

Lab Guide

# Prime Brokerage Hands-on

Cloud Pak for Data v4.7.x

IBM Data and AI

Table of Contents

<b><u>BACKGROUND .....</u></b>	<b><u>3</u></b>
<b><u>ADDITIONAL RESOURCES.....</u></b>	<b><u>4</u></b>
<b><u>PRE-REQ: KNOWLEDGE ACCELERATOR, BUSINESS TERMS, CATEGORIES, CATALOGS.....</u></b>	<b><u>5</u></b>
<b><u>PRE-REQ: ADD SAMPLE DATA CONNECTIONS FOR POSTGRESQL &amp; DB2.....</u></b>	<b><u>10</u></b>
<b><u>CREATE A PROJECT FOR PRIME BROKERAGE ASSETS .....</u></b>	<b><u>15</u></b>
<b><u>METADATA IMPORT (MDI).....</u></b>	<b><u>21</u></b>
<b><u>METADATA IMPORT - DISCOVER.....</u></b>	<b><u>22</u></b>
<b><u>METADATA ENRICHMENT .....</u></b>	<b><u>28</u></b>
<b><u>METADATA IMPORT - LINEAGE.....</u></b>	<b><u>35</u></b>
<b><u>VIEW POSTGRESQL LINEAGE IN THE CATALOG.....</u></b>	<b><u>37</u></b>
<b><u>DATA STAGE FLOWS IMPORT.....</u></b>	<b><u>40</u></b>
<b><u>YOU WILL FIND THE DATASTAGE FLOWS IN ZIP FORMAT: .....</u></b>	<b><u>40</u></b>
<b><u>CONFIGURE DATASTAGE WITH DATABAND.....</u></b>	<b><u>44</u></b>

## BACKGROUND

**A data fabric is an architectural approach to simplify data access in an organization to facilitate self-service data consumption.** This architecture is agnostic to data environments, processes, utility, and geography. The goal is to automate data discovery, governance, , data pipeline's and consumption enabling enterprises to use data to maximize their value chain.

IBM's Cloud Pak for Data provides a data fabric solution for faster, trusted AI outcomes by connecting the right data, at the right time, to the right people, from anywhere it's needed. A single, unified platform that spans hybrid and multi-cloud environments to ingest, explore, prepare, manage, govern, and serve petabyte-scale data for business-ready AI.

Data is useful only if its context, content, and quality are trusted. To keep it that way, data must be continuously evaluated, and the appropriate remediation must be applied if required. Data traceability requires to trace back the data from its source to destination to identify complete record of data journey. With Watson Knowledge Catalog coupled with Databand on Cloud Pak for Data, data stewards can configure recurring jobs to continuously track changes to the content and structure of data and analyze only the data that changed.

The purpose of this guide is to provide DataStage workflows creation with sample connections and demonstration steps for technologies supported by Cloud Pak for Data's data lineage capabilities. The guide will walk through the steps for creating a connection, importing metadata from various technologies, enriching the data assets, and visualizing data lineage directly within a catalog.

This lab guide will cover the following Data Fabric components and what's new in Cloud Pak for Data v4.7.x including:

- Metadata Import & Metadata Enrichment
- Data Governance
- Data Lineage
- Data Integration Workflows
- Data Lineage

\*At the end of each section, you will find the assets (scripts, csv files, etc.) needed to recreate steps in your own environment.

Please note: This guide will be continuously updated as sample connections to additional technologies become available.

## ADDITIONAL RESOURCES

1. Cloud Pak for Data 4.7.x TechZone Environment

<https://techzone.ibm.com/collection/PakInstaller>

2. Installation Guide

Will provide in session

3. Workflow Jobs of DataStage

<https://github.ibm.com/poholkar/Banking-PRIME-BROKERAGE>

## PRE-REQ: KNOWLEDGE ACCELERATOR, BUSINESS TERMS, CATEGORIES, CATALOGS.

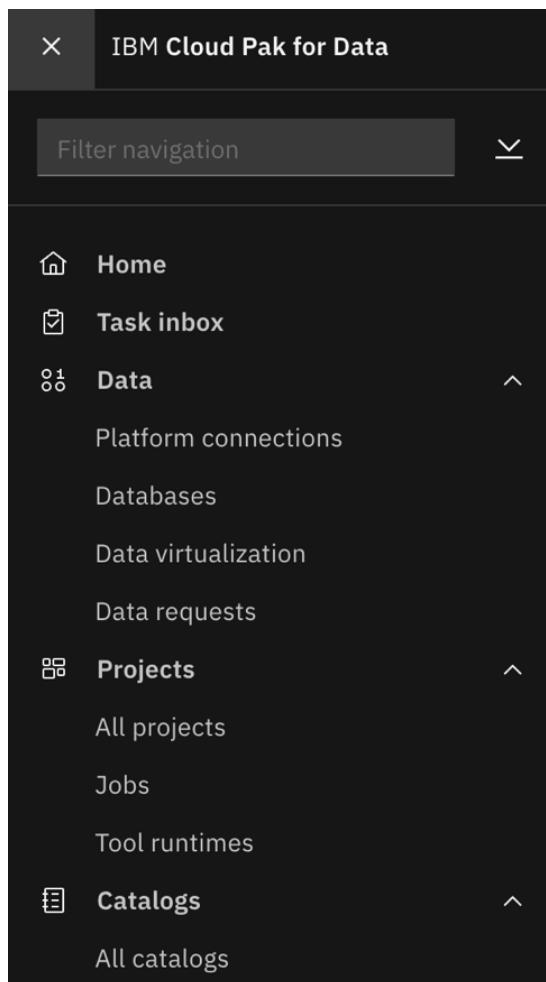
To prepare for the labs, complete the following steps.

1. **Enable Knowledge accelerator(KA) for pre-created, extensive, curated glossaries in Finance.**  
**Follow below steps to enable KA:**

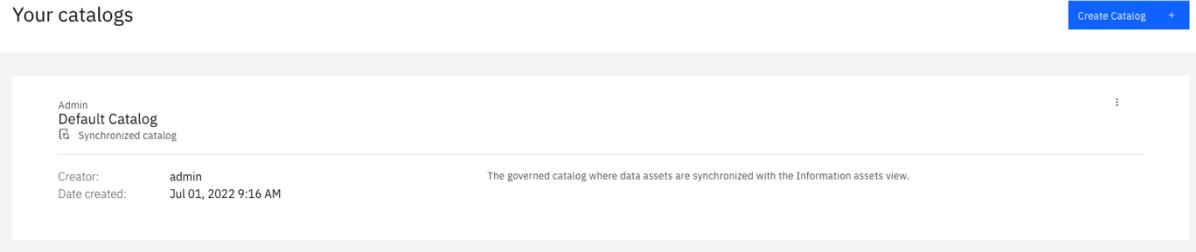
- Get your IBM Cloud Pak for Data URL from Techzone
  - export CPD\_URL={CPD\_URL}
- Generate your token using your password
  - curl -k -X POST \$CPD\_URL/icp4d-api/v1/authorize -H 'cache-control: no-cache' -H 'content-type: application/json' -d '{"username":"admin","password":"<admin password>"}'
- Set the TOKEN variable, copy and paste the token value obtained in the above step into the following command:
  - export TOKEN={TOKEN}
- Set the KA\_IND variable for the Knowledge Accelerator for Finance
  - export KA\_IND= kafs20
- To check what is going to be imported in advance, run the following command
  - curl -X POST "\$CPD\_URL/v1/knowledge\_accelerators/\$KA\_IND/import\_items?ids=core-bpi&merge\_option=all" -k -s -H "Authorization: Bearer \$TOKEN"
- Import the selected Knowledge Accelerator core vocabulary using the following command:
  - curl -X POST "\$CPD\_URL/v1/knowledge\_accelerators/\$KA\_IND/import\_items?ids=core-bpi&merge\_option=all&do\_import=true" -k -s -H "Authorization: Bearer \$TOKEN"
- When you issue the command to import the Knowledge Accelerator items, a process id is returned in the command prompt, for example:
  - "process\_id": "CPD\_URL/v3/governance\_artifact\_types/import/status/6e23b62f-e20d-4fd8-88d7-06f92950fdb9" where the {process id} is 6e23b62f-e20d-4fd8-88d7-06f92950fdb9
- Set a new environment PROCESS\_ID variable to store the process id value from the import response:
  - export PROCESS\_ID={process id}
- Run the following command to check the import status:
  - curl -X GET "\$CPD\_URL/v3/governance\_artifact\_types/import/status/\$PROCESS\_ID" -k -s -H "Authorization: Bearer \$TOKEN"

