopai integration using the Central Management Console or Promotion Management Wizard.

Note: Version supported: up to 4.3

Tool Permissions Prerequisites

Warning: Missing permissions could end up in broken lineages.

Enable read permission for Cloudera Octopai Windows NT User to the Business Objects BIAR files folder.

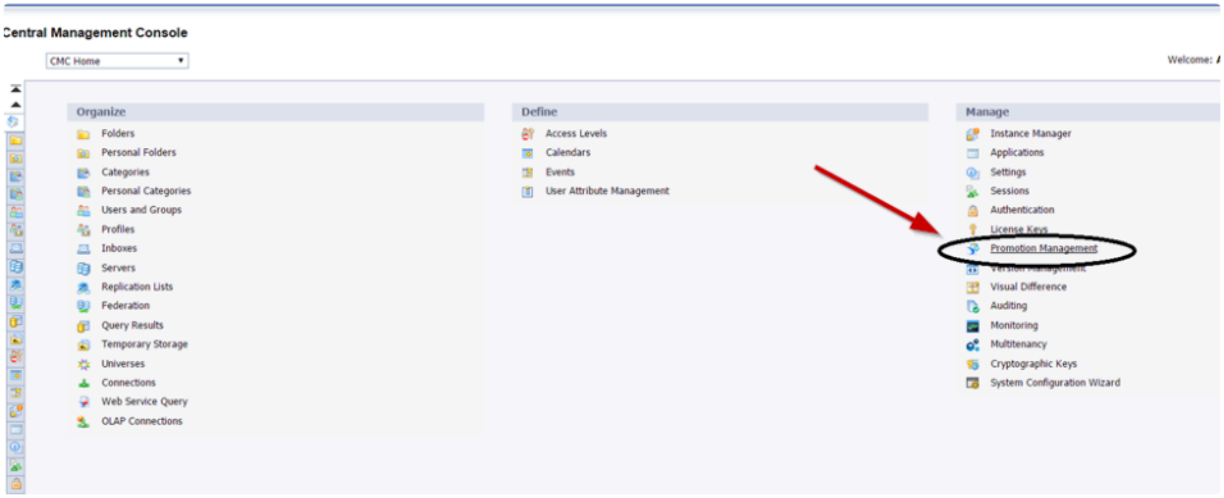
How to extract LCMBIAR files

You can extract LCMBIAR files using one of the following methods:

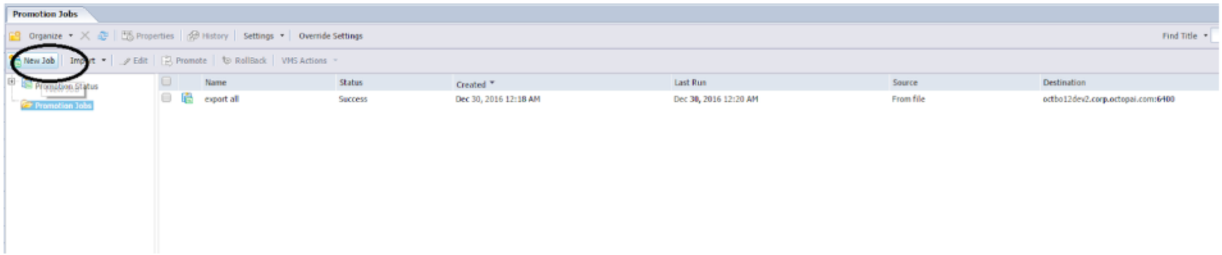
Method 1: Export LCMBIAR with the Central Management Console

1. Log in to CMC: http://bo-server-name.com:8080/BOE/CMC

2. Click on promotion management:

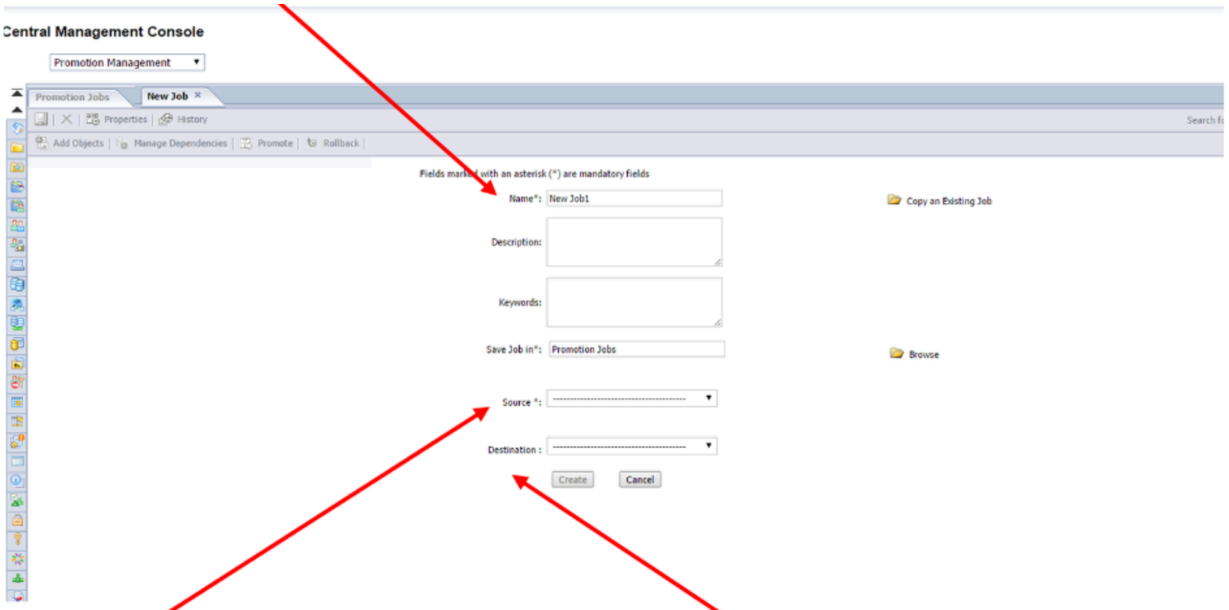


3. Create a new job:



Name	Status	Created	Last Run	Source	Destination
export all	Success	Dec 30, 2016 12:18 AM	Dec 30, 2016 12:20 AM	From file	octbo12dev2.corp.octopai.com:6400

4. Fill in a job name and details:



Fields marked with an asterisk (*) are mandatory fields

Name*: New Job1

Description:

Keywords:

Save Job in*: Promotion Jobs

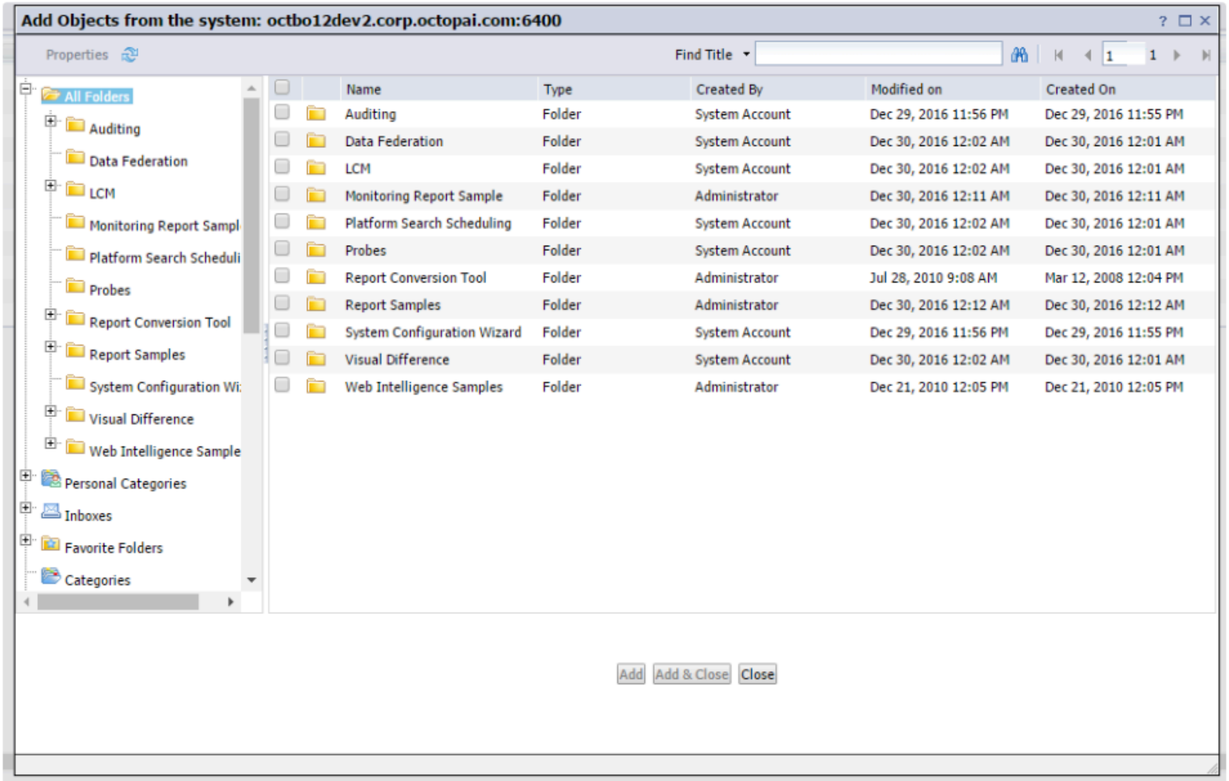
Source *:

Destination :

Source - fill in the login details

Destination - Output for LCMBIAR file

5. Click on "Create":



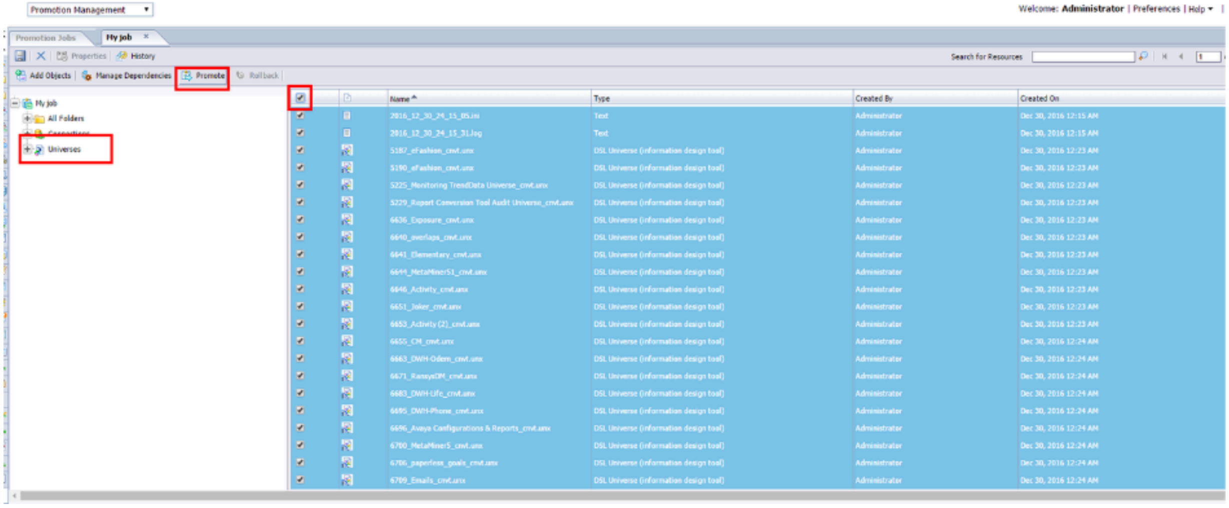
Add Objects from the system: octbo12dev2.corp.octopai.com:6400

□		Name	Type	Created By	Modified on	Created On
□		Auditing	Folder	System Account	Dec 29, 2016 11:56 PM	Dec 29, 2016 11:55 PM
□		Data Federation	Folder	System Account	Dec 30, 2016 12:02 AM	Dec 30, 2016 12:01 AM
□		LCM	Folder	System Account	Dec 30, 2016 12:02 AM	Dec 30, 2016 12:01 AM
□		Monitoring Report Sample	Folder	Administrator	Dec 30, 2016 12:11 AM	Dec 30, 2016 12:11 AM
□		Platform Search Scheduling	Folder	System Account	Dec 30, 2016 12:02 AM	Dec 30, 2016 12:01 AM
□		Probes	Folder	System Account	Dec 30, 2016 12:02 AM	Dec 30, 2016 12:01 AM
□		Report Conversion Tool	Folder	Administrator	Jul 28, 2010 9:08 AM	Mar 12, 2008 12:04 PM
□		Report Samples	Folder	Administrator	Dec 30, 2016 12:12 AM	Dec 30, 2016 12:12 AM
□		System Configuration Wizard	Folder	System Account	Dec 29, 2016 11:56 PM	Dec 29, 2016 11:55 PM
□		Visual Difference	Folder	System Account	Dec 30, 2016 12:02 AM	Dec 30, 2016 12:01 AM
□		Web Intelligence Samples	Folder	Administrator	Dec 21, 2010 12:05 PM	Dec 21, 2010 12:05 PM

6. On the left side panel choose Universes and then select all by clicking the upper checkbox
7. Click add
8. Repeat steps from #6 above for **All Connections**
9. Repeat steps from #6 above for **All folders (reports)**

Note: There is a 500 MB file size limit for LCMBIAR files, In the event report files are larger than 500 MB please divide this step as necessary by breaking up the selection of folders and creating separate output files

10. Click on close.
11. Select all and click on the promote button, then choose - **export now**

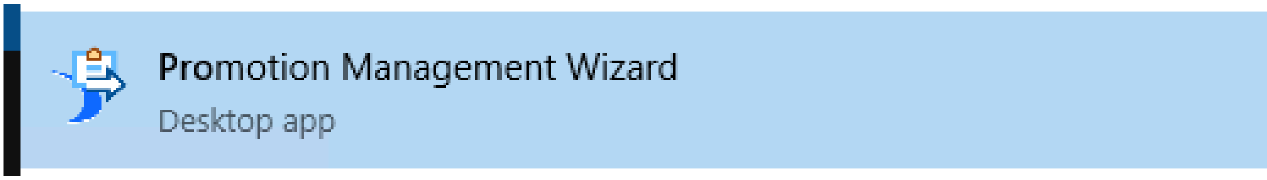


☑		Name	Type	Created By	Created On
☑		2016_12_30_24_15_05.ini	Text	Administrator	Dec 30, 2016 12:15 AM
☑		2016_12_30_24_15_31.log	Text	Administrator	Dec 30, 2016 12:15 AM
☑		5187_eFashion_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:23 AM
☑		5190_eFashion_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:23 AM
☑		5225_Monitoring TrendData Universe_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:23 AM
☑		5229_Report Conversion Tool Audit Universe_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:23 AM
☑		6636_Exposure_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:23 AM
☑		6640_overlaps_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:23 AM
☑		6641_Elementary_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:23 AM
☑		6644_MetaMiner51_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:23 AM
☑		6646_Activity_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:23 AM
☑		6651_Joker_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:23 AM
☑		6653_Activity (2)_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:23 AM
☑		6655_CM_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:24 AM
☑		6663_DWH-Odem_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:24 AM
☑		6671_RansysDM_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:24 AM
☑		6683_DWH-Life_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:24 AM
☑		6695_DWH-Phone_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:24 AM
☑		6696_Avaya Configurations & Reports_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:24 AM
☑		6700_MetaMiner5_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:24 AM
☑		6706_paperless_goals_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:24 AM
☑		6709_Emails_cnvt.unx	DSL Universe (information design tool)	Administrator	Dec 30, 2016 12:24 AM

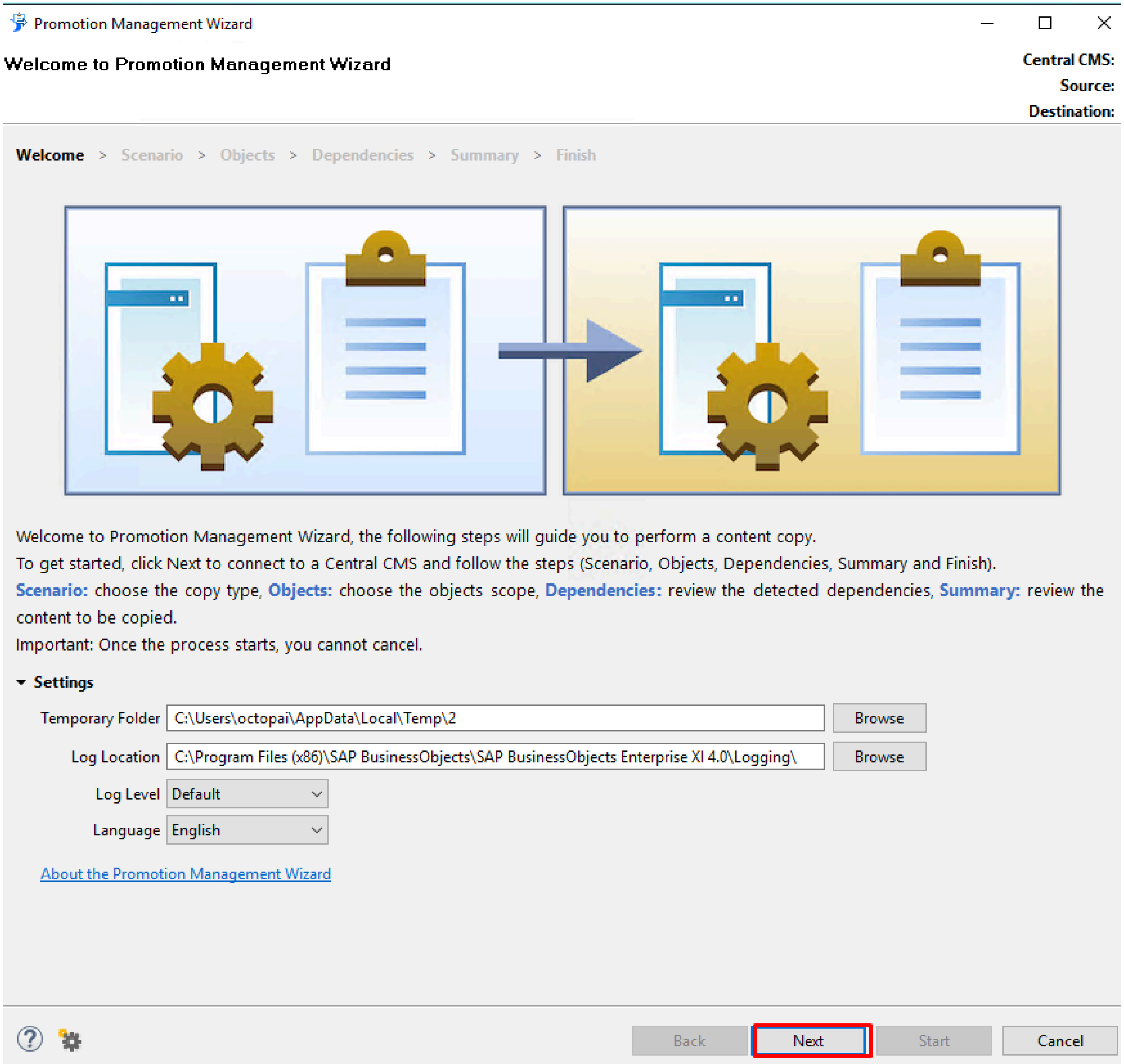
Note: The file (*.lcmbiar) will appear in the destination-defined folder.Use this file as a source for your Cloudera Octopai metadata BO connection in the Cloudera Octopai Client.