2. Log into Cloud Pak for Data homepage with the credentials: **admin\_id/xxxx**

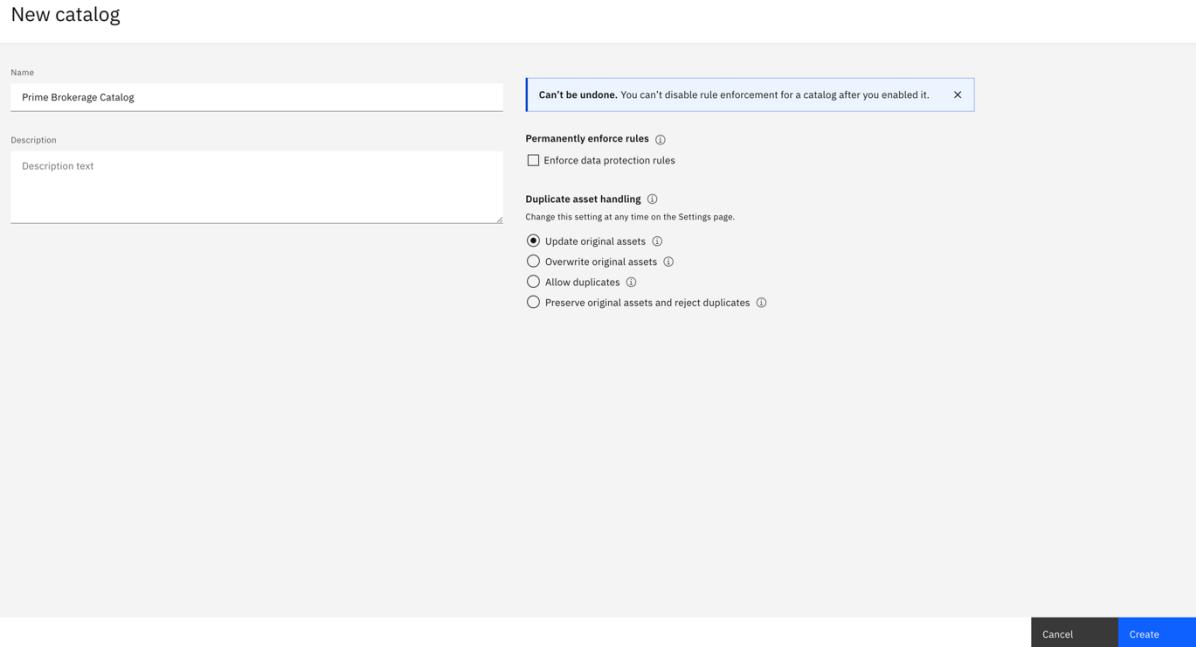
**While remaining in the admin role, navigate to Catalogs → All Catalogs**



3. Click **Create Catalog**.



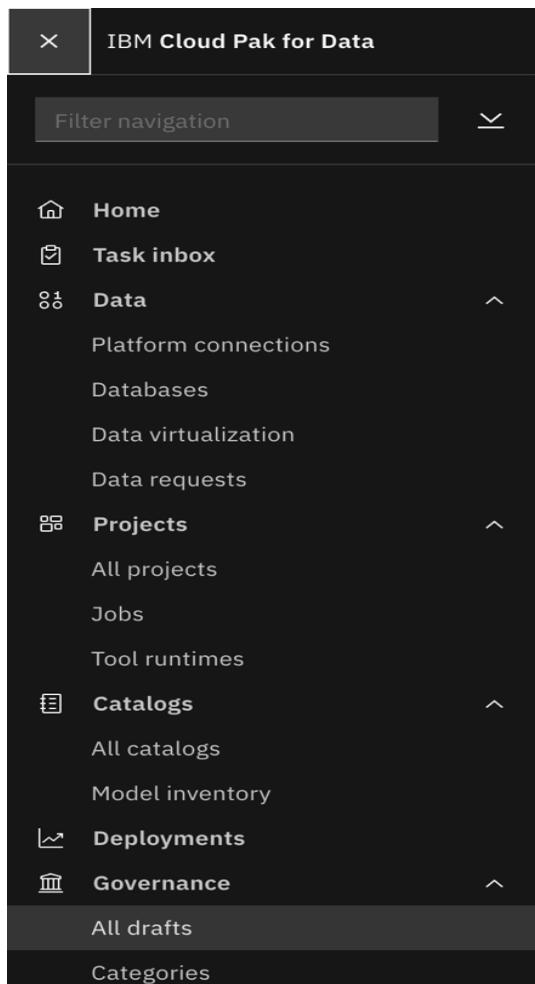
- Enter a name for your catalog (ex. **Prime Brokerage Catalog\_(Yourinitials)**).
- Select **Update original assets**.



After the catalog is created, proceed to the next steps. You will reference the same catalog when publishing lineage assets.

For the Business Glossary of financial domain, we are using knowledge accelerator. The IBM Knowledge Accelerator for Financial Services is a comprehensive resource that encompasses a rich array of terms and concepts, providing a nuanced understanding of the intricate nature of the business information handled by Financial Institutions in their daily operations.

4. Navigate to **Governance → Categories**.



You can view the populated business glossary of Knowledge accelerator (KA) for Financial Services

**Categories**

Manage category roles 

Name	Description	Subcategories	Modified by
[uncategorized]	This is the system default if a standard category is not assigned.	0  System	
Healthcare_Data	Category that organizes all of the data governance artifacts related to the HealthCare project.	0  admin 	
Knowledge Accelerator for Cross Industry	IBM Knowledge Accelerator for Cross Industry Copyright IBM Corp. 2020, 2022. Documentation for this offering...	2  admin	
Knowledge Accelerator for Financial Services	IBM Knowledge Accelerator for Financial Services. Copyright IBM Corp. 2020, 2022. Documentation for this offering...	4  admin	
Locations	Locations Category (LC) is defined as parent for locations reference datasets.	0  System	
Person	Data governance artifacts related to a person like a Customer, Employee, Partner, Loan Applicant, Account Holder etc.	2  admin	

The KA hence populates different artifacts like Business terms, Classifications, Data Classes for finance domain.

## 5. Navigate to Governance → Business Terms

### Business terms

Add business term 

Published	Draft	
Q Find business terms	I	
		Sort by: Name  Show: All  Edit 
30-Year Treasury Bond	Identifies a Treasury Bond as having a 30 year maturity period. A 30-Year Treasury Bond is a U.S. Treasury debt obligation that has a maturity of 30 years. The 30-year...	 Product Last modified: Mar 7, 2023, 10:48 PM by admin
Abandonment Option Flag	Identifies securities which provide the ability to liquidate an investment before the date of expected or allowed termination.	 Product Last modified: Mar 7, 2023, 10:48 PM by admin
ABCP Program	Identifies a service which is a program for the creation of Asset-Backed Commercial Paper. Underlying exposures of the ABCP are held in a bankruptcy-remote SPE. The...	 Product Last modified: Mar 7, 2023, 10:48 PM by admin
ABCP Program Product Sponsor	Identifies an Involved Party that is the sponsor of an Asset Backed Commercial Paper (ABCP) program Product, which is recorded as a service.	 Involved Party Last modified: Mar 7, 2023, 10:48 PM by admin
Absolute Point In Time	Identifies a Point In Time that specifies the exact calendar date and/or time when something is agreed to transpire; for example, electronic transmissions to a payment...	 Condition Last modified: Mar 7, 2023, 10:48 PM by admin
Accept Or Reject Date	The date on which the Financial Institution either accepted to honor the Claim, or rejected the Claim.	 Event Last modified: Mar 7, 2023, 10:48 PM by admin
Acceptance Indicator	Indicates the degree of acceptance that the Financial Institution applied to the Claim. For example: (A) Accepted (P) Accepted Partially (i.e. insufficient cover) (R) Rejected.	 Event Last modified: Mar 7, 2023, 10:48 PM by admin
Acceptance Service	Identifies a Trade Finance Service in which the Financial Institution agrees to pay the face value of a Bill of Exchange if the issuer fails to pay. The Bill of Exchange can either...	 Product Last modified: Mar 7, 2023, 10:48 PM by admin

## PRE-REQ: ADD SAMPLE DATA CONNECTIONS FOR POSTGRESQL & DB2

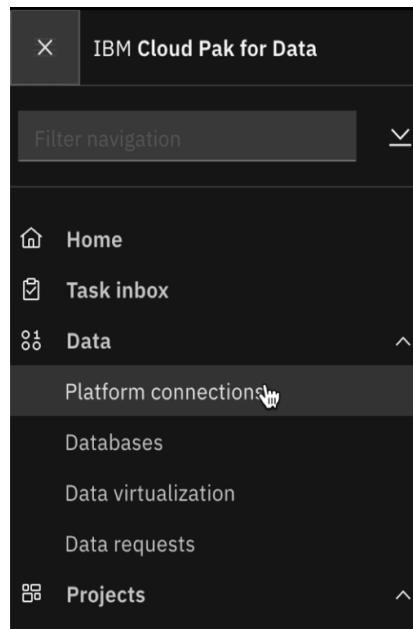
Next, we will add sample data connections for PostgreSQL and Db2.

\*To load the data into your own PostgreSQL or Db2 instance, you can find the data assets and scripts below:

- PostgreSQL – bfsi\_asset schema
- <https://ibm.box.com/s/snoeugbzlfcnuci0cxpzwmkn5gni>
- Db2 – RISKREPORT\_(initialxx) Schema, DEPOT\_POSITION csv files

<HTTPS://IBM.ENT.BOX.COM/FOLDER/215248216782>

1. While logged in as admin, navigate to **Data → Platform connections**.



## 2. Add a new connection for PostgreSQL.

New connection

Select the type of data source that you want to connect to.