Method 2: Export LCMBIAR with the Promotion Management Wizard

Go into the BO Server and open the program called " **Promotion Management Wizard** ".



When the following window opens, click **Next**


Promotion Management Wizard
Welcome to Promotion Management Wizard
Central CMS:
Source:
Destination:
Welcome > Scenario > Objects > Dependencies > Summary > Finish
Browse
Browse
About the Promotion Management Wizard
Back
Next
Start
Cancel

Welcome to Promotion Management Wizard, the following steps will guide you to perform a content copy.
To get started, click Next to connect to a Central CMS and follow the steps (Scenario, Objects, Dependencies, Summary and Finish).
Scenario: choose the copy type, **Objects:** choose the objects scope, **Dependencies:** review the detected dependencies, **Summary:** review the content to be copied.
Important: Once the process starts, you cannot cancel.

▾ Settings

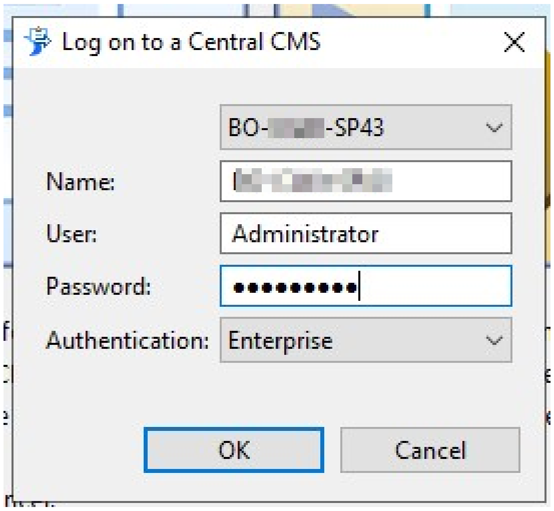
Temporary Folder: C:\Users\octopai\AppData\Local\Temp\2

Log Location: C:\Program Files (x86)\SAP BusinessObjects\SAP BusinessObjects Enterprise XI 4.0\Logging\

Log Level: Default

Language: English

Login to the CMS (BO Server) that you want to extract from.



Log on to a Central CMS

BO-[redacted]-SP43

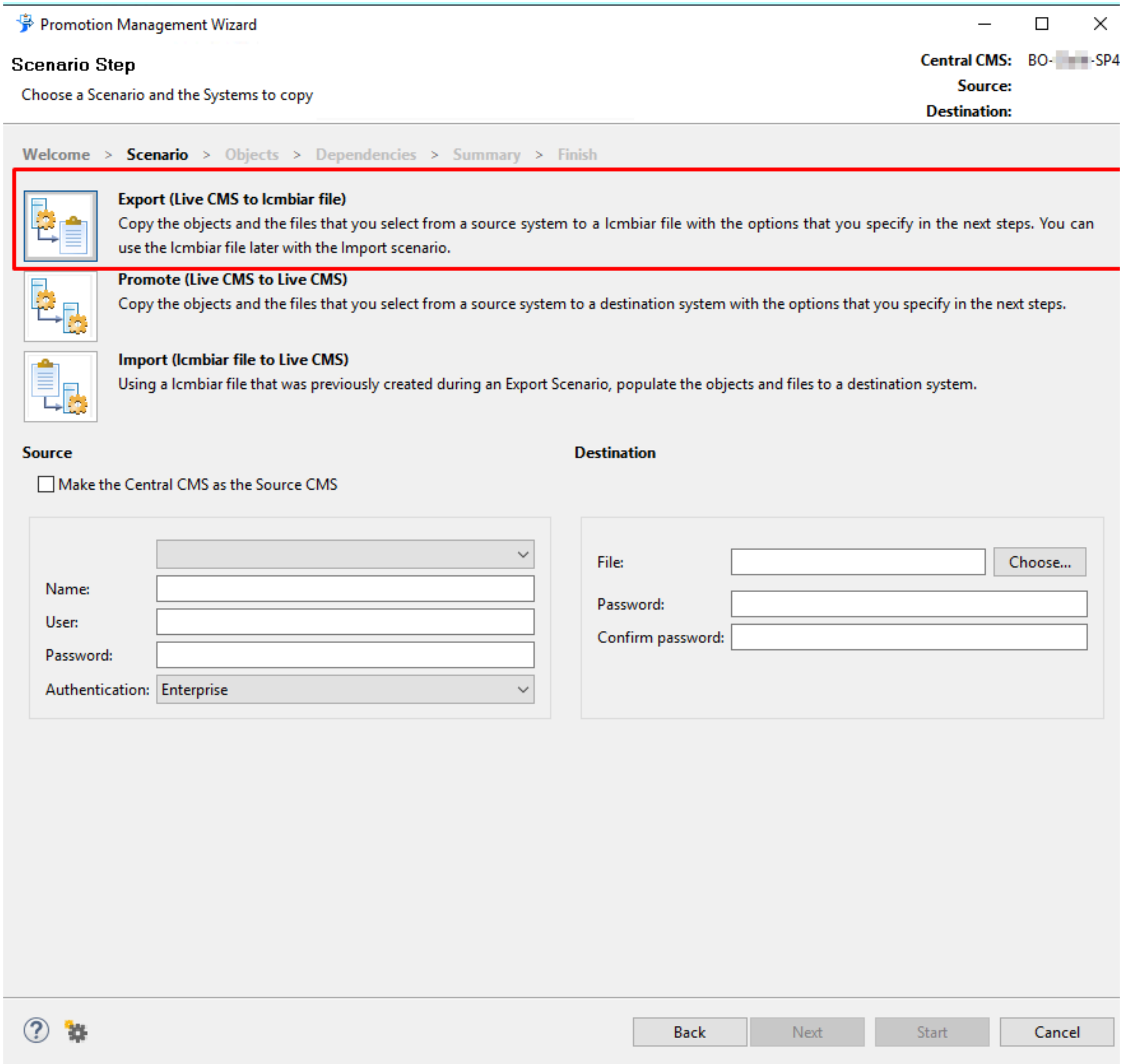
Name: [redacted]

User: Administrator

Password: ●●●●●●●●●

Authentication: Enterprise

Click on “ **Export (Live CMS to lcmbiar file)** ”



Promotion Management Wizard

Scenario Step

Choose a Scenario and the Systems to copy

Central CMS: BO-[redacted]-SP4
Source:
Destination:

Welcome > Scenario > Objects > Dependencies > Summary > Finish

Export (Live CMS to lcmbiar file)
Copy the objects and the files that you select from a source system to a lcmbiar file with the options that you specify in the next steps. You can use the lcmbiar file later with the Import scenario.

Promote (Live CMS to Live CMS)
Copy the objects and the files that you select from a source system to a destination system with the options that you specify in the next steps.

Import (lcmbiar file to Live CMS)
Using a lcmbiar file that was previously created during an Export Scenario, populate the objects and files to a destination system.

Source

□ Make the Central CMS as the Source CMS

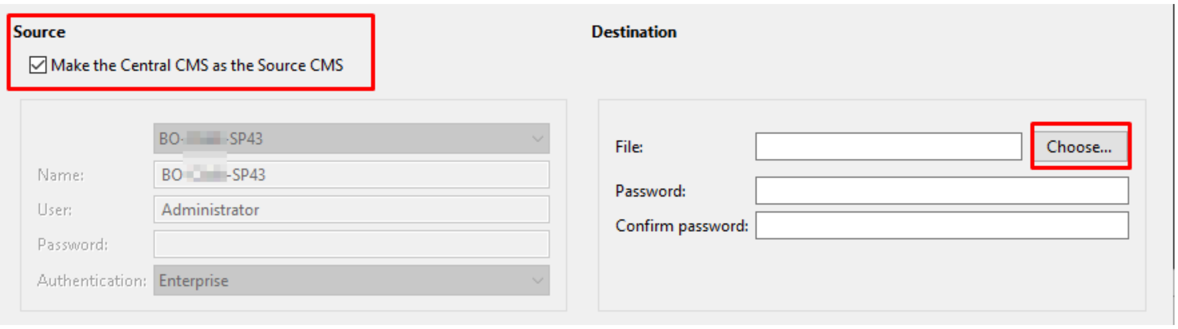
Name:
User:
Password:
Authentication: Enterprise

Destination

File:
Password:
Confirm password:

Fill up the information:

- Click on the Source tab to **"Make the Central CMS as the Source CMS"** which means to use the local CMS as our source.
- In the Destination tab choose where to save the file and give it a password. **(Password is Mandatory).**



Source

☑ Make the Central CMS as the Source CMS

BO-[redacted]-SP43

Name: BO-[redacted]-SP43

User: Administrator

Password:

Authentication: Enterprise

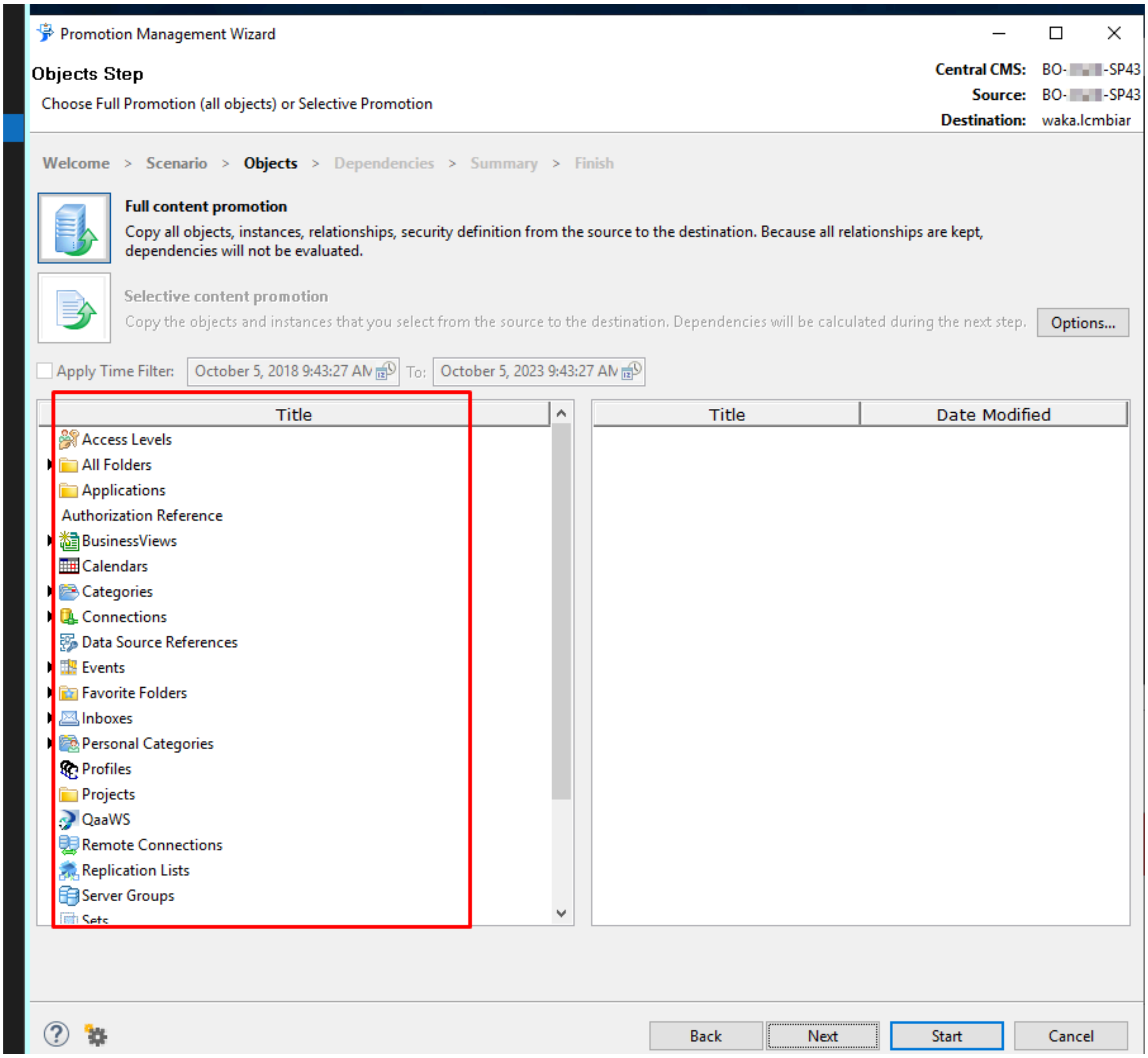
Destination

File:

Password:

Confirm password:

Choose the objects you want to export whether it's Universes, Reports, Connections.



Promotion Management Wizard

Objects Step

Choose Full Promotion (all objects) or Selective Promotion

Central CMS: BO-[redacted]-SP43
Source: BO-[redacted]-SP43
Destination: waka.lcmbiar

Welcome > Scenario > **Objects** > Dependencies > Summary > Finish

Full content promotion

Copy all objects, instances, relationships, security definition from the source to the destination. Because all relationships are kept, dependencies will not be evaluated.

Selective content promotion

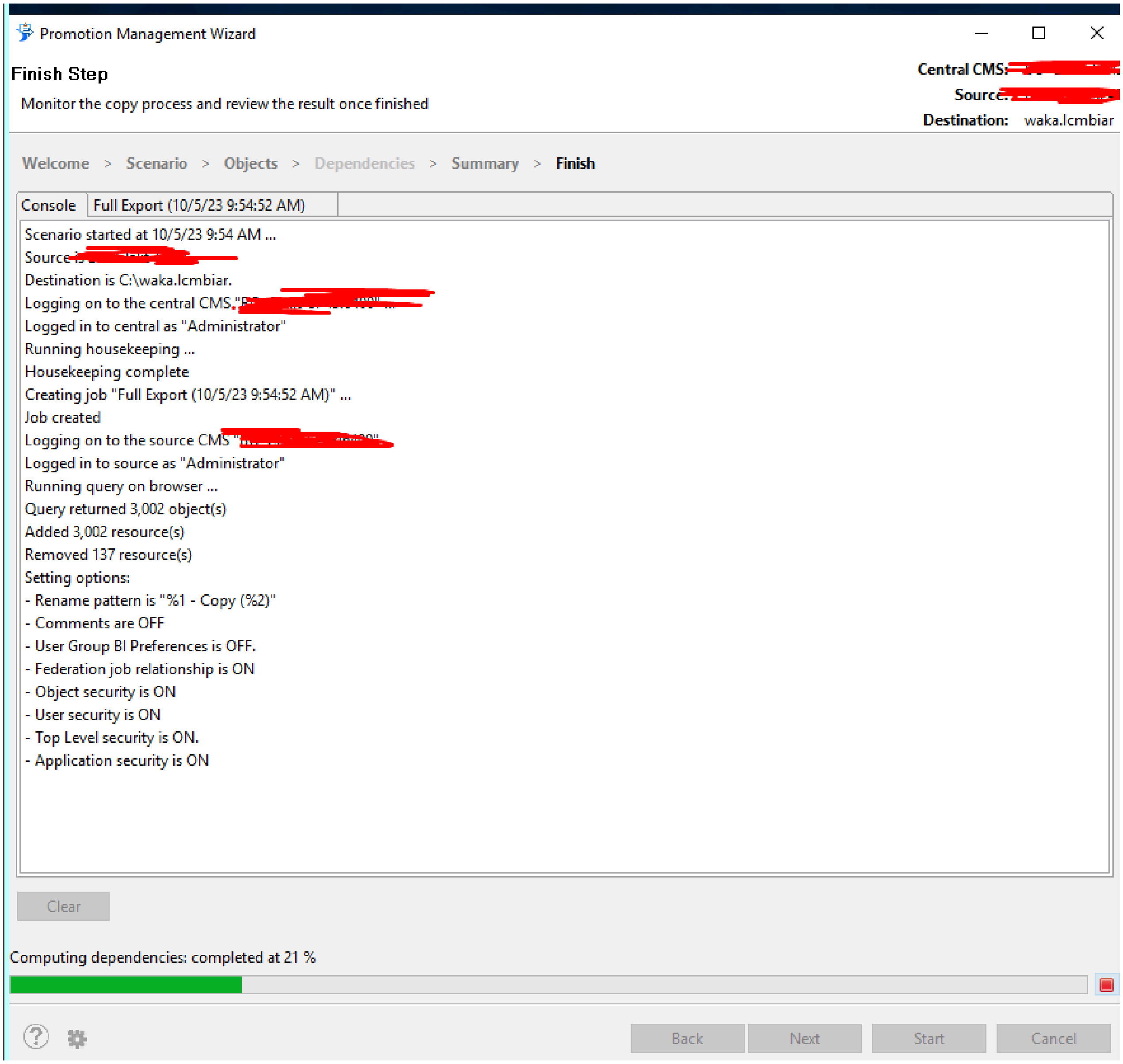
Copy the objects and instances that you select from the source to the destination. Dependencies will be calculated during the next step.

□ Apply Time Filter: October 5, 2018 9:43:27 AM To: October 5, 2023 9:43:27 AM

Title
Access Levels
All Folders
Applications
Authorization Reference
BusinessViews
Calendars
Categories
Connections
Data Source References
Events
Favorite Folders
Inboxes
Personal Categories
Profiles
Projects
QaaWS
Remote Connections
Replication Lists
Server Groups
Sets

Title	Date Modified

Then afterward select **Start** .


Promotion Management Wizard
Welcome > Scenario > Objects > Dependencies > Summary > Finish
Console
Full Export (10/5/23 9:54:52 AM)
Clear
Back
Next
Start
Cancel

Finish Step

Monitor the copy process and review the result once finished

Central CMS: [redacted]
Source: [redacted]
Destination: waka.lcmbiar

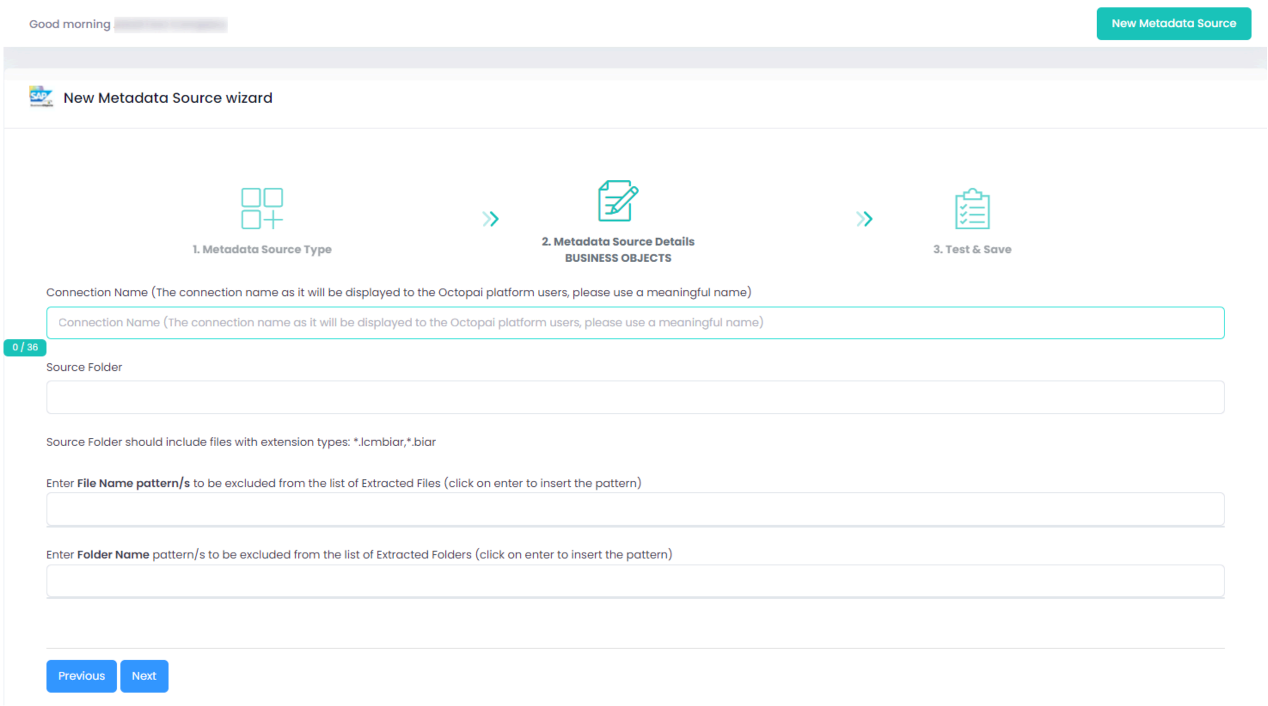
```
Scenario started at 10/5/23 9:54 AM ...
Source is [redacted]
Destination is C:\waka.lcmbiar.
Logging on to the central CMS "[redacted]" ...
Logged in to central as "Administrator"
Running housekeeping ...
Housekeeping complete
Creating job "Full Export (10/5/23 9:54:52 AM)" ...
Job created
Logging on to the source CMS "[redacted]"
Logged in to source as "Administrator"
Running query on browser ...
Query returned 3,002 object(s)
Added 3,002 resource(s)
Removed 137 resource(s)
Setting options:
- Rename pattern is "%1 - Copy (%2)"
- Comments are OFF
- User Group BI Preferences is OFF.
- Federation job relationship is ON
- Object security is ON
- User security is ON
- Top Level security is ON.
- Application security is ON
```

Computing dependencies: completed at 21 %

All done!Once the file has been created, upload the file to the Cloudera Octopai Portal.

Setting up SAP-BO Metadata Source

Metadata Sources are set on the Cloudera Octopai Client



Good morning [redacted]

New Metadata Source wizard

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Source Folder

Source Folder should include files with extension types: *.lcmbiar,*.biar

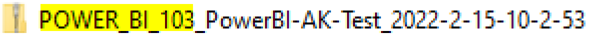
Enter **File Name pattern/s** to be excluded from the list of Extracted Files (click on enter to insert the pattern)

Enter **Folder Name** pattern/s to be excluded from the list of Extracted Folders (click on enter to insert the pattern)

How to verify the extracted Metadata File

Access the Cloudera Octopai Target Folder (TGT)

1. Go to the TGT Folder located on the Server where the Cloudera Octopai Client is installed.By default: **C:\Program Files (x86)\Octopai\Service\TGT**
2. Open the zip file having the Connector NameExample:



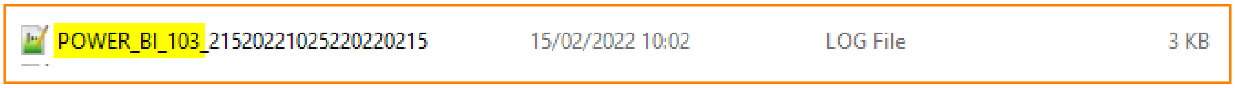
POWER_BI_103_PowerBI-AK-Test_2022-2-15-10-2-53

3. Verify its contentQuantity & Quality of inner files

Troubleshoot

Error during the extraction:

- Check the permissions
- Send the log with the connector number and name to Cloudera Support - C:\Program Files (x86)\Octopai\Service\log



POWER_BI_103_21520221025220220215 15/02/2022 10:02 LOG File 3 KB

Power BI & Power BI Report Server

Learn how to configure Power BI and Power BI Report Server to integrate with Cloudera Octopai.

Note: Version supported: up to 2023

Tool Permissions Prerequisites

Note: Missing permissions could end up in broken lineages.

On-Prem Version:

- Local Folder - Read Permission for Cloudera Octopai Windows NT User on Power BI local folder - *.pbix files.
- Shared folder (SMB/Remote) - Full Control permission for Cloudera Octopai Windows NT User on Power BI shared folder (SMB/Remote) - *.pbix files.
- Power BI Web (Cloud App) - Azure Registered Application with relevant permissions (Please use the provided Power BI Settings for the Cloudera Octopai Extraction file) and a Power BI Premium user.
- **Note:** Multi-Factor Authentication (MFA) is not supported.

CLOUD Version:

- Power BI Web (Cloud App) - Client Secret Authentication (Open the Power BI Settings for Cloud Version Guide).
- Azure Registered Application with relevant permissions (Open the Power BI Web App Registration Guide).
- Tenant ID.
- Client Secret.

Power BI Report Server:

- ‘Browse’ and 'Content Manager' Role assigned for reports you would like extracted.

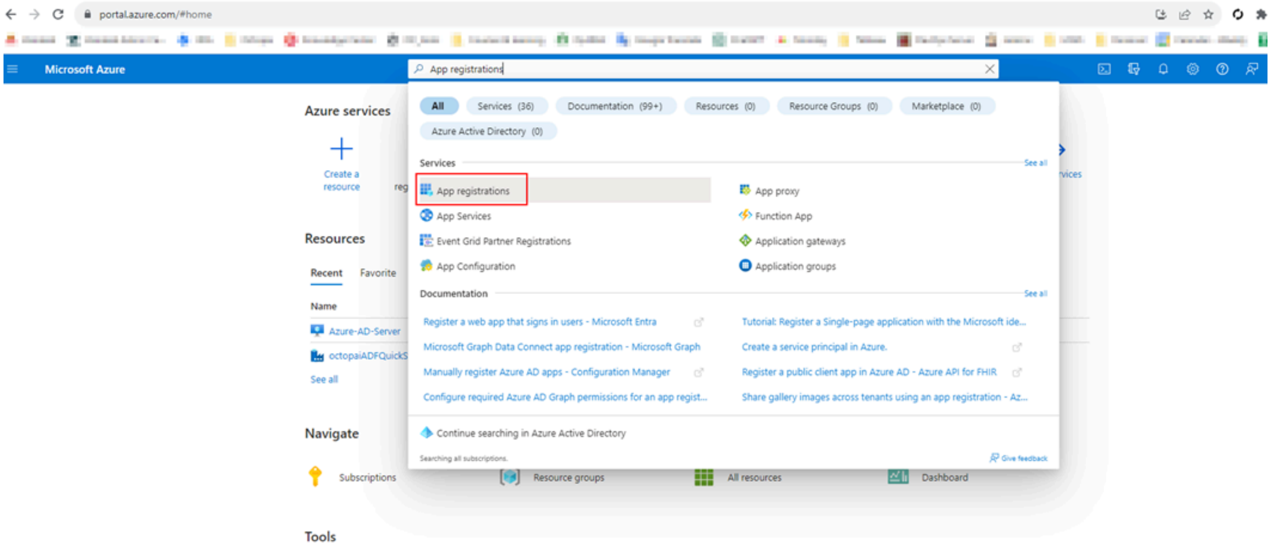
Permissions and configurations setup in Microsoft Azure

Important: Important clarification ahead of this guide:

*** Changes in Azure and PowerBI portals typically take 15-30 minutes to apply but may take up to 24 hours in some cases.**

Go to https://portal.azure.com/

Then, go to App registrations.

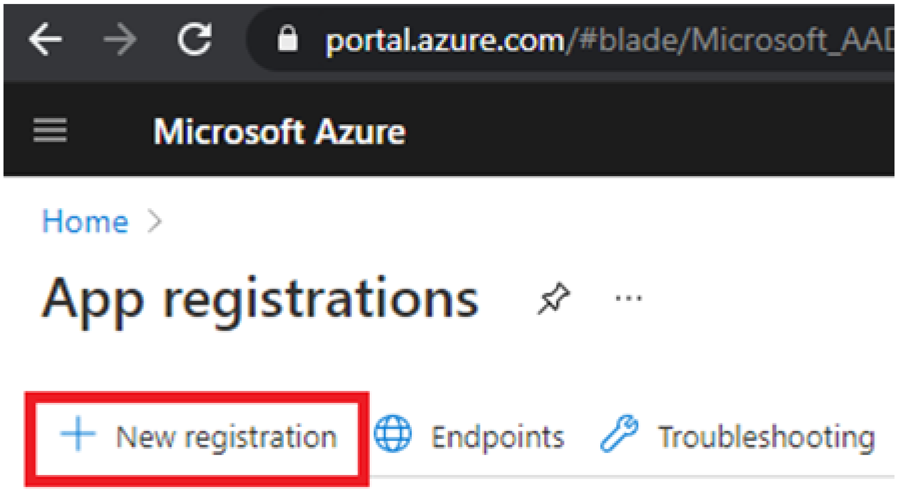


portal.azure.com/#home

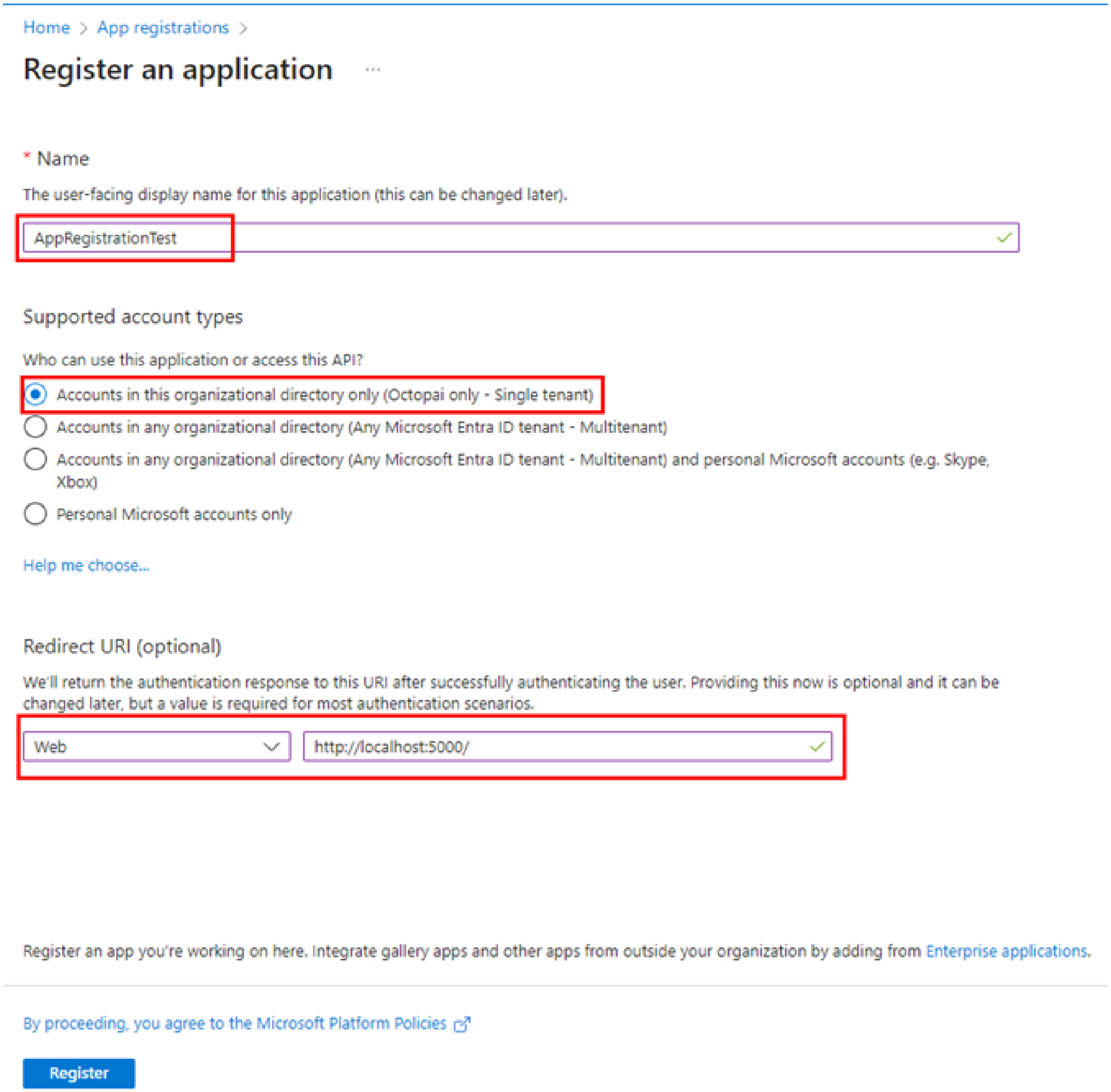
App registrations

Microsoft Azure
All
Services (36)
Documentation (99+)
Resources (0)
Resource Groups (0)
Marketplace (0)
Azure Active Directory (0)
Azure services
Create a resource
Services
See all
App registrations
App Services
Event Grid Partner Registrations
App Configuration
App proxy
Function App
Application gateways
Application groups
Documentation
Register a web app that signs in users - Microsoft Entra
Microsoft Graph Data Connect app registration - Microsoft Graph
Manually register Azure AD apps - Configuration Manager
Configure required Azure AD Graph permissions for an app regist...
Tutorial: Register a Single-page application with the Microsoft ide...
Create a service principal in Azure.
Register a public client app in Azure AD - Azure API for FHIR
Share gallery images across tenants using an app registration - Az...
Continue searching in Azure Active Directory
Searching all subscriptions.
Give feedback
Resources
Recent
Favorite
Name
Azure-AD-Server
octopaiADFQuickS
See all
Navigate
Subscriptions
Resource groups
All resources
Dashboard
Tools

Click New Registration.



Configure as below and click Register.



Home > App registrations >

Register an application ...

* Name

The user-facing display name for this application (this can be changed later).

AppRegistrationTest ✓

Supported account types

Who can use this application or access this API?

● Accounts in this organizational directory only (Octopai only - Single tenant)

○ Accounts in any organizational directory (Any Microsoft Entra ID tenant - Multitenant)

○ Accounts in any organizational directory (Any Microsoft Entra ID tenant - Multitenant) and personal Microsoft accounts (e.g. Skype, Xbox)

○ Personal Microsoft accounts only

Help me choose...