**Provider**

- IBM
- Third-party
- User-defined

**Compatible services**

- Catalogs
- Dashboard
- Data Replication
- Data Virtualization
- Data discovery and Data quality
- DataStage
- Metadata import (discovery)
- Metadata import (lineage)
- Watson Studio

**All connection types**

Provider	Connection Type	Description
Amazon RDS for MySQL	Google Cloud Storage	IBM Db2 Warehouse
Amazon RDS for Oracle	Greenplum	IBM Informix
Amazon RDS for PostgreSQL	HDFS via Execution Engine for Hadoop	IBM Match 360
Amazon Redshift	Hive via Execution Engine for Hadoop	IBM MQ
Amazon S3	HTTP	IBM Netezza Performance Server
Apache Cassandra	IBM Cloud Compose for MySQL	IBM Netezza Performance Server (optimized)
Apache Cassandra (optimized)	IBM Cloud Data Engine	IBM Planning Analytics
Apache Derby	IBM Cloud Databases for DataStax	IBM SPSS Analytic Server
Apache HBase	IBM Cloud Databases for MongoDB	IBM Watson Query
Apache HDFS	IBM Cloud Databases for PostgreSQL	Impala via Execution Engine for Hadoop

**Selected connection type**

- PostgreSQL

**Details**

PostgreSQL is an open source and customizable object-relational database. For the PostgreSQL connection type on IBM Cloud, select IBM Cloud Databases for PostgreSQL.

**Compatible services**

- Catalogs
- Dashboard
- DataStage
- Data Virtualization
- Metadata import (lineage)
- Metadata import (discovery)
- Data Replication
- Watson Studio

**Cancel** **Back** **Select**

Name: [PostGreDB](#)

Database: [postgres](#)

Hostname: XXXXXX

Port: [5432](#)

Username: [postgres](#)

Password: [XXXXXX](#)

4. Test the connection then click **Create**.

Create connection: PostgreSQL

Enter the connection information.

Test connection

Connection overview

Connection details

Credentials

Certificates

Other properties

The test was successful.  
Click Create to save the connection information.

postgres

Hostname or IP address\* 169.60.148.140

Port\* 5432

Credentials

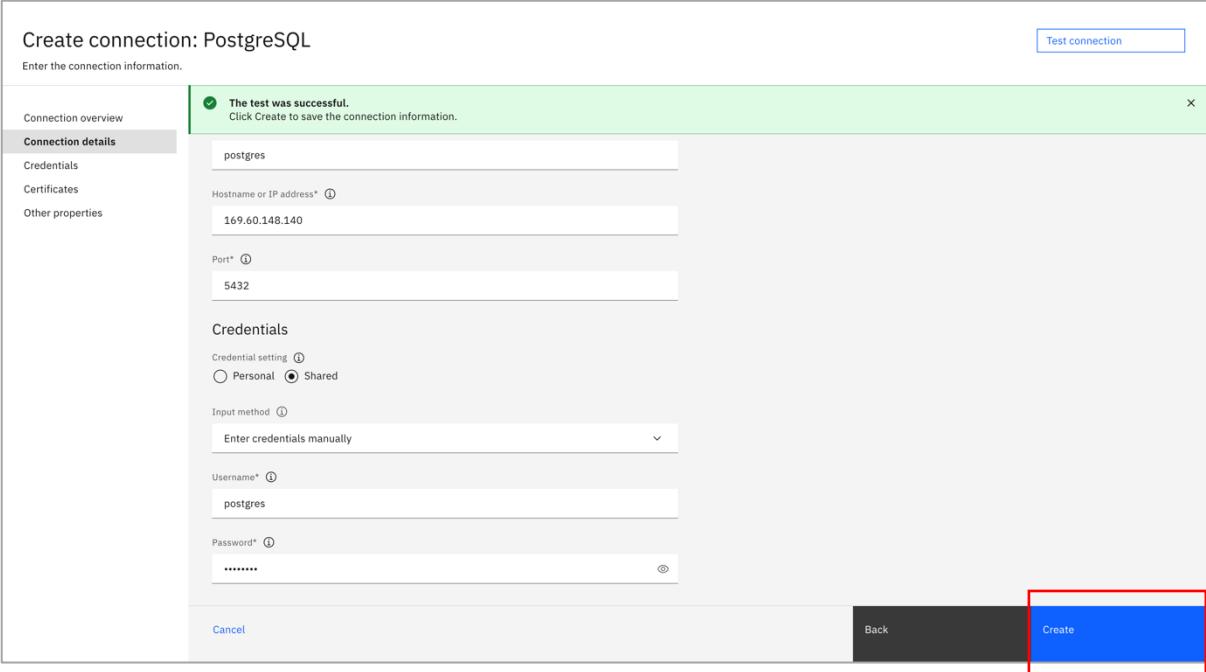
Credential setting ①  
 Personal  Shared

Input method ①  
Enter credentials manually

Username\* postgres

Password\* ⋮

Cancel Back Create



Next, we will create connection for a Db2 database.

The screenshot shows the 'New connection' dialog box. On the left, there's a sidebar with 'Provider' (IBM, Third-party, User-defined) and 'Compatible services' (Catalogs, Dashboard, Data Replication, Data Virtualization, Data discovery and Data quality, DataStage, Metadata import (discovery), Watson Studio). The main area has a search bar with 'db2' and a list of connection types: Db2 (optimized), IBM Db2 Event Store, IBM Db2 Hosted, IBM Db2 on Cloud (selected and highlighted with a blue border), IBM Db2 for i, IBM Db2 for z/OS, IBM Db2 Big SQL, IBM Db2 Warehouse. On the right, a 'Selected connection type' panel shows 'IBM Db2 on Cloud' with 'Details' (an SQL database managed by IBM Cloud) and 'Compatible services' (Catalogs, Dashboard, DataStage, Data Virtualization, Metadata import (lineage), Metadata import (discovery), Data Replication, Watson Studio). At the bottom, there are 'Cancel', 'Back', and a large blue 'Select' button.

## 5 . Credentials

Name: [Db2OnCloud](#)

Database: [bludb](#)

Hostname: [XXXXXX](#)

Port: [30315](#)

Username: [f353587e](#)

Password: [XXXXXXXX](#)

**Check port is SSL-enabled – do not need to paste certificate**

### Create connection: IBM Db2 on Cloud

Enter the connection information.

[Test connection](#)

The test was successful.  
Click Create to save the connection information.

Hostname or IP address\* [?](#)  
a5617cc5-bda2-422d-8f73-b01685c09b55.bpe60pb01oinge4psd0.databases.appdomain.cloud

Port [?](#)  
30315

Credentials

Credential setting [?](#)  
 Personal  Shared

Input method [?](#)  
Enter credentials manually

Username\* [?](#)  
f353587e

Password\* [?](#)  
\*\*\*\*\*

Mask sensitive credentials retrieved through API calls [?](#)

[Cancel](#) [Create](#)

Once created the two added databases will reflect in Platform connections.

### Platform connections

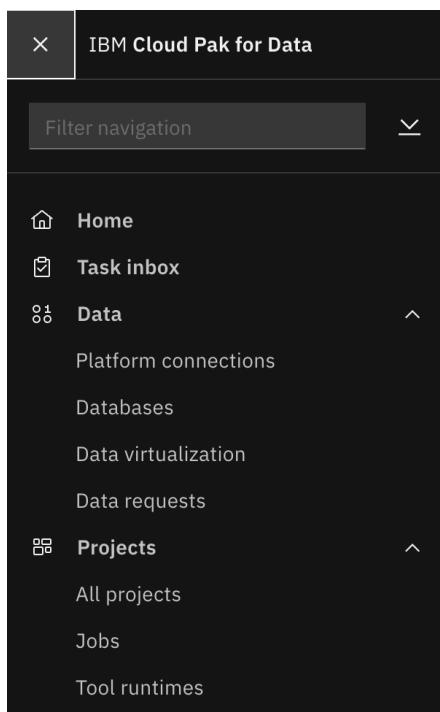
[Supported connectors](#)

Connected data sources				
Manage existing data source connections or create new connections that can be used across the platform.				
Filter by: All connectors				
Name	Connector	Created by	Modified by	Last updated
Db2OnCloud	IBM Db2 on Cloud	Admin	admin	Aug 02, 2023
PostGreDB	PostgreSQL	Admin	admin	Aug 02, 2023

# CREATE A PROJECT FOR PRIME BROKERAGE ASSETS

A [project](#) is a collaborative workspace where you work with data and other assets to accomplish a particular goal. Projects include resources such as collaborators, data assets, jobs, integrations, documentation, notifications and more.

1. **Navigate to Projects → All Projects** from the Cloud Pak for Data homepage.



2. **Click New Project.**

A screenshot of the "Projects" page. The page has a light gray header with the word "Projects". Below the header is a search bar with a magnifying glass icon and the placeholder text "Find a project". To the right of the search bar is a blue button labeled "New project" with a plus sign. The main area is a table with the following columns: "Name", "Date created", "Your role", and "Collaborators". There is one row in the table, showing a project named "RetestMultiCloud" created "2 weeks ago" by "Admin". The "Collaborators" column shows a red circular icon with the number "1".