Redirect URI (optional)

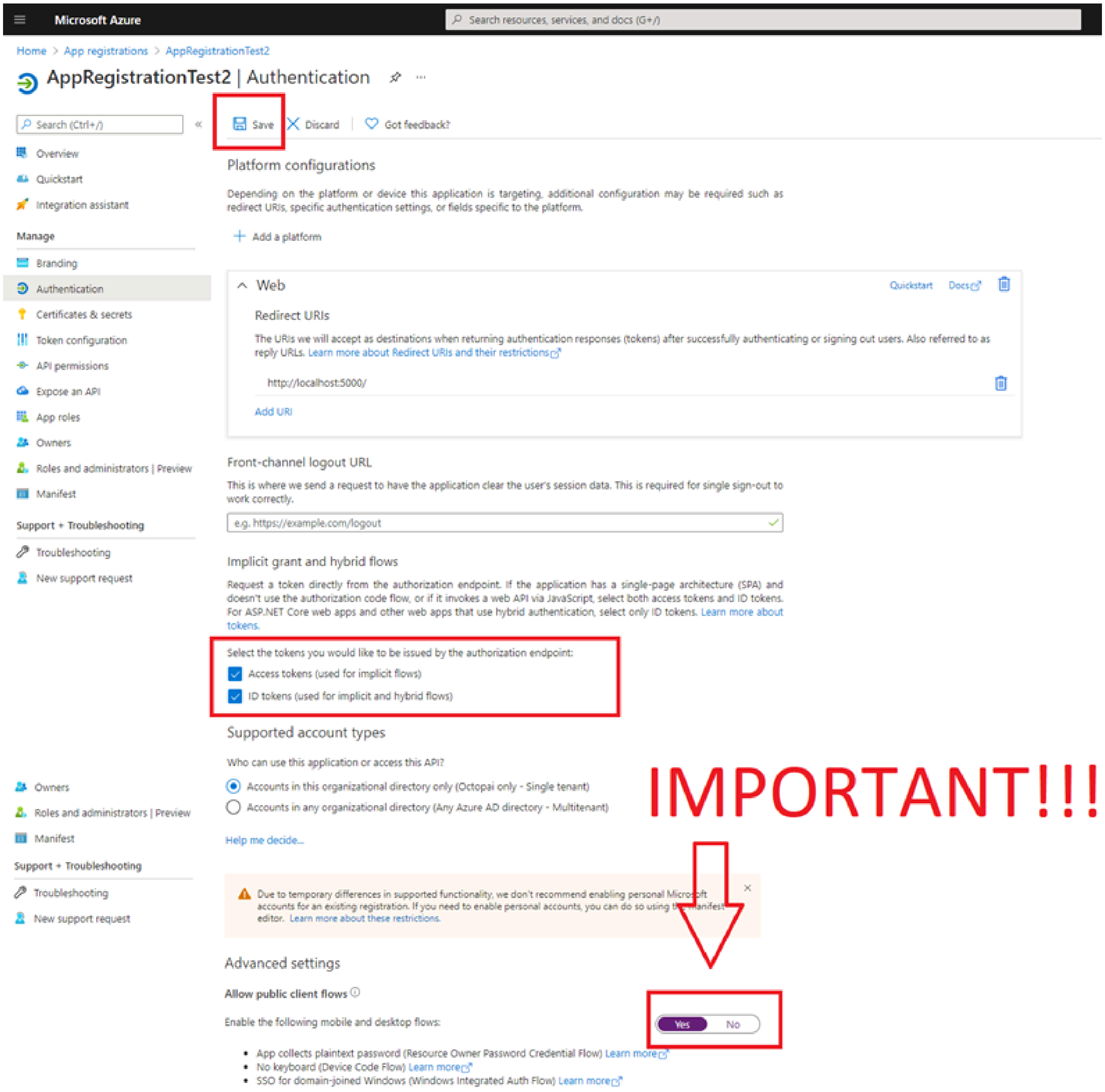
We'll return the authentication response to this URI after successfully authenticating the user. Providing this now is optional and it can be changed later, but a value is required for most authentication scenarios.

Web | http://localhost:5000/ ✓

Register an app you're working on here. Integrate gallery apps and other apps from outside your organization by adding from Enterprise applications.

By proceeding, you agree to the Microsoft Platform Policies

Go to Authentication tab > Configure as following > Click Save



Platform configurations

Depending on the platform or device this application is targeting, additional configuration may be required such as redirect URIs, specific authentication settings, or fields specific to the platform.

+ Add a platform

Web   Quickstart   Docs

Redirect URIs

The URIs we will accept as destinations when returning authentication responses (tokens) after successfully authenticating or signing out users. Also referred to as reply URLs. Learn more about Redirect URIs and their restrictions

http://localhost:5000/

Add URI

Front-channel logout URL

This is where we send a request to have the application clear the user's session data. This is required for single sign-out to work correctly.

e.g. https://example.com/logout

Implicit grant and hybrid flows

Request a token directly from the authorization endpoint. If the application has a single-page architecture (SPA) and doesn't use the authorization code flow, or if it invokes a web API via JavaScript, select both access tokens and ID tokens. For ASP.NET Core web apps and other web apps that use hybrid authentication, select only ID tokens. Learn more about tokens.

Select the tokens you would like to be issued by the authorization endpoint:

☑ Access tokens (used for implicit flows)

☑ ID tokens (used for implicit and hybrid flows)

Supported account types

Who can use this application or access this API?

● Accounts in this organizational directory only (Octopai only - Single tenant)

○ Accounts in any organizational directory (Any Azure AD directory - Multitenant)

Help me decide...

Due to temporary differences in supported functionality, we don't recommend enabling personal Microsoft accounts for an existing registration. If you need to enable personal accounts, you can do so using the manifest editor. Learn more about these restrictions.

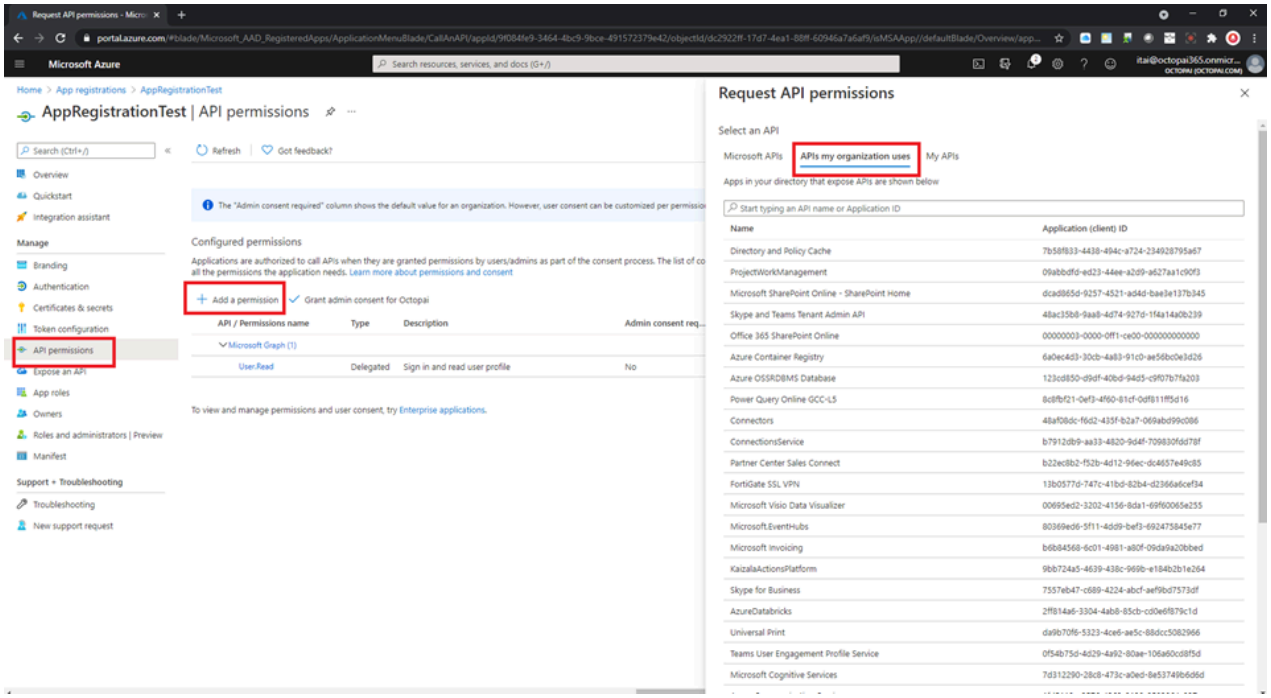
Advanced settings

Allow public client flows

Enable the following mobile and desktop flows: [✓ Yes] No

- App collects plaintext password (Resource Owner Password Credential Flow) Learn more
- No keyboard (Device Code Flow) Learn more
- SSO for domain-joined Windows (Windows Integrated Auth Flow) Learn more

Click the API Permissions tab > Click Add permission > Click APIs my organization uses.


Request API permissions - Micro
portal.azure.com/#blade/Microsoft_AAD_RegisteredApps/ApplicationMenuBlade/CallAnAPI/appId/9f084fe9-3464-4bc9-9bce-491572379e42/objectId/dc2922ff-17d7-4ea1-88ff-60946a7a6af9/isMSAApp//defaultBlade/Overview/app...
Microsoft Azure
Search resources, services, and docs (G+/)
Home > App registrations > AppRegistrationTest
AppRegistrationTest | API permissions
Search (Ctrl+/)
Overview
Quickstart
Integration assistant
Manage
Branding
Authentication
Certificates & secrets
Token configuration
API permissions
Expose an API
App roles
Owners
Roles and administrators | Preview
Manifest
Support + Troubleshooting
Troubleshooting
New support request
Refresh
Got feedback?
+ Add a permission
✓ Grant admin consent for Octopai
Request API permissions
Select an API
Microsoft APIs
APIs my organization uses
My APIs
Start typing an API name or Application ID

The "Admin consent required" column shows the default value for an organization. However, user consent can be customized per permissio

Configured permissions

Applications are authorized to call APIs when they are granted permissions by users/admins as part of the consent process. The list of co all the permissions the application needs. Learn more about permissions and consent

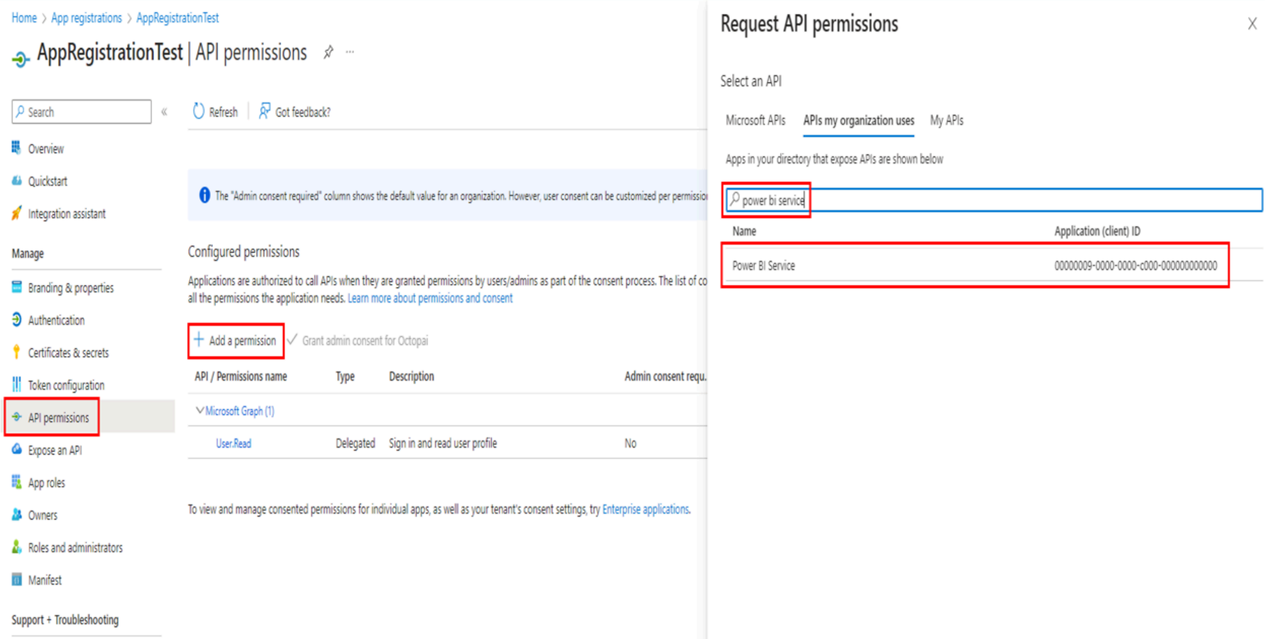
API / Permissions name	Type	Description	Admin consent req...
Microsoft Graph (1)			
User.Read	Delegated	Sign in and read user profile	No

To view and manage permissions and user consent, try Enterprise applications.

Apps in your directory that expose APIs are shown below

Name	Application (client) ID
Directory and Policy Cache	7b58f833-4438-494c-a724-234928795a67
ProjectWorkManagement	09abbdfd-ed23-44ee-a2d9-a627aa1c90f3
Microsoft SharePoint Online - SharePoint Home	dcad865d-9257-4521-ad4d-bae3e137b345
Skype and Teams Tenant Admin API	48ac35b8-9aa8-4d74-927d-1f4a14a0b239
Office 365 SharePoint Online	00000003-0000-0ff1-ce00-000000000000
Azure Container Registry	6a0ec4d3-30cb-4a83-91c0-ae56bc0e3d26
Azure OSSRDBMS Database	123cd850-d9df-40bd-94d5-c9f07b7fa203
Power Query Online GCC-L5	8c8fbf21-0ef3-4f60-81cf-0df811ff5d16
Connectors	48af08dc-f6d2-435f-b2a7-069abd99c086
ConnectionsService	b7912db9-aa33-4820-9d4f-709830fdd78f
Partner Center Sales Connect	b22ec8b2-f52b-4d12-96ec-dc4657e49c85
FortiGate SSL VPN	13b0577d-747c-41bd-82b4-d2366a6cef34
Microsoft Visio Data Visualizer	00695ed2-3202-4156-8da1-69f60065e255
Microsoft.EventHubs	80369ed6-5f11-4dd9-bef3-692475845e77
Microsoft Invoicing	b6b84568-6c01-4981-a80f-09da9a20bbed
KaizalaActionsPlatform	9bb724a5-4639-438c-969b-e184b2b1e264
Skype for Business	7557eb47-c689-4224-abcf-aef9bd7573df
AzureDatabricks	2ff814a6-3304-4ab8-85cb-cd0e6f879c1d
Universal Print	da9b70f6-5323-4ce6-ae5c-88dcc5082966
Teams User Engagement Profile Service	0f54b75d-4d29-4a92-80ae-106a60cd8f5d
Microsoft Cognitive Services	7d312290-28c8-473c-a0ed-8e53749b6d6d

Search for Power bi Service > Click on Power bi Service.



The "Admin consent required" column shows the default value for an organization. However, user consent can be customized per permissio

Configured permissions

Applications are authorized to call APIs when they are granted permissions by users/admins as part of the consent process. The list of co all the permissions the application needs. Learn more about permissions and consent

API / Permissions name	Type	Description	Admin consent requ.
Microsoft Graph (1)			
User.Read	Delegated	Sign in and read user profile	No

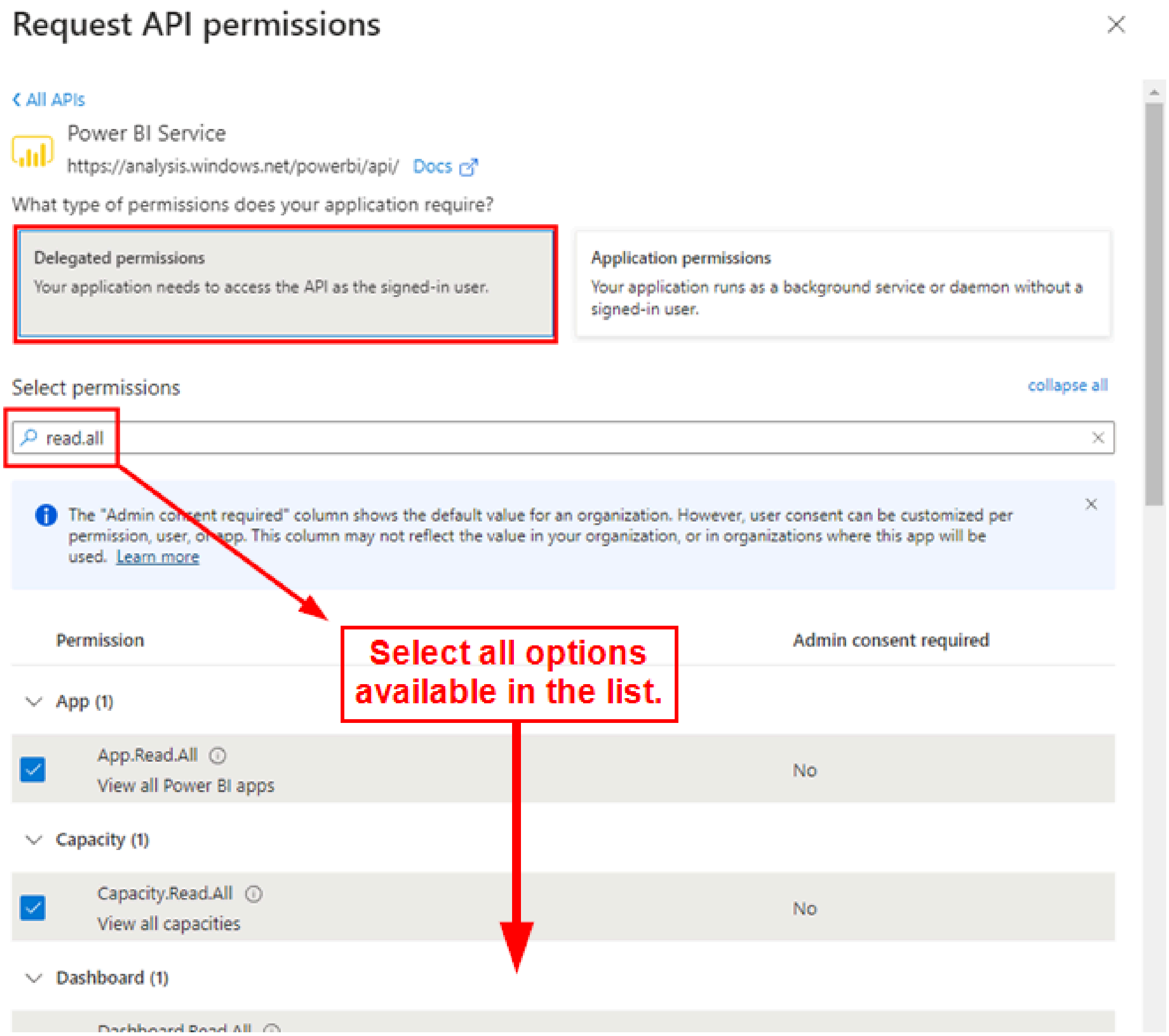
To view and manage consented permissions for individual apps, as well as your tenant's consent settings, try Enterprise applications.