3. Enter a name for your project (ex. “Prime Brokerage\_Your initials”) and click on create.

4. Add collaborators to the project. Select Project you create in step 3 → click on Manage → click on Access control

The screenshot shows the 'General' tab selected in the sidebar. The 'Manage' button is highlighted with a red box. The 'Access control' section is also highlighted with a red box. The main area displays project details like Name (Prime Brokerage), Description (What's the purpose of this project?), Project ID (3811a030-b18a-4b25-bc03-5ae022c42a6e), and storage usage (0 Bytes). Below this is a 'Danger zone' section with 'Leave project' and 'Delete project' buttons, both of which are highlighted with red boxes.

5. Next Click on Add collaborators

The screenshot shows the 'Access control' tab selected in the sidebar. The 'Add collaborators' button is highlighted with a red box. The main area displays a list of collaborators with one entry: 'admin (you)' with a role of 'Admin'. A search bar labeled 'Find collaborators' is also visible.

6. Next click on Add users and type admin in search bar to select.

The screenshot shows the 'Access control' page within the 'Manage' tab. On the left, there's a sidebar with 'Project' sections: General, Access control (which is selected and highlighted in grey), and Environments. The main area has a header 'Access control' and a sub-section 'Collaborators (1)'. Below it is a search bar 'Find collaborators'. A table lists one collaborator: 'admin (you)' with a role of 'Admin'. In the top right corner of the 'Add collaborators' dropdown, a red box highlights the 'Add users' button.

7. Once selected add Editor as a Role and select Add

The screenshot shows the 'Add users as collaborators' dialog. At the top, there's a search bar with 'admin' typed in. Below it, a table shows a single result: 'Search results 1' with 'admin' and a checked checkbox. To the right, a 'Selected users' section shows '1' user selected, with a thumbnail of 'admin' and the role 'Editor'. A dropdown menu for 'Role' is open, showing 'Editor' selected. At the bottom right of the dialog, a red box highlights the 'Add' button.

8. Once added you will see your and admin userid in collaborators

The screenshot shows the 'Access control' section of a project. On the left, there's a sidebar with 'Project', 'General', 'Access control' (which is selected and highlighted in grey), and 'Environments'. The main area has a header 'Access control' and 'Collaborators (2)'. Below that is a search bar 'Find collaborators'. A table lists two users:

Name	Role
PP pooja (you)	Admin
AA ..	Editor

9. Let us add the platform connections already been created in our project .

Click the **Assets** tab, then **New Asset**.

The screenshot shows the 'Assets' tab selected in the top navigation bar. Below it is a search bar 'Find assets'. At the bottom right of the screen, there are buttons for 'Import assets' and 'New asset'.

10. Click the **Connection** tab and then select platform connection.

### New asset

Select a tool based on what type of asset you want and how you want to work.

The screenshot shows the 'New asset' dialog. On the left, a sidebar lists 'Tool type' categories: 'All types' (selected), 'Data access tools', 'Automated builders', 'Graphical builders', 'Code editors', and 'Component editors'. The main area has a message 'Connected data moved' and a search bar 'Find tools by name or description'. Under 'Data access tools', there are three options: 'Connection' (highlighted with a red arrow), 'Metadata import', and 'Model'. Below that is a section for 'Automated builders' with 'AutoAI' and 'Metadata enrichment'.

## New connection

Create a new connection or select an existing connection from the list of platform connections.

New From platform

**Provider**

- IBM
- Third-party
- User-defined

**Compatible services**

- Catalogs
- Dashboard
- Data Replication
- Data Virtualization
- Data discovery and Data quality
- DataStage
- Metadata import (discovery)

Find platform connections

All connection types

Db2OnCl... PostGreDB

### 11. Select the platform connections Db2OnCloud and PostgreDb one at a time.

#### New connection

[Supported connection types](#)

New From platform

**Provider**

- IBM
- Third-party
- User-defined

**Compatible services**

- Catalogs
- Dashboard
- Data Replication
- Data Virtualization
- Data discovery and Data quality
- DataStage
- Metadata import (discovery)
- Metadata import (lineage)
- Watson Studio

Find platform connections

All connection types

Db2OnCl... PostGreDB

**Selected platform connection**

Db2OnCloud

**Connection type**  
IBM Db2 on Cloud

**Description**

Cancel Select

**12. Create the connection for the asset.**

Create connection: IBM Db2 on Cloud

Enter the connection information.

**Connection overview**

Name: Db2OnCloud

Description: Connection description

**Connection details**

Database\*: BLUDB

Hostname or IP address\*: a5617cc5-bda2-422d-8f73-b01685c09b55.bpe60pb0loinge4psd0.databases.appdomain

**Create**

**13. Once created both the connections should be visible in all assets.**

Projects / Prime Brokerage

Overview Assets Jobs Manage

Find assets Import assets New asset

**2 assets**

All assets

Name	Last modified	⋮
PostGreDB Connection	Now admin (You)	⋮
Db2OnCloud Connection	Now admin (You)	⋮

Asset types Data access 2

## METADATA IMPORT (MDI)

There are two options for MDI:

- 1) Discover: Import metadata associated with the data into a project or a catalog to inventory, evaluate, and catalog these assets. Technical metadata in this context describes the characteristics of data objects. This metadata helps users decide whether the data is appropriate for the task at hand, whether they can trust the data, and how to work with the data.
- 2) Get Lineage: Import metadata from data integration and reporting tools into a catalog. Capture lineage information from various data sources to provide business lineage information for imported assets. You add information to a catalog about where your data comes from, how it changes, and where it moves over time. In addition, you can access more advanced lineage information for the imported asset in MANTA Automated Data Lineage. Additional lineage details include technical data lineage, historical data Lineage, and indirect data lineage.

*Note: To import lineage, a valid MANTA license key must be installed. Lineage feature requirements:*  
<https://www.ibm.com/docs/en/cloud-paks/cp-data/4.7.x?topic=metadata-getting-lineage>

# METADATA IMPORT - DISCOVER

1. Search for the **Metadata Import (MDI)** asset type.

New asset

Select the tool to create an operational or configuration asset.

Tool type

All types

Federated Learning		Jupyter notebook editor	
<input type="checkbox"/> Automated builders	<input type="checkbox"/> Graphical builders	<input type="checkbox"/> Create a federated learning experiment to train a common model on a set of remote data sources. Share training results without sharing data.	<input type="checkbox"/> Create a notebook in which you run Python, R, or Scala code to prepare, visualize, and analyze data, or build a model.
<input type="checkbox"/> Code editors	<input type="checkbox"/> Component editors		
<input type="checkbox"/> Data access tools			

Component editors

Data quality definition	Data quality rule	DataStage component	Parameter set
<input type="checkbox"/> Create abstract rule logic for data analysis that can be used in any number of data quality rules.	<input type="checkbox"/> Create rules to assess the quality of your data by evaluating and validating specific conditions.	<input type="checkbox"/> Create reusable components for DataStage flows, such as, subflows, libraries, and data definitions.	<input type="checkbox"/> Collect multiple job parameters with specified values to reuse in jobs.

Data access tools

Connected data	Connection	Metadata import	Model
<input type="checkbox"/> Data in an external data source that is accessed through a connection.	<input type="checkbox"/> Supply the information necessary to connect to a data source.	<input type="checkbox"/> Import asset metadata from a connection into a project or a catalog.	<input type="checkbox"/> Add an existing PMML (predictive model markup language) file (.xml) from your local system as a model

Show descriptions (1)

2. Click **Discover** as the MDI goal, then **Next**

New asset

## Create a metadata import

### Define goal

What is the intended goal of your import?

#### Discover

Import and view assets of various types in a project or a catalog. If you import assets into a project, you can also set up and run metadata enrichment for those data assets.

Import can include  
Technical metadata  
Access to data

#### Get lineage

Add information to a catalog about where your data comes from, how it changes, and where it moves over time.