Apps in your directory that expose APIs are shown below

power bi service

Name	Application (client) ID
Power BI Service	00000009-0000-0000-c000-000000000000

Click Delegated permissions > Search "Read.All" and check ALL related permissions > click Add permission.



Request API permissions

Power BI Service

https://analysis.windows.net/powerbi/api/ Docs

What type of permissions does your application require?

Delegated permissions
Your application needs to access the API as the signed-in user.

Application permissions
Your application runs as a background service or daemon without a signed-in user.

Select permissions

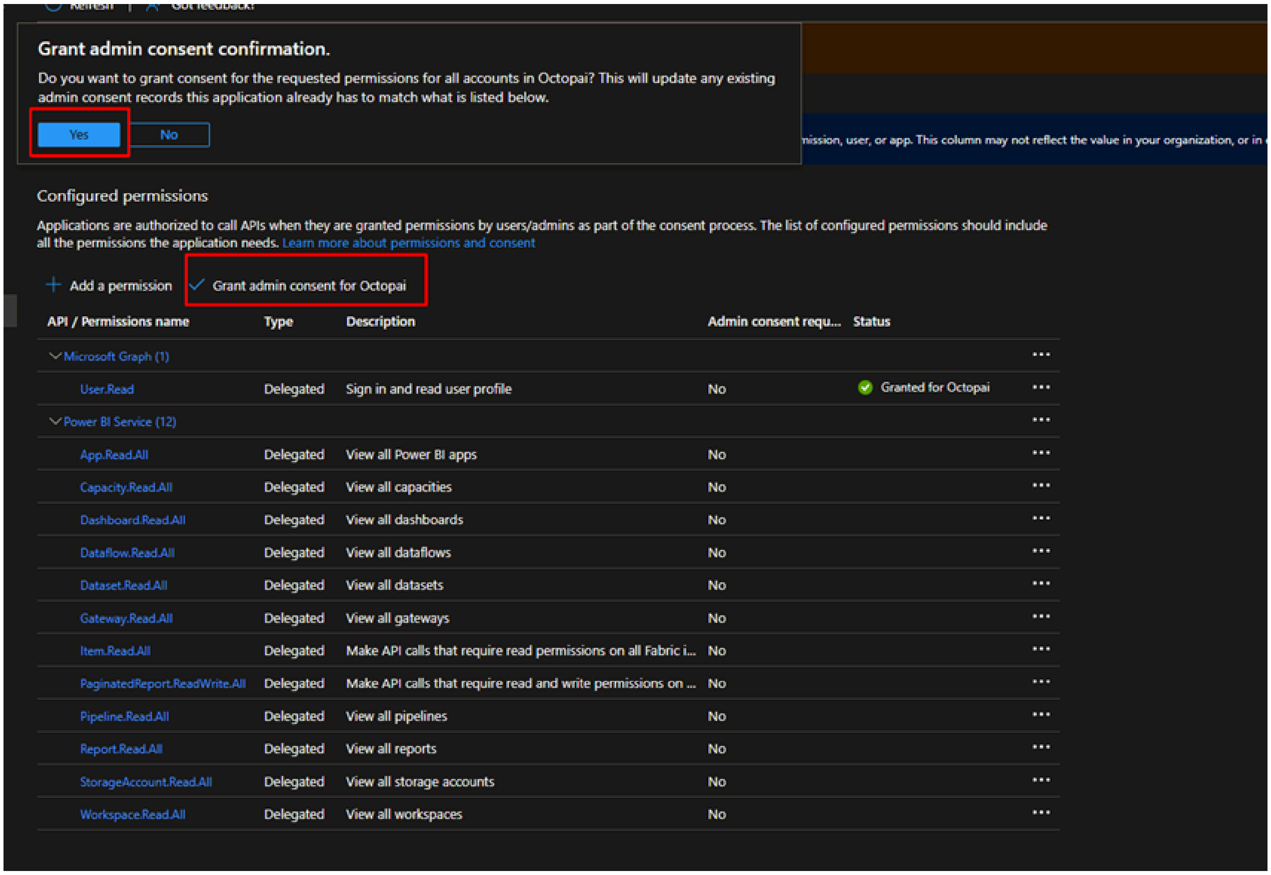
read.all

The "Admin consent required" column shows the default value for an organization. However, user consent can be customized per permission, user, or app. This column may not reflect the value in your organization, or in organizations where this app will be used. Learn more

Select all options available in the list.

	Permission	Admin consent required
	App (1)	
☑	App.Read.All View all Power BI apps	No
	Capacity (1)	
☑	Capacity.Read.All View all capacities	No
	Dashboard (1)	
	Dashboard.Read.All	

Click Grant admin consent for <Your domain name> > Click Yes



Grant admin consent confirmation.

Do you want to grant consent for the requested permissions for all accounts in Octopai? This will update any existing admin consent records this application already has to match what is listed below.

mission, user, or app. This column may not reflect the value in your organization, or in

Configured permissions

Applications are authorized to call APIs when they are granted permissions by users/admins as part of the consent process. The list of configured permissions should include all the permissions the application needs. Learn more about permissions and consent

API / Permissions name	Type	Description	Admin consent requ...	Status
Microsoft Graph (1)				
User.Read	Delegated	Sign in and read user profile	No	Granted for Octopai
Power BI Service (12)				
App.Read.All	Delegated	View all Power BI apps	No	
Capacity.Read.All	Delegated	View all capacities	No	
Dashboard.Read.All	Delegated	View all dashboards	No	
Dataflow.Read.All	Delegated	View all dataflows	No	
Dataset.Read.All	Delegated	View all datasets	No	
Gateway.Read.All	Delegated	View all gateways	No	
Item.Read.All	Delegated	Make API calls that require read permissions on all Fabric i...	No	
PaginatedReport.ReadWrite.All	Delegated	Make API calls that require read and write permissions on ...	No	
Pipeline.Read.All	Delegated	View all pipelines	No	
Report.Read.All	Delegated	View all reports	No	
StorageAccount.Read.All	Delegated	View all storage accounts	No	
Workspace.Read.All	Delegated	View all workspaces	No	

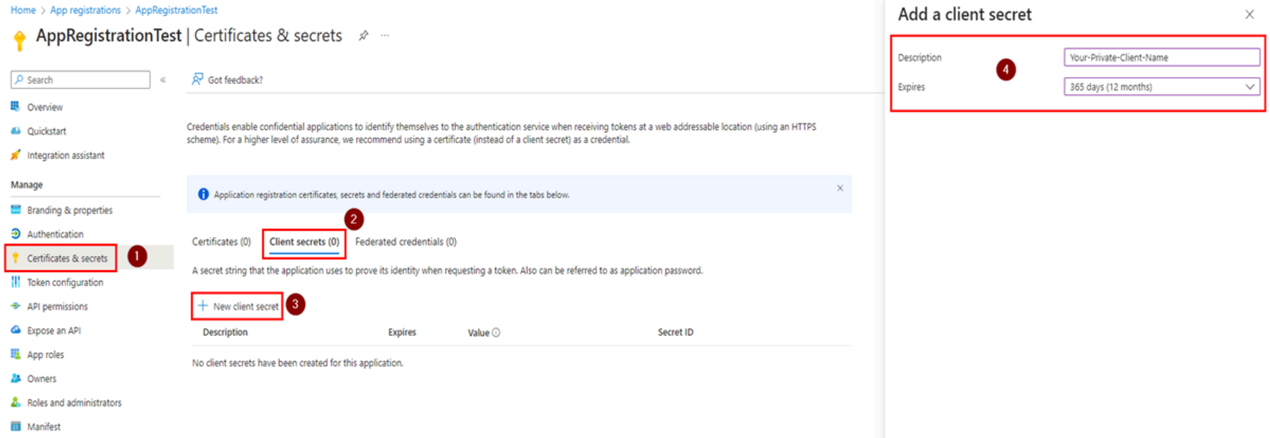
That summarizes the guidelines for creating the AppRegistration and granting it the proper permissions.

Security Group and PowerBI Portal configuration setup

Important clarification ahead of this guide:

- Please make sure you completed **Permissions and configurations setup in Microsoft Azure**.
- The workspace license mode must be Premium Per User.

In the Azure portal, go to Certificates & secrets > New client secret.



Credentials enable confidential applications to identify themselves to the authentication service when receiving tokens at a web addressable location (using an HTTPS scheme). For a higher level of assurance, we recommend using a certificate (instead of a client secret) as a credential.

Application registration certificates, secrets and federated credentials can be found in the tabs below.

A secret string that the application uses to prove its identity when requesting a token. Also can be referred to as application password.

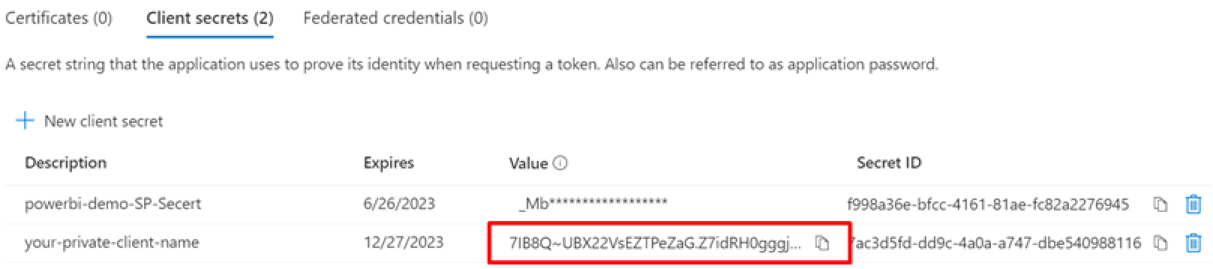
Description	Expires	Value	Secret ID

No client secrets have been created for this application.

Add a client secret

Description	Your-Private-Client-Name
Expires	365 days (12 months)

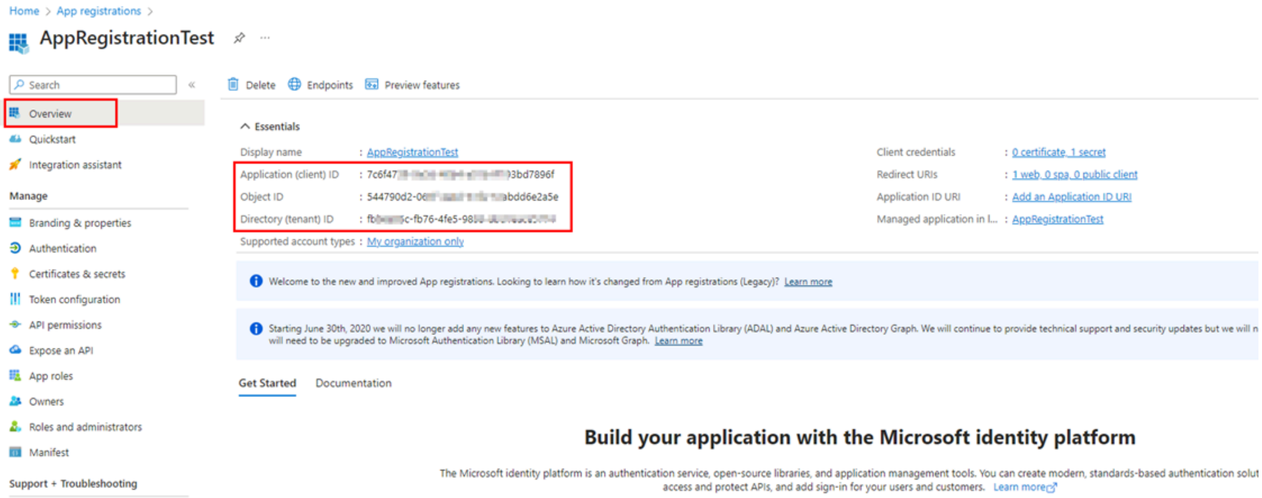
Please copy the Secret value, we will need it later for the OC authentication.



A secret string that the application uses to prove its identity when requesting a token. Also can be referred to as application password.

Description	Expires	Value ⓘ	Secret ID
powerbi-demo-SP-Secert	6/26/2023	_Mb******************	f998a36e-bfcc-4161-81ae-fc82a2276945
your-private-client-name	12/27/2023	7IB8Q~UBX22VsEZTPeZaG.Z7idRH0gggj...	7ac3d5fd-dd9c-4a0a-a747-dbe540988116

In the Overview tab, we will see the relevant information for the Cloudera Octopai Client.



Display name : AppRegistrationTest
Application (client) ID : 7c6f47[redacted]3bd7896f
Object ID : 544790d2-0[redacted]abdd6e2a5e
Directory (tenant) ID : fb[redacted]c-fb76-4fe5-98[redacted]
Supported account types : My organization only

Client credentials : 0 certificate, 1 secret
Redirect URIs : 1 web, 0 spa, 0 public client
Application ID URI : Add an Application ID URI
Managed application in l... : AppRegistrationTest

Welcome to the new and improved App registrations. Looking to learn how it's changed from App registrations (Legacy)? Learn more

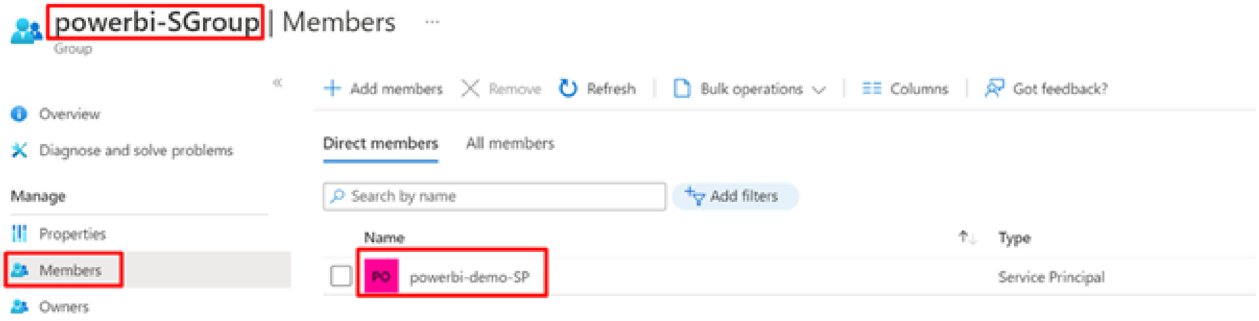
Starting June 30th, 2020 we will no longer add any new features to Azure Active Directory Authentication Library (ADAL) and Azure Active Directory Graph. We will continue to provide technical support and security updates but we will n will need to be upgraded to Microsoft Authentication Library (MSAL) and Microsoft Graph. Learn more

Build your application with the Microsoft identity platform

The Microsoft identity platform is an authentication service, open-source libraries, and application management tools. You can create modern, standards-based authentication solut access and protect APIs, and add sign-in for your users and customers. Learn more

Create Security Group: We will need to create a group in https://portal.azure.com/ under groups OR ADD your Service Principal to an existing group.

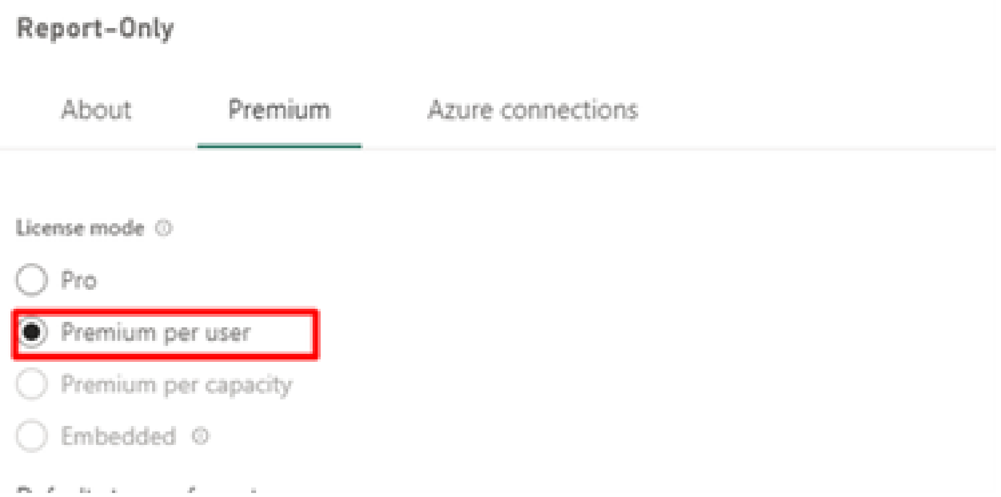
(App Registration name = Principal service name)



Name	Type
PO powerbi-demo-SP	Service Principal

Power BI Settings (https://app.powerbi.com/home):

Required license type: Workspace License mode must be Premium Per User.



Report-Only

About Premium Azure connections

License mode

○ Pro

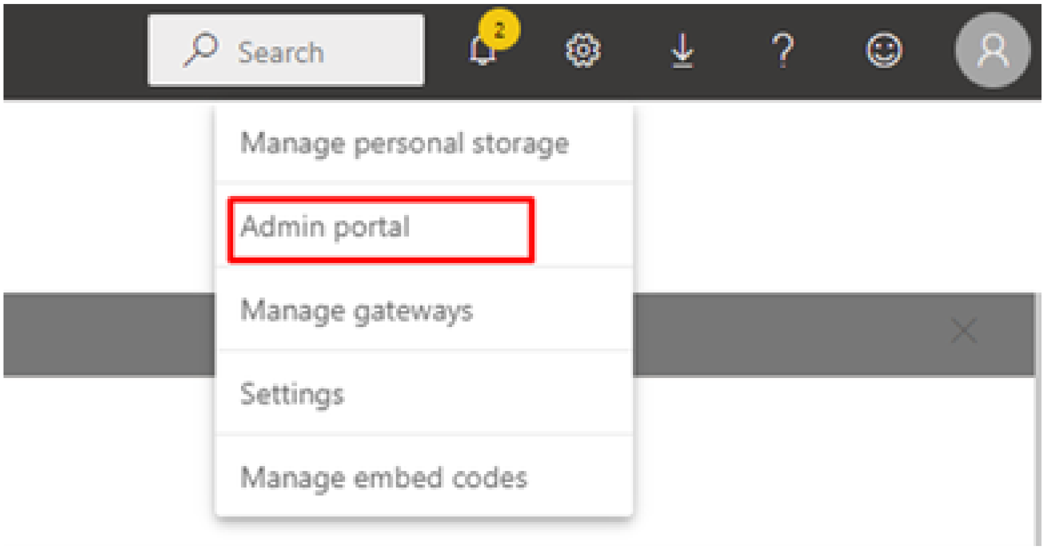
● Premium per user

○ Premium per capacity

○ Embedded

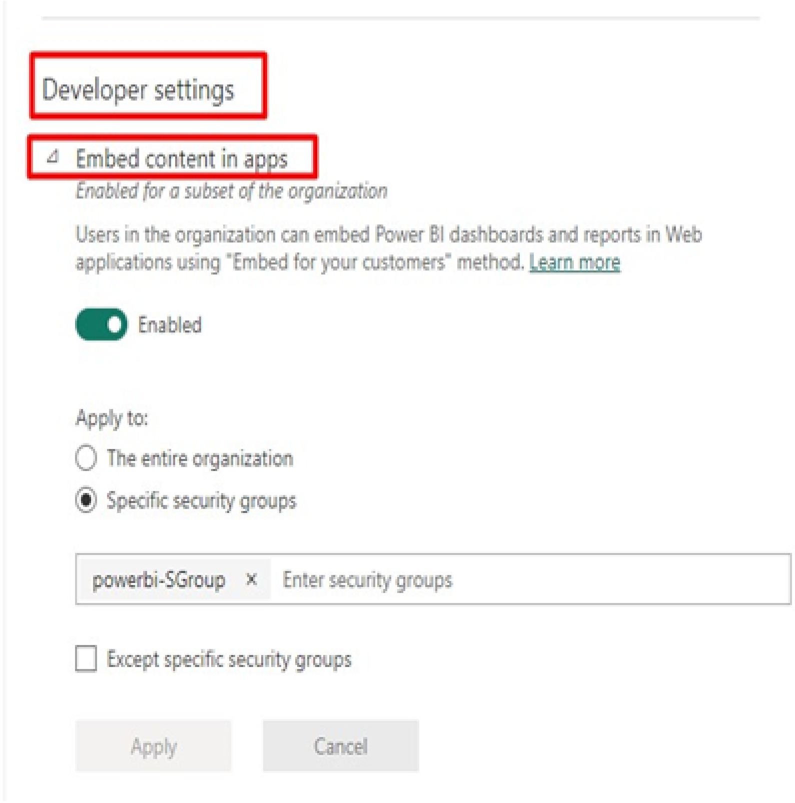
Admin Tenant settings: https://app.powerbi.com/home

Go to Settings - > Admin portal # Tenant settings.



Note: Note: Changing the following steps takes approximately 15 min to be refreshed.

We need to enable the following: Under Tenant settings go to Developer Settings. Enable “Embed content in apps” and “Allow service principals to use Power BI APIs”. Choose the security group that the Service Principal is in.



Developer settings

◿ Embed content in apps

Enabled for a subset of the organization

Users in the organization can embed Power BI dashboards and reports in Web applications using "Embed for your customers" method. Learn more

Enabled

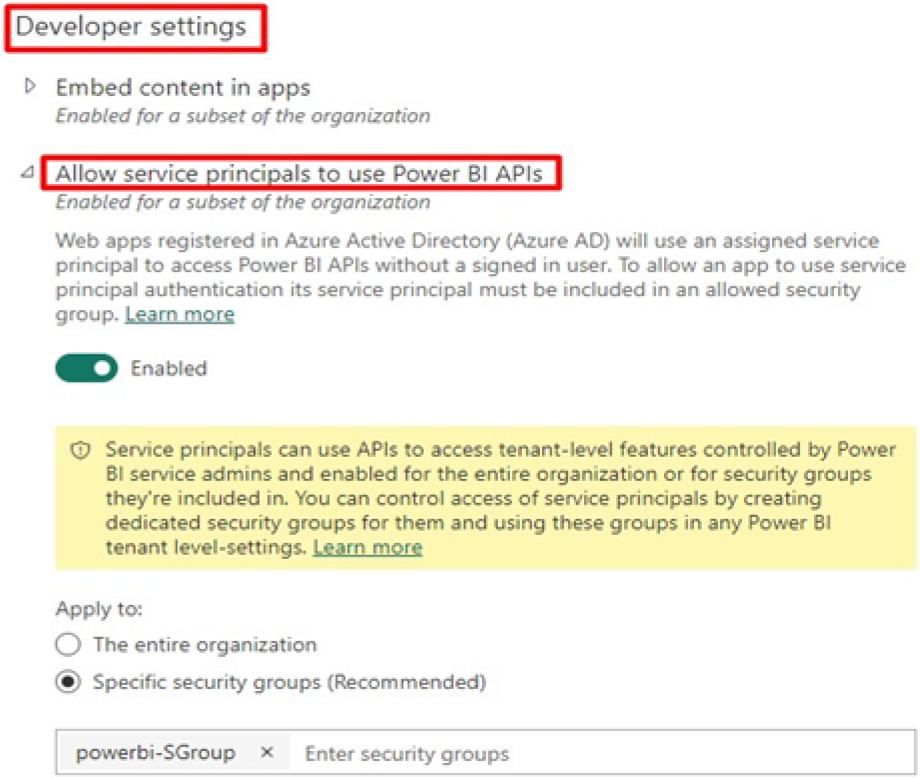
Apply to:

○ The entire organization

● Specific security groups

powerbi-SGroup × Enter security groups

□ Except specific security groups



Developer settings

▷ Embed content in apps

Enabled for a subset of the organization

◿ Allow service principals to use Power BI APIs

Enabled for a subset of the organization

Web apps registered in Azure Active Directory (Azure AD) will use an assigned service principal to access Power BI APIs without a signed in user. To allow an app to use service principal authentication its service principal must be included in an allowed security group. Learn more

Enabled

Service principals can use APIs to access tenant-level features controlled by Power BI service admins and enabled for the entire organization or for security groups they're included in. You can control access of service principals by creating dedicated security groups for them and using these groups in any Power BI tenant level-settings. Learn more

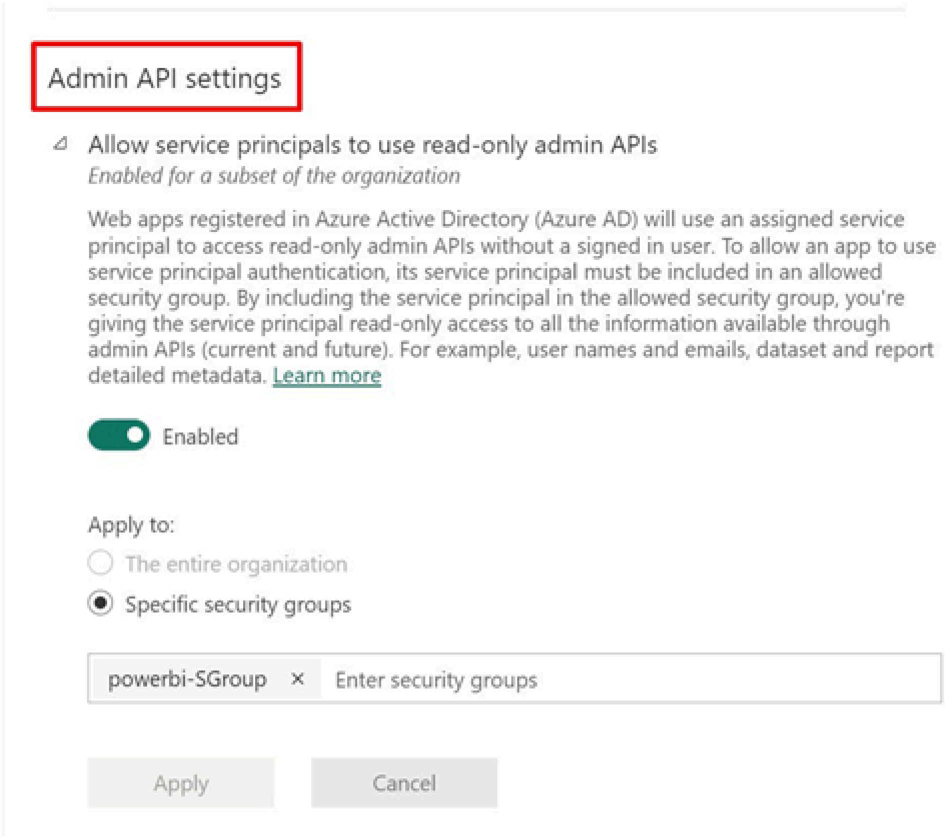
Apply to:

○ The entire organization

● Specific security groups (Recommended)

powerbi-SGroup × Enter security groups

Under Admin API settings enable and add the specific group that the Service Principal is in.



Admin API settings

◿ Allow service principals to use read-only admin APIs

Enabled for a subset of the organization

Web apps registered in Azure Active Directory (Azure AD) will use an assigned service principal to access read-only admin APIs without a signed in user. To allow an app to use service principal authentication, its service principal must be included in an allowed security group. By including the service principal in the allowed security group, you're giving the service principal read-only access to all the information available through admin APIs (current and future). For example, user names and emails, dataset and report detailed metadata. Learn more

Enabled

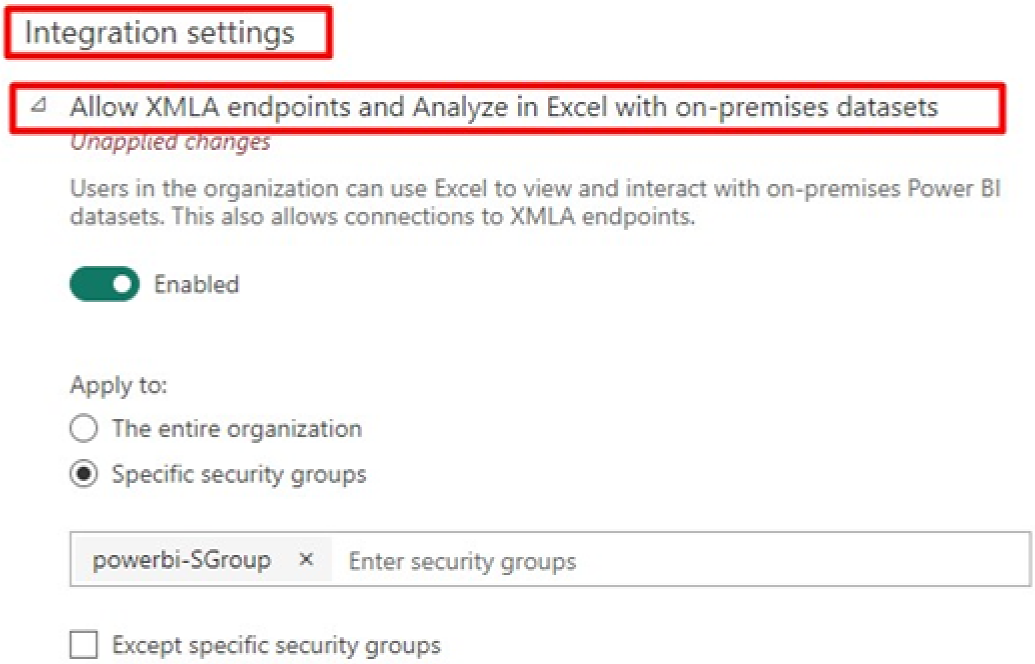
Apply to:

○ The entire organization

● Specific security groups

powerbi-SGroup × Enter security groups

Under Integration settings enable and add the specific group that the Service Principal is in.



Integration settings

◿ Allow XMLA endpoints and Analyze in Excel with on-premises datasets

Unapplied changes

Users in the organization can use Excel to view and interact with on-premises Power BI datasets. This also allows connections to XMLA endpoints.

Enabled

Apply to:

○ The entire organization

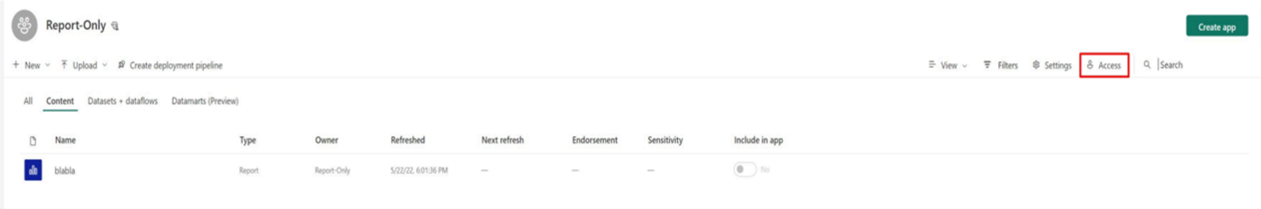
● Specific security groups

powerbi-SGroup × Enter security groups

□ Except specific security groups

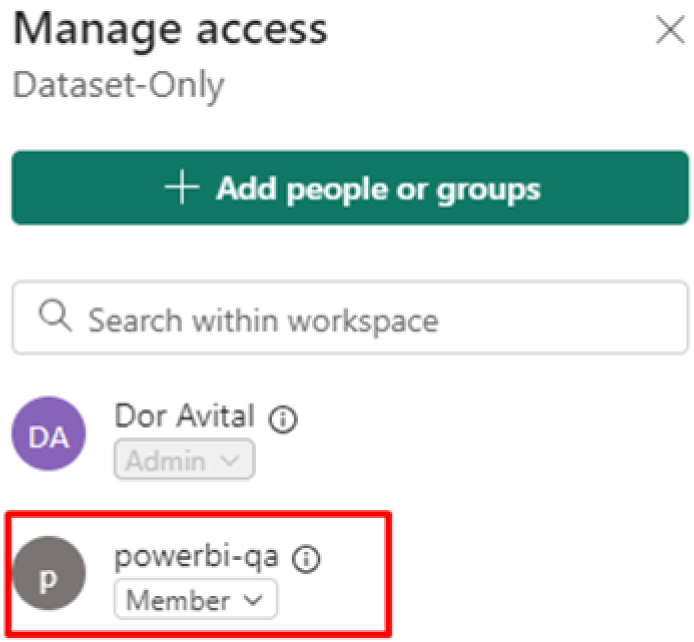
Workspace settings:

Go to your desired workspace and click Access.



Name	Type	Owner	Refreshed	Next refresh	Endorsement	Sensitivity	Include in app
blabla	Report	Report-Only	5/22/22, 6:01:36 PM	—	—	—	No

Add the security group you’ve created or the Service Principal to the workspace and give it a Member role.



Dor Avital: Admin

powerbi-qa: Member

Note: Changes in Azure and PowerBI portals typically take 15-30 minutes to apply but may take up to 24 hours in some cases.

If you are experiencing trouble with extracting your metadata using the Cloudera Octopai Client, please ensure you have followed all steps in this guide and wait for Microsoft to apply the new configurations across your organization.

Power BI Report Server

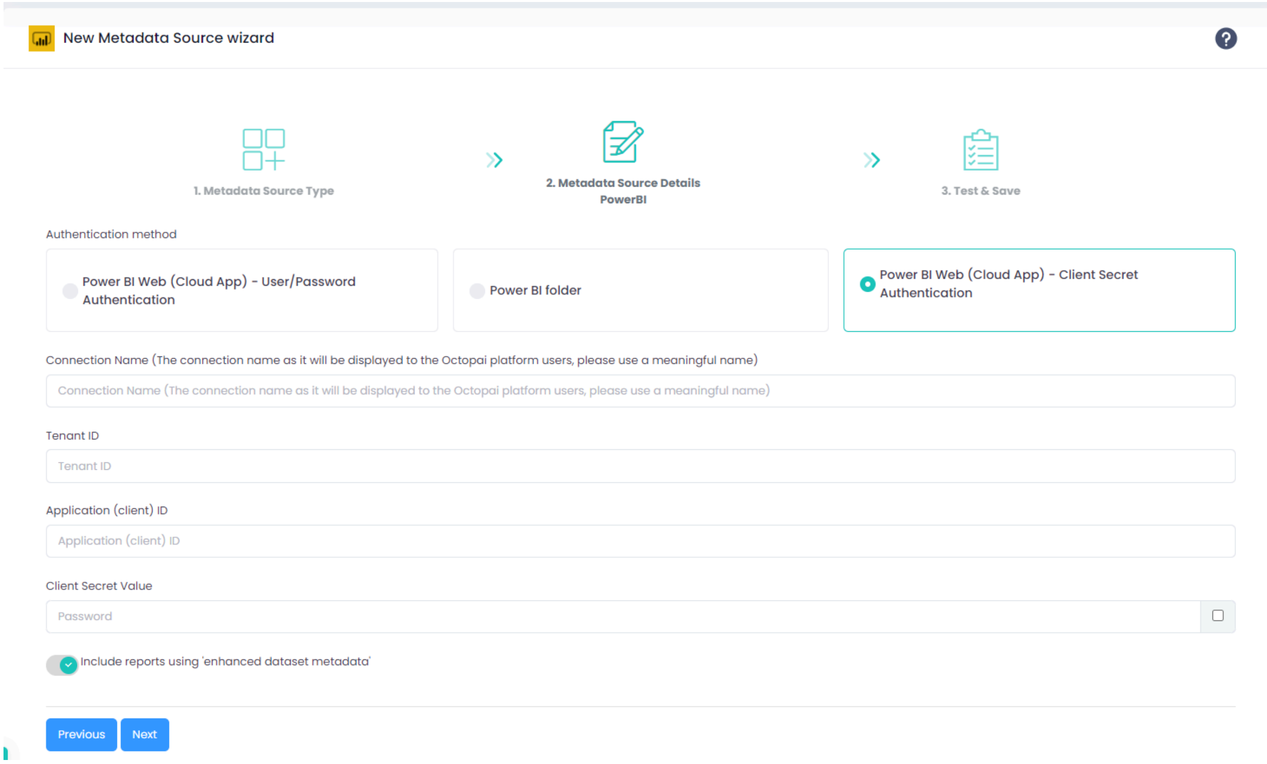
Please assign your (application) user with ‘Browse’ and 'Content Manager' roles for reports you would like to see in Cloudera Octopai.