Import can include  
Technical metadata  
Lineage metadata

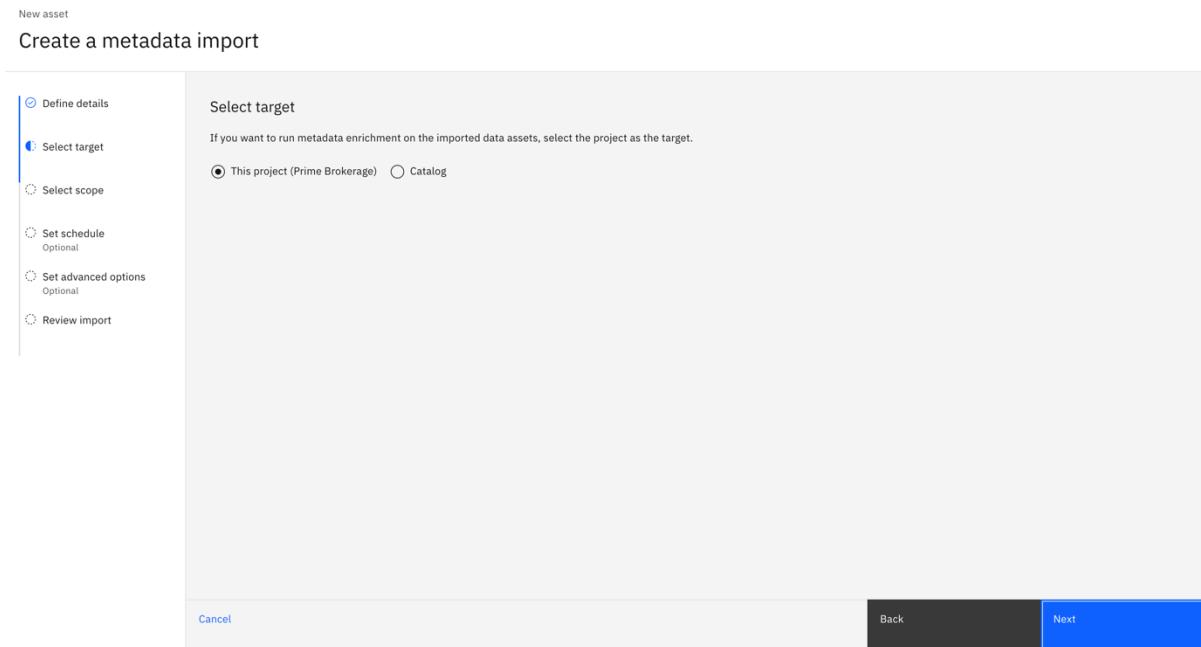
3. Enter a name and description for the MDI(metadata-discover-DB2). Click Next.

metadata-discover-pb

### Edit metadata import

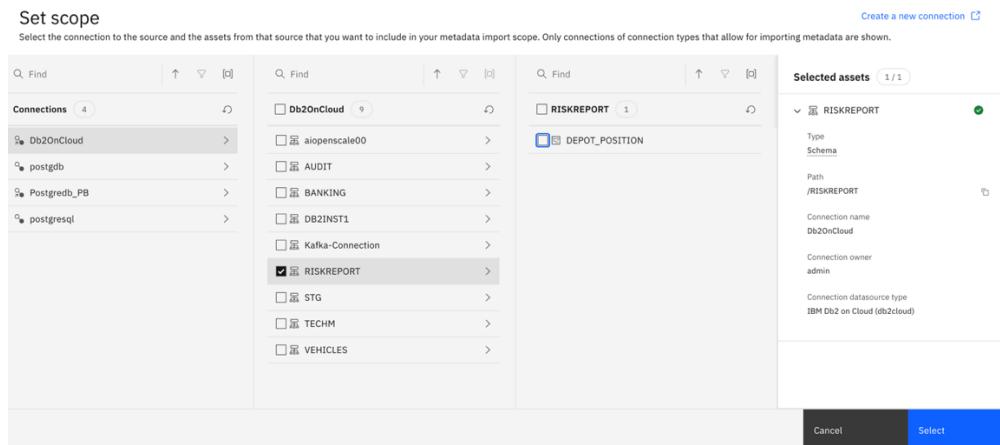
<ul style="list-style-type: none"> <li><input checked="" type="radio"/> Define details</li> <li><input type="radio"/> Select target</li> <li><input type="radio"/> Select scope</li> <li><input type="radio"/> Set schedule           <small>Optional</small></li> <li><input type="radio"/> Set advanced options           <small>Optional</small></li> <li><input type="radio"/> Review import</li> </ul>	<p><b>Define details</b></p> <p>Specify some basic information about your metadata import to make it easy to identify.</p> <p>Name <input type="text" value="metadata-discover-pb"/></p> <p>Description (optional) <input type="text" value="What's the purpose of this metadata import?"/></p> <p>Tags (optional) <small>Add tags to make assets easier to find.</small> <input type="text" value="Search tags"/></p>
<a href="#">Cancel</a> <span style="margin: 0 10px;">Back</span> <span style="background-color: #0070C0; color: white; padding: 5px 10px; border-radius: 5px;">Next</span>	

4. Select the current project as the target. Click Next.



5. Click **Select connection** then **add the platform connection** you added in previous step for db2.

6. Select the connection, then the **RISKREPORT** schema. Click **Select**.



7. Click **Next**. Leave the schedule off and set advanced options selected default then click **Next**.

New asset

### Create a metadata import - Discover

Define details

Select target

Select scope

Set schedule  
Optional

Set advanced options  
Optional

Review import

**Selected assets**  
1 assets

Find assets

	Name	Type	Context
<input type="checkbox"/>	RISKREPORT	schema	/RISKREPORT

Edit data scope

Cancel Back Next

New asset

### Create a metadata import

Define details

Select target

Select scope

Set schedule  
Optional

Set advanced options  
Optional

Review import

**Advanced options**

Update on reimport

By default, all properties are updated when assets are reimported. If you want any of the following properties to remain unchanged, clear the respective check box.

Asset name

Asset description

Column description

Delete on reimport

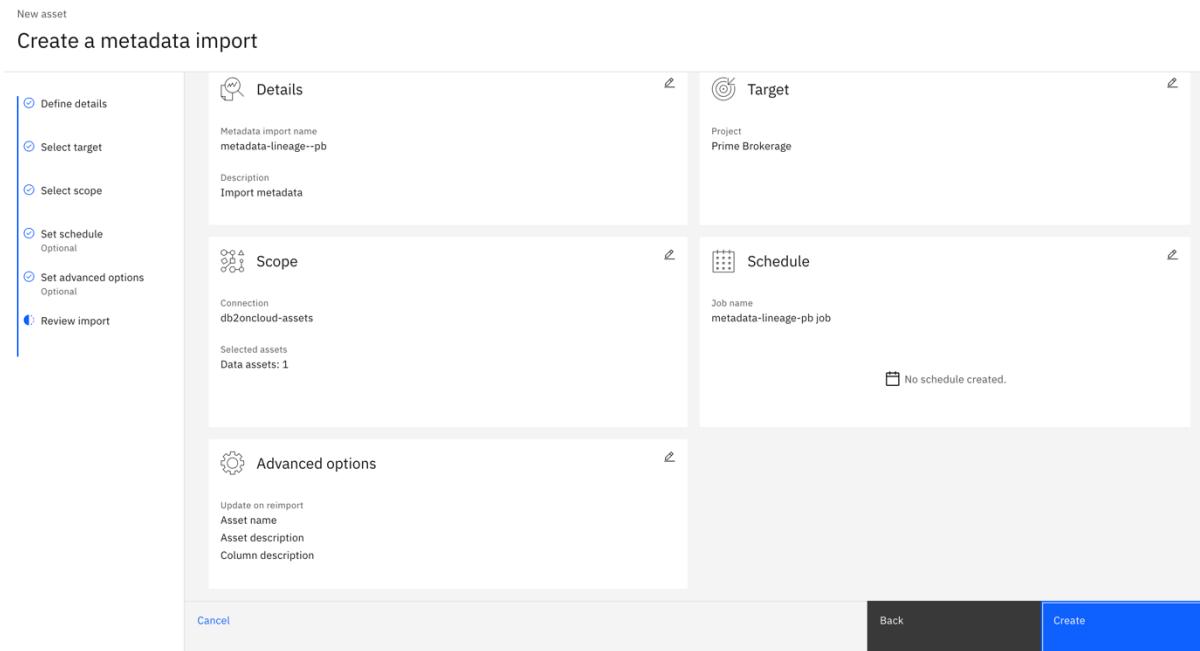
By default, no assets are deleted from the target project or catalog when you rerun the import. If you want the following types of assets to be deleted, select the respective check box.

Asset not found in the data source

Asset removed from the import scope

Cancel Back Next

8. Click **Create** to run the MDI.



Projects / Prime Brokerage /

✓ Metadata import complete. 10 assets were processed successfully. x

metadata-discover-pb Edit metadata import Reimport assets

Metadata import

Imported assets  
10 assets

Find assets

<input type="checkbox"/>	Name	Asset type	Format	Context	Last imported	Status
<input type="checkbox"/>	trade_sum_trans2	Data	Relational table	prime_brokerage/trade_sum_t...	Mar 24, 2023, 11:57 AM	<span>✓ Imported</span>
<input type="checkbox"/>	trade_sum_trans3	Data	Relational table	prime_brokerage/trade_sum_t...	Mar 24, 2023, 11:57 AM	<span>✓ Imported</span>
<input type="checkbox"/>	depot_position	Data	Relational table	prime_brokerage/depot_positi...	Mar 23, 2023, 01:29 PM	<span>✓ Imported</span>
<input type="checkbox"/>	depot_pos_trans1	Data	Relational table	prime_brokerage/depot_pos_t...	Mar 23, 2023, 01:29 PM	<span>✓ Imported</span>
<input type="checkbox"/>	trade_sum_qty	Data	Relational table	prime_brokerage/trade_sum_...	Mar 23, 2023, 01:29 PM	<span>✓ Imported</span>
<input type="checkbox"/>	trade	Data	Relational table	prime_brokerage/trade	Mar 23, 2023, 01:29 PM	<span>✓ Imported</span>
<input type="checkbox"/>	pending_trade	Data	Relational table	prime_brokerage/pending_tr...	Mar 23, 2023, 01:29 PM	<span>✓ Imported</span>
<input type="checkbox"/>	pending_trade_trans1	Data	Relational table	prime_brokerage/pending_tr...	Mar 23, 2023, 01:29 PM	<span>✓ Imported</span>
<input type="checkbox"/>	customer_profile	Data	Relational table	prime_brokerage/customer_pr...	Mar 23, 2023, 01:29 PM	<span>✓ Imported</span>
<input type="checkbox"/>	customer_settled_stocks_sum	Data	Relational table	prime_brokerage/customer_s...	Mar 23, 2023, 01:29 PM	<span>✓ Imported</span>