For the Cloudera Octopai Client connector, use the following URL structure:

http:\\comapny-url.com OR http:\\<IP>

Setting up Power BI Metadata Source

Metadata Sources are set on the Cloudera Octopai Client



New Metadata Source wizard

Authentication method

○ Power BI Web (Cloud App) - User/Password Authentication ○ Power BI folder ● Power BI Web (Cloud App) - Client Secret Authentication

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Tenant ID

Application (client) ID

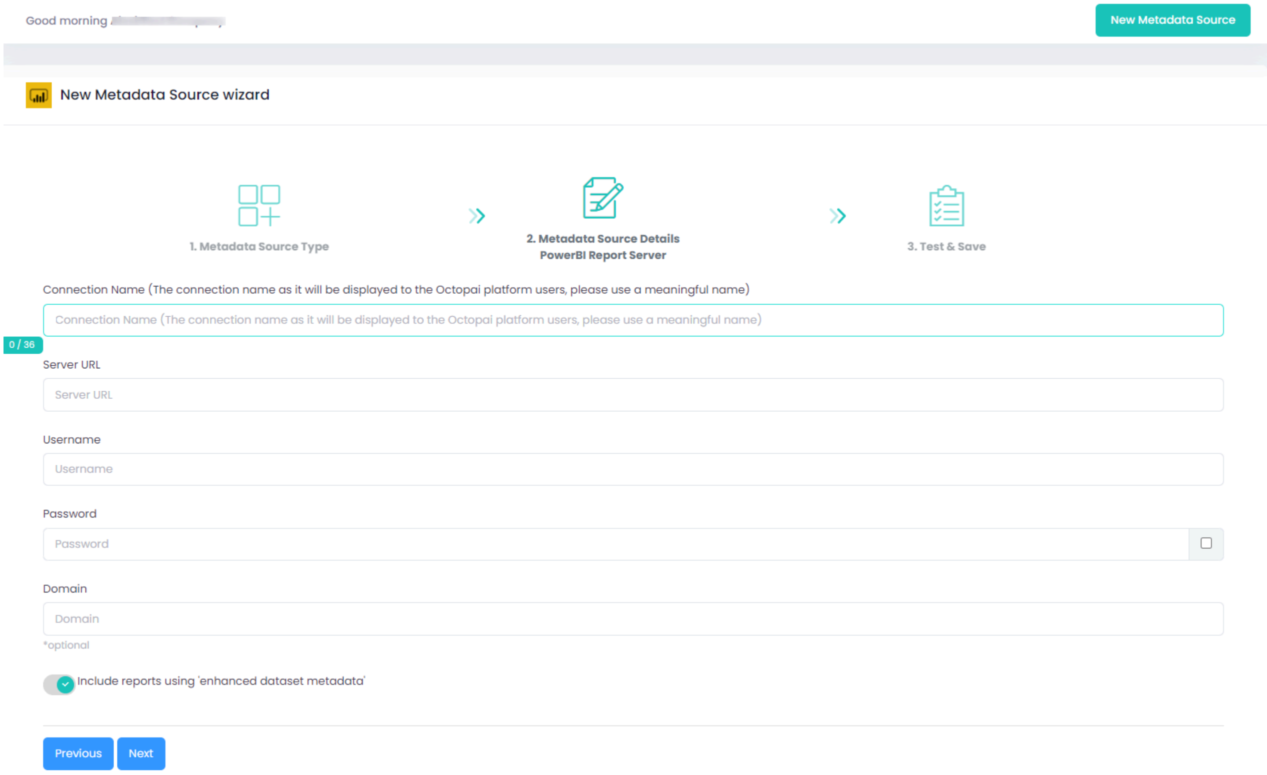
Client Secret Value

Password □

Include reports using 'enhanced dataset metadata'

1. Connection Name - Give a meaningful name, as it will be displayed to the Cloudera Octopai platform users.
2. Tenant ID - Available in the 'App registrations' section under the application you created.
3. Application (Client) ID - Available in the 'App registrations' section under the application you created.
4. Client Secret - Generated in 'App registrations > Certificates & secrets'.

Figure 37: Power BI Report Server



Good morning [redacted]

New Metadata Source wizard

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

0 / 36

Server URL

Username

Password

Password □

Domain

*optional

Include reports using 'enhanced dataset metadata'

1. **Connection Name** - Give a meaningful name, as it will be displayed to the Cloudera Octopai platform users.
2. **Server URL** - The Report Server's URL.
3. **Username** - The user that is granted the needed roles.
4. **Password** - The user's password.
5. **Domain (optional)** - Only if the user is in a domain.

DBT Files

Set up DBT project files as a textual metadata source in Cloudera Octopai Data Lineage and review the permissions required for ingestion.

Tool Permissions Prerequisites

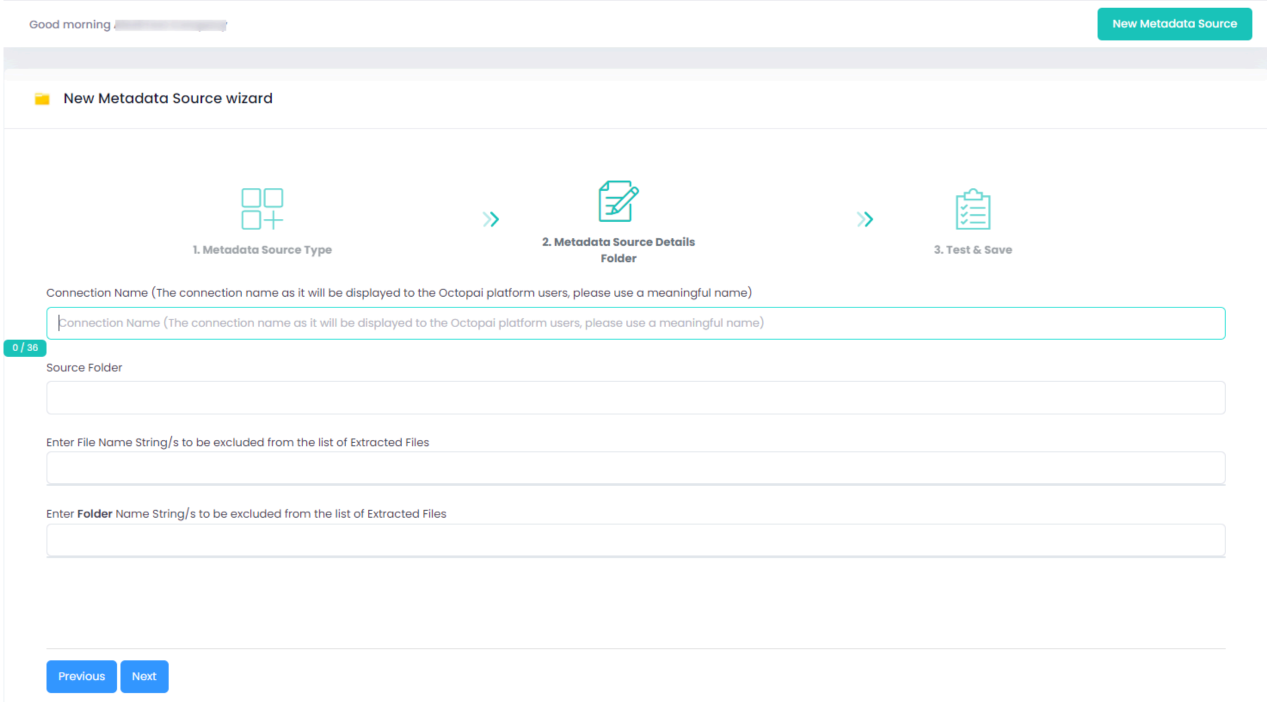
⚠ **Warning:** Missing permissions could end up in broken lineages.

- Enable read permission for Cloudera Octopai Windows NT User to the following path: \run\<projectName>\models
- The above folder should contain the DBT *.sql files.

Setting up Textual Files Metadata Source

Metadata Sources are set on the Cloudera Octopai Client

Please use the Textual Files metadata source and fill in the \run<projectName>\models*.sql path on the source folder attribute



Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Source Folder

Enter File Name String/s to be excluded from the list of Extracted Files

Enter **Folder** Name String/s to be excluded from the list of Extracted Files

Universal Connector

Instructions for configuring Cloudera Octopai Universal Connector metadata source and required permissions.

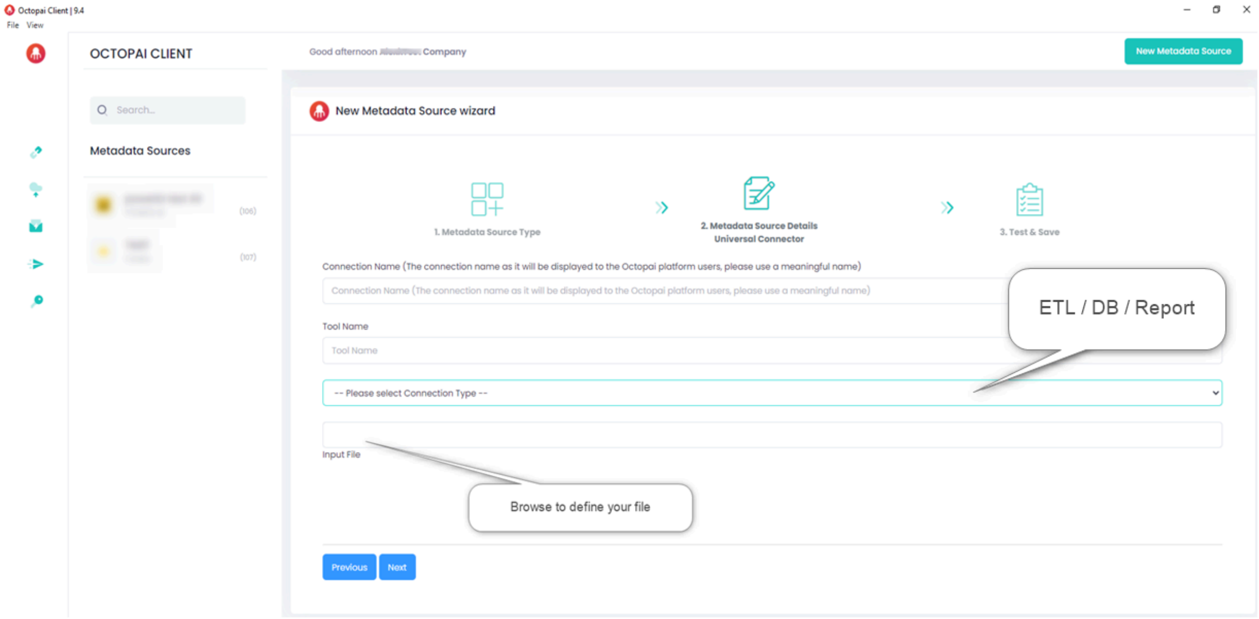
Tool Permissions Prerequisites

Warning: Missing permissions could end up in broken lineages.

Enable read permission for Cloudera Octopai Windows NT User to the folder (which contains the filled template files).

Setting up Universal Connector Metadata Source

Metadata Sources are set on the Cloudera Octopai Client



Good afternoon [redacted] Company

New Metadata Source wizard

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Tool Name

Tool Name

-- Please select Connection Type --

Input File

ETL / DB / Report

Browse to define your file

Related Information

Enhancing Data Connectivity: Cloudera Octopai Universal Connector for Databases & ETLs Tools Guide

Enhancing Data Connectivity: Cloudera Octopai Universal Connector for Reporting Tools Guide

Textual Files / Folder connector

Set up Cloudera Octopai Textual Files connector by configuring metadata sources and validating permissions.

Supported file types

The following file types are supported:

- Upload (Discovery Display): text file, CSV file, SQL file, log file, XML file, JSON file, bash file, output file, BTEQ file, BTQ file, control file, report file, Perl file, KornShell file, C# file, Siddhi file, Python file
- Fully Analyzed (All Modules Supported): SQL file, BTEQ file, BTQ file, control file, Siddhi file, Python file

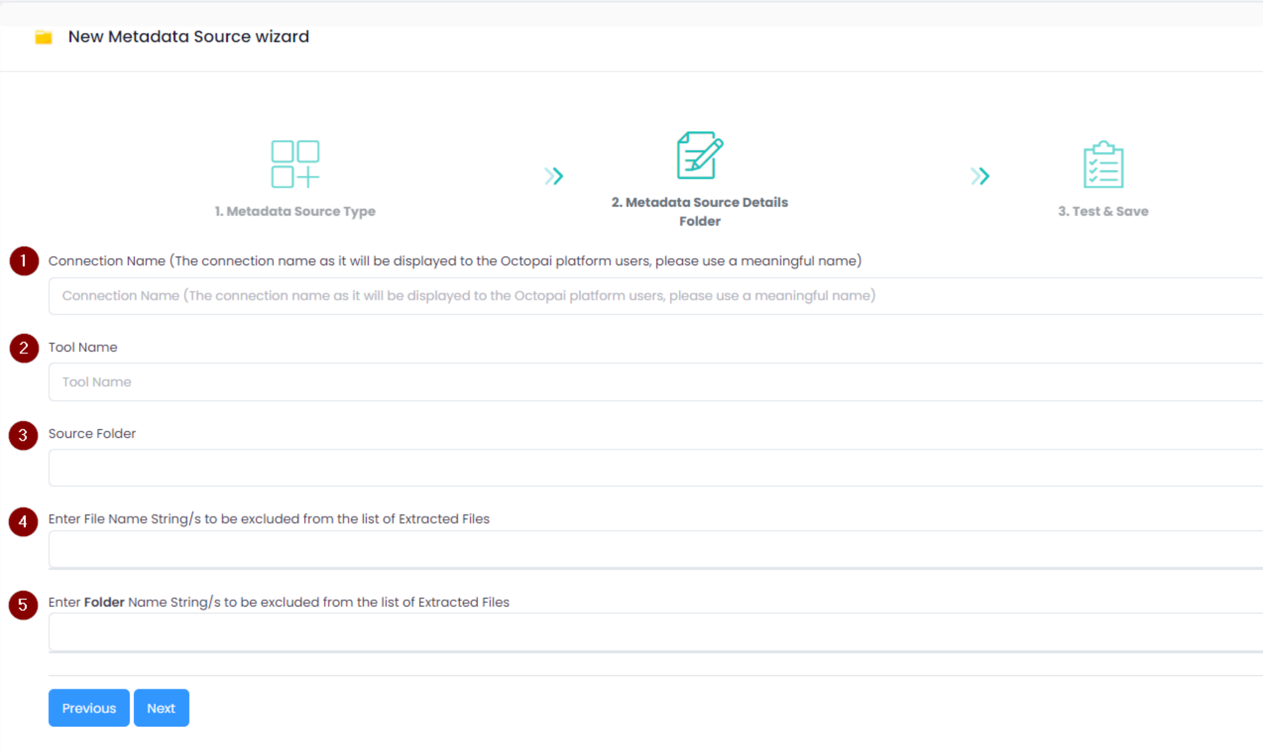
Tool Permissions Prerequisites

⚠ **Warning:** Missing permissions could end up in broken lineages.

Enable read permission for Cloudera Octopai Windows NT User to the folder (which contains the textual files).

Setting up Textual Files Metadata Source

Metadata Sources are set on the Cloudera Octopai Client



New Metadata Source wizard

1 Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

Connection Name (The connection name as it will be displayed to the Octopai platform users, please use a meaningful name)

2 Tool Name

Tool Name

3 Source Folder

4 Enter File Name String/s to be excluded from the list of Extracted Files

5 Enter **Folder** Name String/s to be excluded from the list of Extracted Files

Legend:

1. Connection Name: Provide a meaningful name for this connection to help you easily identify it later in the Cloudera Octopai application.
2. Tool Name: Specify the tool you are using. For example, enter "Python" if you are uploading Python scripts.
3. Source Folder: Click the blank field to select the folder containing the files you want to upload.
4. Exclude Files or Strings: Enter file names, suffixes, strings, or patterns you want to exclude from the export.
5. Exclude Folders: Specify folder names to exclude from the export if you have multiple subfolders.

Enhancing Data Connectivity: Cloudera Octopai Universal Connector for Databases & ETLs Tools Guide

The Cloudera Octopai Universal Connector for Databases & ETLs Tools integrate metadata from diverse systems into the Data Intelligence Platform, enabling lineage, data discovery, and full visibility of your data ecosystem.

As data demands evolve, data teams continuously seek a better understanding of their data ecosystem. The need for analysis and visualization of additional systems is growing. As a result, Cloudera Octopai is consistently expanding its extensive coverage of out-of-the-box supported technologies in our Data Intelligence Platform.

However, as your needs progress, it is crucial to provide an overview of the complete data landscape with various systems and data flows.

New data systems often lack automation support, and many organizations rely on custom-built data processes. A lineage tool must cover these processes to deliver a complete and accurate picture.

Therefore, Cloudera Octopai has developed the Universal Connector, empowering you to add your metadata from these types of systems into Cloudera Octopai's Data Intelligence platform to get the full picture - complete lineage, data discovery and a data catalog.

You get unlimited ingestion capabilities to enrich the platform with additional lineage, allowing you to add the final piece of the puzzle and get full visibility of your data ecosystem.

This flexibility allows you to adapt quickly to your changing data landscape, and consistently get a complete view regardless of what data systems you're using.

How it is done

Use the Cloudera Octopai templates below to ingest your metadata into the platform. The rest is fully automated.

What Cloudera Octopai offers

This metadata, along with the metadata automatically ingested from out-of-the-box supported systems, is analyzed using machine learning. In turn Cloudera Octopai provides you with end-to-end column-level lineage, inner system lineage, cross system lineage, data discovery and a data catalog of your entire data landscape accessible to all data users in the organization.

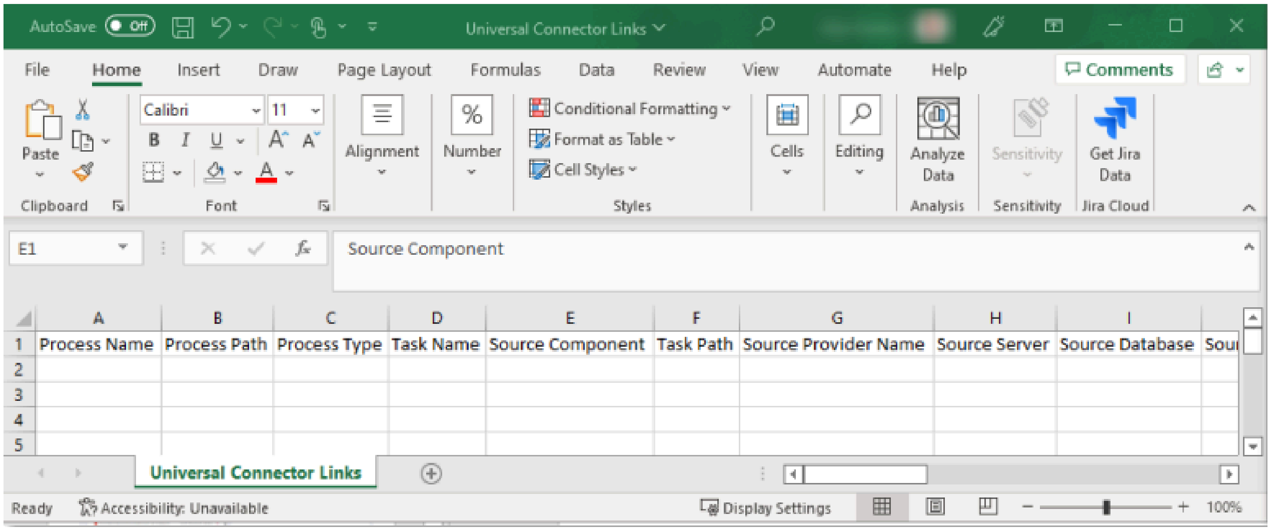
The benefits:

- No blind spots – perform changes with confidence.
- Get a clear picture of data transformations.
- Increase visibility of the organization's complete data ecosystem.
- Future-proof your expanding data landscape by providing access to unlimited data systems.
- Add links to our out-of-the-box technologies.

How to use the template files

1. Download the template files:
   - Universal Connector Links
   - Universal Connector Objects
2. Fill in the required fields in the template files using the information provided in the tables below, see Universal Connector Links on page 211 and Universal Connector Objects on page 213.

Universal Connector Links



E1 — Source Component

	A	B	C	D	E	F	G	H	I
1	Process Name	Process Path	Process Type	Task Name	Source Component	Task Path	Source Provider Name	Source Server	Source Database	Sou
2									
3									
4									
5									

Column Name	Description	Required
Process Name	Name of the process that wraps the task, for example “Workflow” in Informatica or “Package” in SSIS	No
Process Path	Path of the process – for example, the path where the SSIS package is stored, including the package name and suffix (aaa\bbb\ccc \Package Name.dtsx).	No
Process Type	The type of process – job, map, package, and so forth.	Yes
Process Description	Short process description to be identified clearly in the lineages.	No
Task Name	The task name – the atomic unit that holds the data flow within the process.	Yes
Task Path	The path of the task – the location of the atomic unit that runs the process (for example, aaa\bbb\ccc\Package Name\container\Task Name).	No
Source Component	Name of the logic component in the ETL tool. Example: for Informatica, the name of the aggregator in the map. When there is no component, enter the table name.	No
Source Provider Name	Provider of source object (for example, Oracle, SQL Server).	No
Source Server	Server name of the source object.	No
Source Database	Database name of the source object.	Yes
Source Schema	Schema name of the source object.	Yes
Source Object	Name of the source object.	Yes
Source Column	Column name in the source object.	Yes
Source Data Type	Data type of the column.	No
Source Precision	Precision of the column.	No
Source Scale	Scale of the column.	No
Source Object Type	Type of object – table, view, file.	Yes

Column Name	Description	Required
Target Provider Name	Provider of target object (for example, Oracle, SQL Server).	No
Target Component	Name of the logic component in the ETL tool. Example: for Informatica, the name of the aggregator in the map. When there is no component, enter the table name.	No
Target Server	Server name of the target object.	No
Target Database	Database name of the target object.	Yes
Target Schema	Schema name of the target object.	Yes
Target Object	Name of the target object.	Yes
Target Column	Column name in the target object.	Yes
Target Data Type	Data type of the column.	No
Target Precision	Precision of the column.	No
Target Scale	Scale of the column.	No
Target Object Type	Type of object – table, view, file.	Yes
Expression	Formula or transformation between source column and target column.	No
Link Type	DataFlow or ImpactAnalysis.	No (default = DataFlow)
Link Description	Documentation about the link.	No (default = empty string)

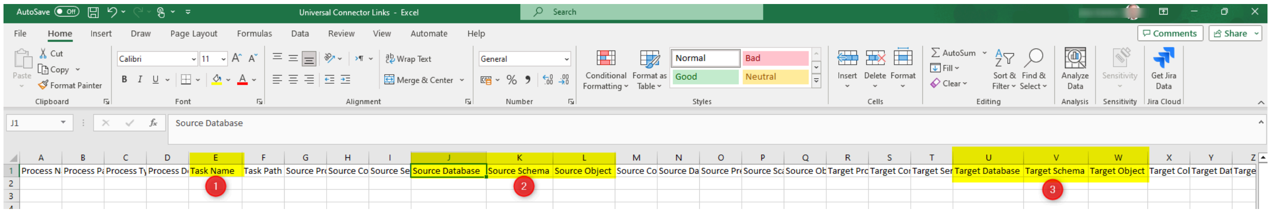
Important:

Note:

- All parameters are string values.
- Ensure the structure of the template files remains unchanged.
- Parameter names (column names) are case sensitive.
- Values are not case sensitive.

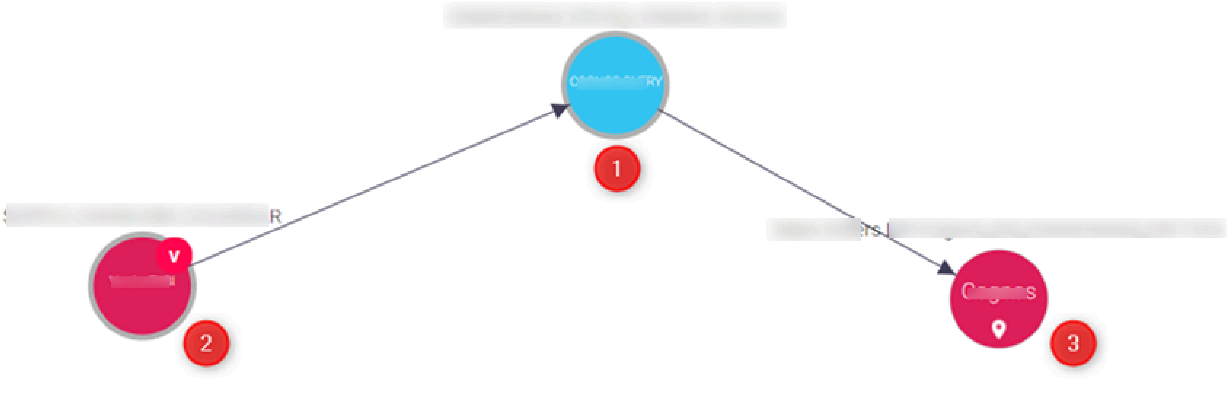
Example for ETL process on cross lineage

The Universal Connector links the source and the target for the task name as the main object.

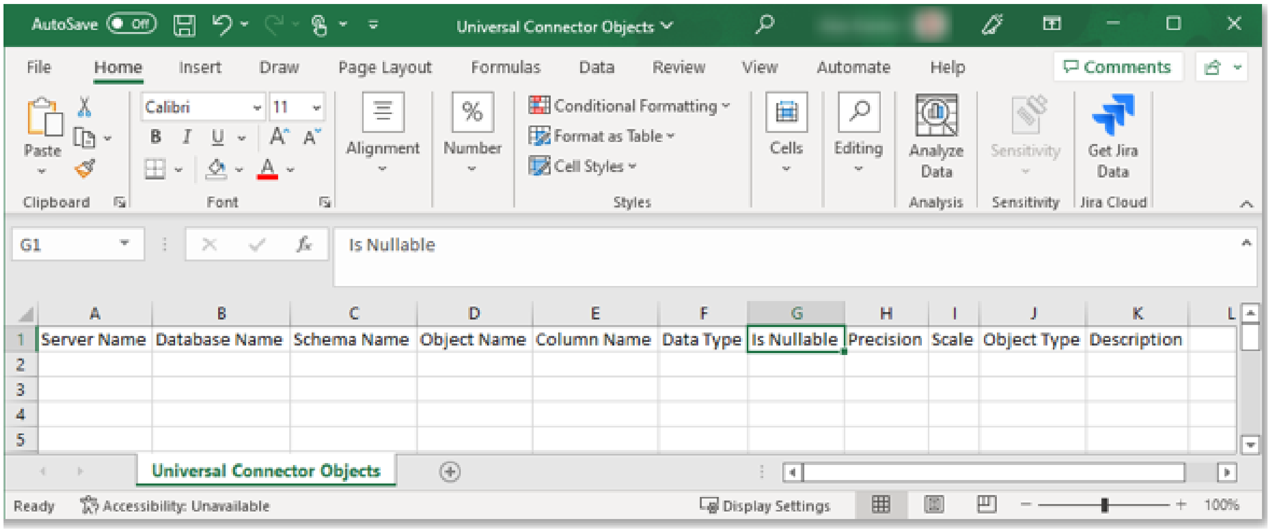


Source Database

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	Process N	Process Pa	Process Ty	Process D	Task Name	Task Path	Source Pr	Source Co	Source Se	Source Database	Source Schema	Source Object	Source Co	Source Da	Source Pre	Source Sca	Source Ob	Target Pro	Target Cor	Target Ser	Target Database	Target Schema	Target Object	Target Col	Target Dat	Targe
2																										
3																										



Universal Connector Objects



G1 | Is Nullable

	A	B	C	D	E	F	G	H	I	J	K
1	Server Name	Database Name	Schema Name	Object Name	Column Name	Data Type	Is Nullable	Precision	Scale	Object Type	Description
2											
3											
4											
5											

Column Name	Description	Required
Provider Name	Provider of object – for example, Oracle, SQL Server.	No
Server Name	Server name of the object.	No
Database Name	Database name of the object.	Yes
Schema Name	Schema name of the object.	Yes
Object Name	Name of the source object.	Yes
Object Description	Documentation about the object.	No (default = empty string)
Column Name	Column name in the source object.	Yes
Column Description	Short column description.	No (default = empty string)
Data Type	Data type of the column.	No
Is Nullable	Indicates whether the column accepts null values.	No
Precision	Precision of the column.	No
Scale	Scale of the column.	No
Object Type	Type of object – table, view, file, and so on.	Yes

Important:

Note:

- All parameters are string values.
- Ensure the structure of the template files remains unchanged.
- Parameter names (column names) are case sensitive.
- Values are not case sensitive.

How to set up the Universal Connector

For step-by-step setup instructions, see Universal Connector.

Enhancing Data Connectivity: Cloudera Octopai Universal Connector for Reporting Tools Guide

The Cloudera Octopai Universal Connector for Reporting Tools enables seamless integration of metadata from diverse systems, providing a comprehensive view of your data ecosystem.

The Cloudera Octopai Platform is dedicated to serving data teams who demand expansive analysis and visualization tools. As an enhancement to our current offerings, we've developed a Universal Connector for Reporting Tools that allows the integration of metadata from a diverse range of systems. This comprehensive integration brings about a more complete view of your data ecosystem with thorough data lineage, discovery, and an all-inclusive data catalog.

The connector allows for the integration of metadata from various types of systems into the Cloudera Octopai platform, thus offering a complete view of your data ecosystem, including full data lineage, discovery, and a comprehensive data catalog. This new connector adds to the capabilities of Cloudera Octopai, with native connectors provided for popular tools like Tableau, Power BI, Cognos, MicroStrategy, Qlik, Looker, SSRS, and more.

This guide will walk you through the steps to successfully utilize this enhancement, breaking down the necessary CSV input structure, detailed SQL queries for populating your database, and the overall workflow of the Universal Connector for Reporting Tools.

How it is done

Use the Cloudera Octopai templates to ingest your metadata into the platform. The rest is fully automated.

What Cloudera Octopai offers

This metadata, along with the metadata automatically ingested from out-of-the-box supported systems, is analyzed using machine learning. In turn Cloudera Octopai provides you with end-to-end column-level lineage, inner system lineage, cross-system lineage, data discovery, and a data catalog of your entire data landscape accessible to all data users in the organization.

The benefits

- No blind spots – perform changes with confidence.
- Get a clear picture of data transformations.
- Increase visibility of the organization's complete data ecosystem.
- Future-proof your expanding data landscape by providing access to unlimited data systems.
- Add links to our out-of-the-box technologies.

How to use the template file

1. Download the template file: Universal Connector Reports
2. Fill in the required fields in the template file using the information provided in the table below.

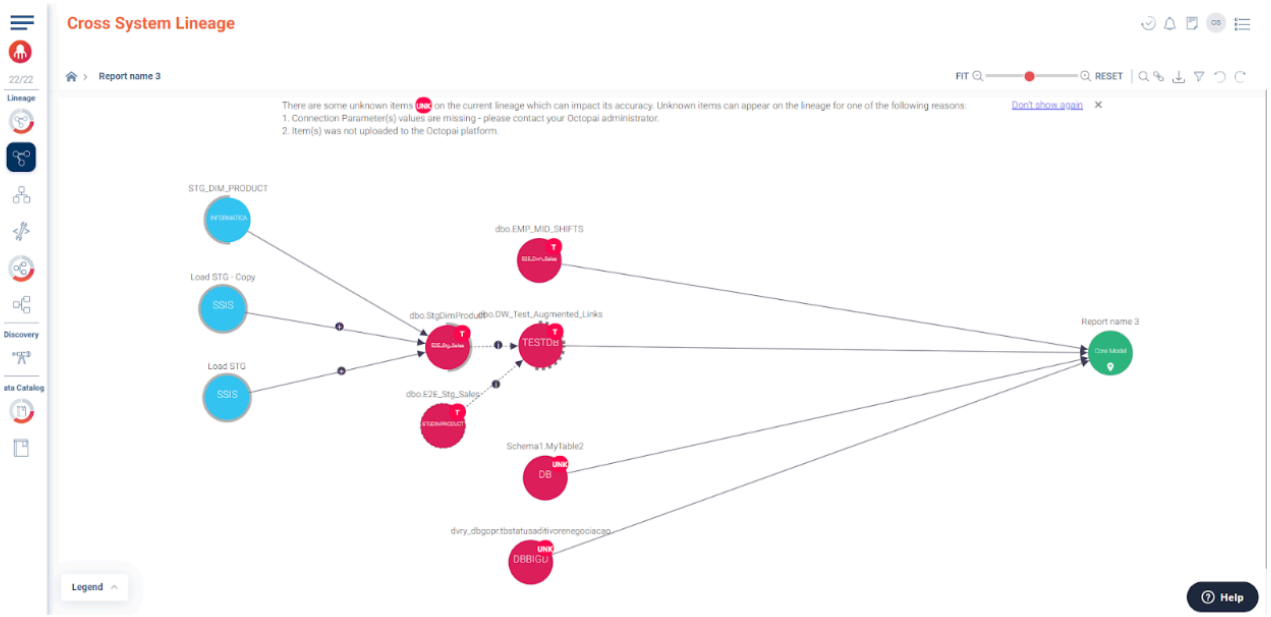
Column Name	Description	Required
MODEL_NAME	The name of the model. This could be the name of the package in Cognos, the RPD name in OBIEE, or the Universe name in Business Object.	Yes
REPORT_PATH	The unique path of the report, per ConnectionID.	Yes
REPORT_NAME	The name of the report.	Yes
SOURCE_PROVIDER	The type of DB that the report connects to.	Yes
SOURCE_SERVER	The server name of the source object.	No
SOURCE_DB	The database name of the source object.	Yes
SOURCE_SCHEMA	The schema name of the source object.	Yes
SOURCE_OBJECT_TYPE	The type of source object, such as Table, View, or Stored Procedure (SP).	Yes
SOURCE_COLUMN	The column name in the source object.	Yes
SOURCE_TABLE	The table name in the source object.	Yes
SOURCE_DATA_TYPE	The data type of the column.	No
SOURCE_PRECISION	The precision of the column.	No
SOURCE_SCALE	The scale of the column.	No
TARGET_LAYER_NAME	The name of the component in the report, such as a table, cross-tab, chart, etc.	Yes
TARGET_SERVER	The server name of the target object.	No
TARGET_DB	The database name of the target object.	No
TARGET_SCHEMA	The schema name of the target object.	No
TARGET_OBJECT_TYPE	The type of target object, for example, Presentation. Default = 'PresentationTable'.	Yes
TARGET_TABLE	The table name in the target object.	No
TARGET_COLUMN	The column name in the target object.	Yes
TARGET_DATA_TYPE	The data type of the column.	No
TARGET_PRECISION	The precision of the column.	No
TARGET_SCALE	The scale of the column.	No
EXPRESSION	The formula or transformation between the source column and target column.	No
LINK_TYPE	Data Flow or Impact Analysis. Default = Data Flow.	No
LINK_DESCRIPTION	Documentation about the link. Default = Null.	No
UPDATED_DATE	The timestamp when a row was inserted. This field is automatically populated.	Yes

Important:

Note:

- All parameters are string values.
- Ensure the structure of the template files remains unchanged.
- Parameter names (column names) are case sensitive.
- Values are not case sensitive.

Example of a report cross-system lineage created as part of an upload of the Universal Connector:



There are some unknown items UNK on the current lineage which can impact its accuracy. Unknown items can appear on the lineage for one of the following reasons:

1. Connection Parameter(s) values are missing - please contact your Octopai administrator.
2. Item(s) was not uploaded to the Octopai platform.

Don't show again ✕

How to set up the Universal Connector

For step-by-step setup instructions, see Universal Connector.