\*\*\* Now Repeat the metadata import steps for PostgreSQL for all tables under '**PRIME\_BROKERAGE**' Schema from step 9 to step 12. create name in a way so you recognise it is for postgredb like for eg- (metadata-discover-postgre)

Below snapshot to select all tables of **PRIME\_BROKERAGE** schema

The screenshot shows the 'Set scope' dialog with three main sections:

- Connections:** A list of available connections: Db2OnCloud, postgdb, Postgredb\_PB, and postgresql.
- Assets:** A list of tables under the 'prime\_brokerage' schema. One item, 'prime\_brokerage', is checked.
- Selected assets:** A summary of the selected asset, including its type (Schema), path (/prime\_brokerage), connection name (Postgredb\_PB), connection owner (admin), and connection datasource type (PostgreSQL (postgresql)).

At the bottom right are 'Cancel' and 'Select' buttons.

9. After MDI completes, click the **Project name** to return to the project assets page.

The screenshot shows the 'Assets' tab of the Project Assets page. On the left, there's a sidebar with '20 assets' and categories: Data access (5), Connections (3), Metadata imports (2), and Data (12). The main area displays a table of assets:

Name	Last modified	Actions
metadata-discover-pb	9 minutes ago	⋮
Metadata import	System	
trade_sum_trans2	9 minutes ago	⋮
application/x-ibm-rel-table	System	
customer_settled_stocks_sum	9 minutes ago	⋮
application/x-ibm-rel-table	System	
customer_profile	9 minutes ago	⋮
application/x-ibm-rel-table	System	

At the top right, there are 'Import assets' and 'New asset' buttons.

# METADATA ENRICHMENT

Metadata enrichment automatically provides context, content, and quality information in a unified framework. Data stewards create asset profiles to understand the meaning of data and to assess its quality. They also add business context to data by assigning terms. Metadata enrichment automates this process thus increasing the data steward's productivity.

By accepting in bulk, the recommended business terms and data classes that have a high ML accuracy (confidence score), using a user-defined confidence threshold, data stewards can on-board much higher volumes of data because manual reviews and approvals of the assigned business metadata is limited to only the assets where the ML confidence rating is low.

## Metadata Enrichment Overview

### 10. Click New asset → Metadata Enrichment.

#### New asset

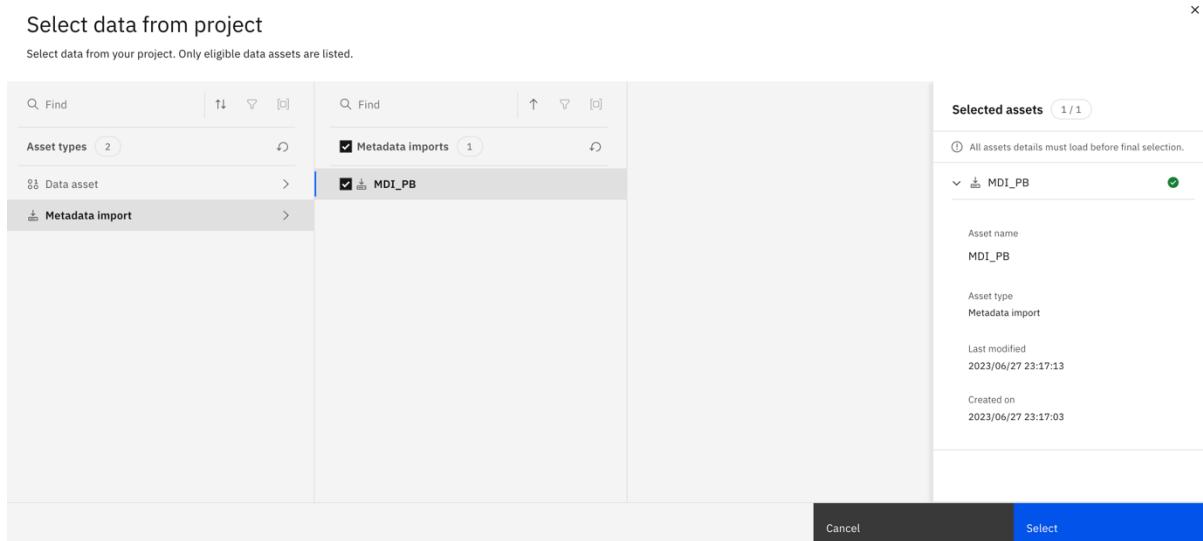
Select the tool to create an operational or configuration asset.

The screenshot shows a user interface for selecting a tool to create a new asset. On the left, a sidebar lists categories: 'Tool type' (selected 'All types'), 'Automated builders' (selected), 'Graphical builders', 'Code editors', 'Component editors', and 'Data access tools'. The main area is titled 'Find tools by name or description'. Under 'Automated builders', there are two cards: 'AutoAI' (described as automatically analyzing tabular data) and 'Metadata enrichment' (described as enriching imported asset metadata with business context, data profiling, and quality assessment). A red arrow points to the 'Metadata enrichment' card. Below this section, there is a 'Graphical builders' section with four cards: 'Dashboard editor', 'Data Refinery', 'Data privacy', and 'DataStage'.

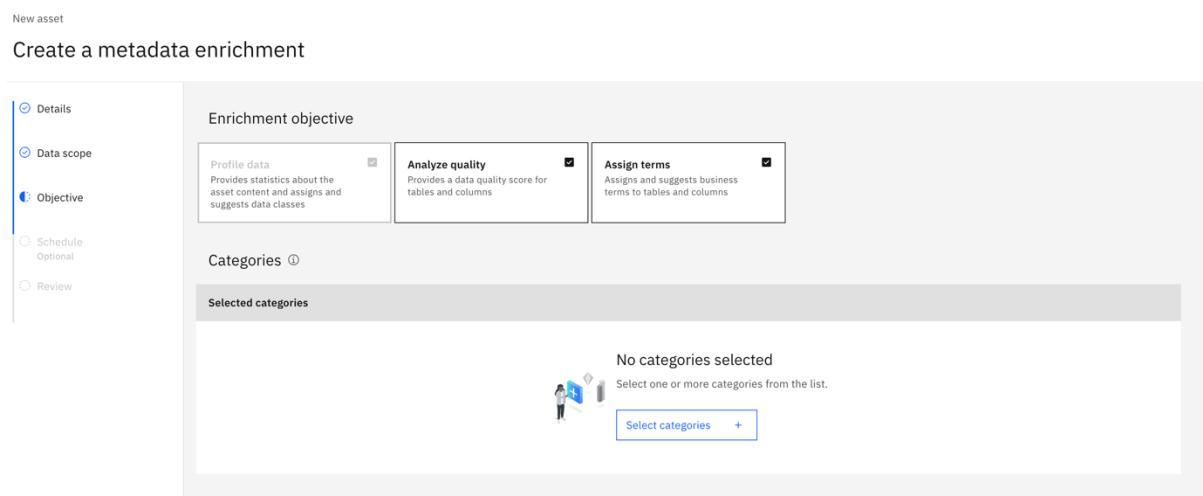
Tool Type	Description
Automated builders	Metadata enrichment
Graphical builders	Dashboard editor, Data Refinery, Data privacy, DataStage
Code editors	
Component editors	
Data access tools	

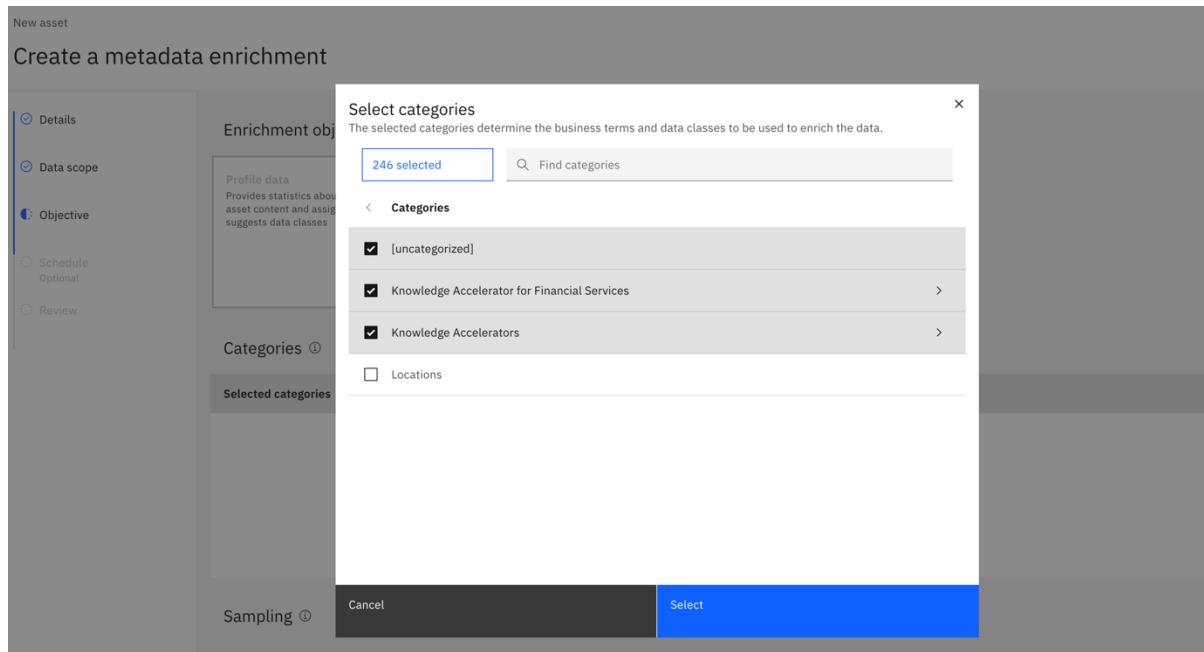
11. Enter a name for the MDE then click Next.

12. Click **Select data from project** then select the previously created MDI. Click **Select**.

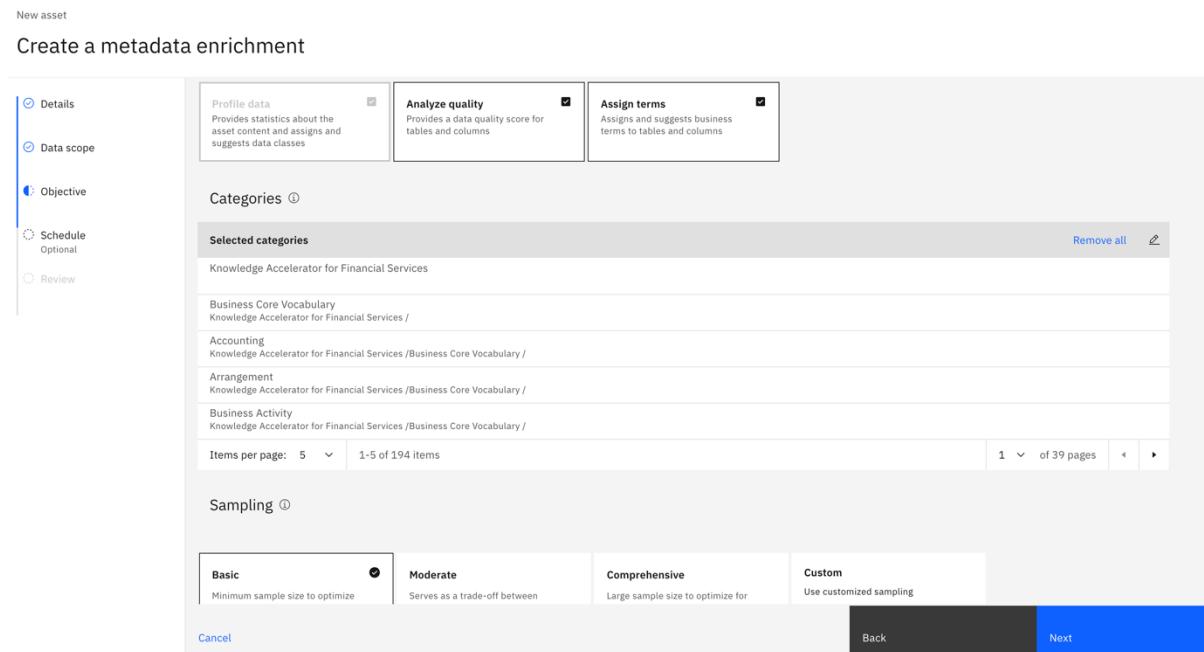


13. Click **Select categories** which is the KA and defaults populated in our pre-reqs.





14. Leave the sampling size as basic.



**15. Click Next.** Leave the schedule and data scope as is then click **Next**

New asset

### Create a metadata enrichment

Details

Schedule enrichment job

Metadata import schedules

Job name  
metadata-enrichment-pb

Schedule off

Data scope of reruns  
 New or modified assets ①  
 All data assets

[Cancel](#) [Back](#) [Next](#)

**16. Click Create.**

New asset

### Create a metadata enrichment

Details

Data scope

Objective

Schedule

Review

**Details**

Metadata enrichment name  
metadata-enrichment-pb.

Description  
—

Tags  
—

**Objective**

Enrichment options  
Profile data | Analyze quality | Assign terms

Categories  
Knowledge Accelerator for Financial Services | Business Core Vocabulary, 192 more

Sampling  
From top: Basic

**Data scope**

Structured data  
1 data asset

**Schedule**

Job name  
metadata-enrichment-pb

No schedule created

Data scope of reruns  
All data assets

[Cancel](#) [Back](#) [Create](#)

17. It will take sometime to assign the Business terms and data classes to your data .After MDE completes, review and **publish the assets** to the catalog you created. (Prime Brokerage Catalog)

metadata-enrichment-pb

Assets (10) Columns

(10) items selected More ▾ Publish Enrich ▾ | Cancel

<input checked="" type="checkbox"/> Assets	Source	Business terms	Data quality	Review status ⓘ	Enrichment status ⓘ	Publish status
<input checked="" type="checkbox"/> customer_profile	new-onpremvm... / prime_br...	- 7 suggested	● 99%	○	● Finished Mar 24, 2023, 2:57 PM	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> customer_settled_stocks_sum	new-onpremvm... / prime_br...	- 2 suggested	● 100%	○	● Finished Mar 24, 2023, 2:57 PM	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> depot_pos_trans1	new-onpremvm... / prime_br...	- 1 suggested	● 100%	○	● Finished Mar 24, 2023, 2:57 PM	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> depot_position	new-onpremvm... / prime_br...	- 1 suggested	● 100%	○	● Finished Mar 24, 2023, 2:57 PM	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> pending_trade	new-onpremvm... / prime_br...	-	● 100%	○	● Finished Mar 24, 2023, 2:57 PM	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> pending_trade_trans1	new-onpremvm... / prime_br...	-	● 97%	○	● Finished Mar 24, 2023, 2:57 PM	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> trade	new-onpremvm... / prime_br...	Trader	● 100%	○	● Finished Mar 24, 2023, 2:57 PM	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> trade_sum_qty	new-onpremvm... / prime_br...	-	● 100%	○	● Finished Mar 24, 2023, 2:57 PM	<input checked="" type="checkbox"/>

Publish to catalog

Select catalog

Catalog name	Your role	Created
Finance Catalog	Admin	2 mo ago admin (You)
Default Catalog	Admin	3 mo ago admin (You)
<b>Prime Brokerage Catalog</b>	Admin	<b>3 mo ago</b> admin (You)
Demo Catalog	Admin	3 mo ago admin (You)
Platform assets catalog	Admin	3 mo ago admin (You)

Cancel      Back      Next

Publish to catalog

Review assets

If any of the following assets already exist in the catalog, they will be combined.

Asset name	Source
customer_profile	new-onpremvm... / prime_br...
customer_settled_stocks_sum	new-onpremvm... / prime_br...
depot_pos_trans1	new-onpremvm... / prime_br...
<b>depot_position</b>	<b>new-onpremvm... / prime_br...</b>
pending_trade	new-onpremvm... / prime_br...
pending_trade_trans1	new-onpremvm... / prime_br...
trade	new-onpremvm... / prime_br...
trade_sum_qty	new-onpremvm... / prime_br...
trade_sum_trans2	new-onpremvm... / prime_br...
trade_sum_trans3	new-onpremvm... / prime_br...

Cancel      Back      Publish

18. Wait a few minutes for the assets to be published. Next, we will run MDI to get the lineage.

19. Return to the project assets page. Click **New asset -> Metadata import**.

The screenshot shows the 'New asset' creation interface in the CP4D application. At the top, there is a breadcrumb navigation bar with the path: Projects / Prime Brokerage /. A red arrow points to the 'Prime Brokerage' part of the path. Below the navigation bar, the title 'metadata-enrichment-pb' is displayed. The main area is titled 'New asset' and contains a sub-instruction: 'Select the tool to create an operational or configuration asset.' On the left, a sidebar titled 'Tool type' lists categories: All types, Automated builders, Graphical builders, Code editors, Component editors, and Data access tools. The 'All types' category is currently selected. The main content area displays various asset types with their descriptions and icons. These include:

- Federated Learning**: Jupyter notebook editor. Description: Create a federated learning experiment to train a common model on a set of remote data sources. Share training results without sharing data.
- Component editors**:
  - Data quality definition**: Create abstract rule logic for data analysis that can be used in any number of data quality rules.
  - Data quality rule**: Create rules to assess the quality of your data by evaluating and validating specific conditions.
  - DataStage component**: Create reusable components for DataStage flows, such as, subflows, libraries, and data definitions.
  - Parameter set**: Collect multiple job parameters with specified values to reuse in jobs.
- Data access tools**:
  - Connected data**: Data in an external data source that is accessed through a connection.
  - Connection**: Supply the information necessary to connect to a data source.
  - Metadata import**: Import asset metadata from a connection into a project or a catalog.
  - Model**: Add an existing PMML (predictive model markup language) file (.xml) from your local system as a model.

At the bottom left of the main content area, there is a 'Show descriptions' button with a circular icon containing a question mark.

## METADATA IMPORT - LINEAGE

- Select **Get Lineage** as the MDI goal.

New asset

### Create a metadata import

**Define goal**

What is the intended goal of your import?

**Discover**

Import and view assets of various types in a project or a catalog. If you import assets into a project, you can also set up and run metadata enrichment for those data assets.

Import can include  
Technical metadata  
Access to data

**Get lineage**

Add information to a catalog about where your data comes from, how it changes, and where it moves over time.

Import can include  
Technical metadata  
Lineage metadata

- Click **Next**. Enter a name and description. Select **your catalog** as the target. Click **Next**.

Projects / Prime Brokerage /

### Create a metadata import

**Define details**

Specify some basic information about your metadata import to make it easy to identify.

Name

Description (optional)

Tags (optional)  
Add tags to make assets easier to find.

**Next**

3. Choose the previously created PostgreSQL connection.

metadata-lineage-pb

### Create a metadata import

Define details

- Define details
- Select target
- Select scope
- Set schedule  
Optional
- Set advanced options  
Optional
- Review import

**Selected assets**  
1 assets

Find assets

Name	Type	Context
new-onpremvm-postgre...	connection	

[Cancel](#)
[Back](#)
Next

metadata-lineage-pb

### Create a metadata import

Define details

- Define details
- Select target
- Select scope
- Set schedule  
Optional
- Set advanced options  
Optional
- Review import

**Details**

Metadata import name  
**metadata-lineage-pb**

**Scope**

Connections  
**new-onpremvm-postgresql**

Selected assets  
Data assets: 0

**Advanced options**

Update on reimport  
Asset name  
Asset description  
Column description

**Target**

Catalog  
**Prime Brokerage Catalog**

**Schedule**

Job name  
**metadata-lineage-pb job**

No schedule created.

[Cancel](#)
[Back](#)
Create 

## VIEW POSTGRESQL LINEAGE IN THE CATALOG

4. After MDI completes (2-5 minutes), navigate to **your catalog**.

The screenshot shows the 'Assets' tab of the Prime Brokerage Catalog. At the top, there are filter options for 'Any asset type', 'Any source', 'Business term', 'Any tag', and a 'Clear all' button. Below the filters, there are three tabs: 'Recently added' (selected), 'Recommended', and 'Highly rated'. The main area displays a grid of data assets:

- max\_trade**: Data asset, 4 columns, owner: admin, added: Mar 24, 2023 12:04 PM, tags: prime\_brokerage.
- premium\_customer**: Data asset, 4 columns, owner: admin, added: Mar 24, 2023 12:04 PM, tags: prime\_brokerage.
- pending\_trade\_trans1**: Data asset, 4 columns, owner: admin, added: Mar 24, 2023 12:04 PM, tags: prime\_brokerage.
- customers\_with\_cash\_trade**: Data asset, 4 columns, owner: admin, added: Mar 24, 2023 12:04 PM, tags: prime\_brokerage.
- trade\_sum\_trans3**: Data asset, 9 columns, owner: admin, added: Mar 24, 2023 12:04 PM, tags: prime\_brokerage.
- trade\_**: Data asset, 4 columns, owner: ac, added: Mar 24, 2023 12:04 PM, tags: prime\_brokerage.

Each asset entry includes a preview icon, a 'Show details' button, and a 'Reviews' section with 0 reviews.

5. Search for the **trade\_sum\_trans3**.  
 6. Click the **Lineage** tab. Here we can view the **business data lineage**.

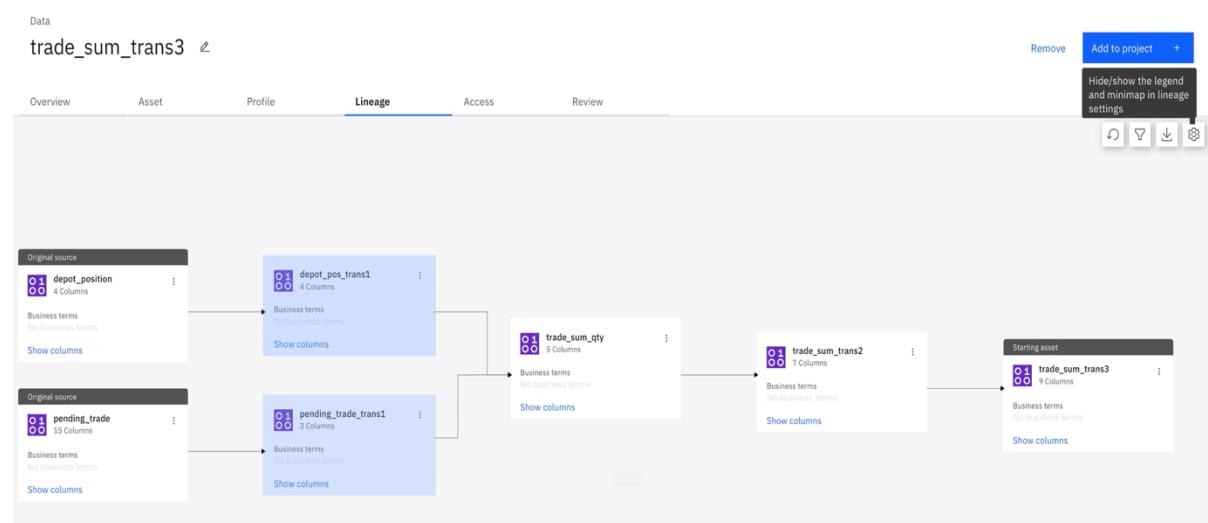
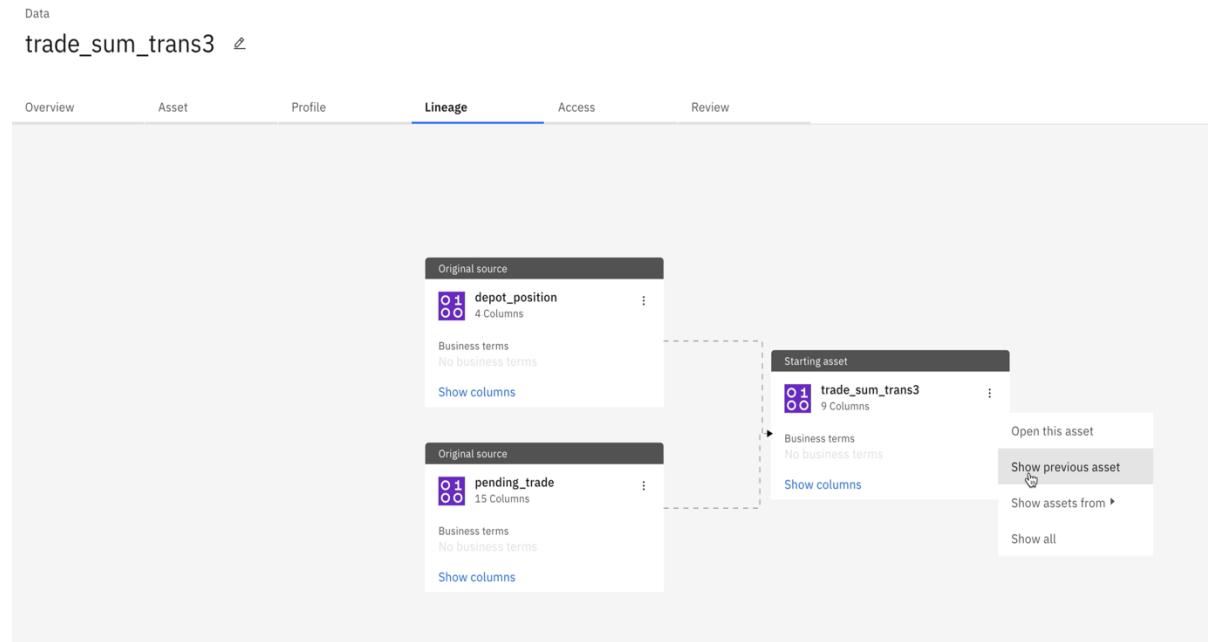
The screenshot shows the details page for the **trade\_sum\_trans3** asset. The top navigation bar includes 'Overview', 'Asset', 'Profile', 'Lineage' (which is selected and highlighted in blue), 'Access', and 'Review'.

The 'Lineage' tab displays the business data lineage. It shows two nodes connected by a dashed arrow:

- Original source**: **depot\_position**, 4 Columns. It has a 'Show columns' button.
- Starting asset**: **trade\_sum\_trans3**, 9 Columns. It has a 'Show columns' button.

Between these nodes is a dashed arrow pointing from left to right, indicating the flow of data from the original source to the starting asset.

7. The complete Business Lineage can be seen by clicking the ellipses -> Show Previous asset.

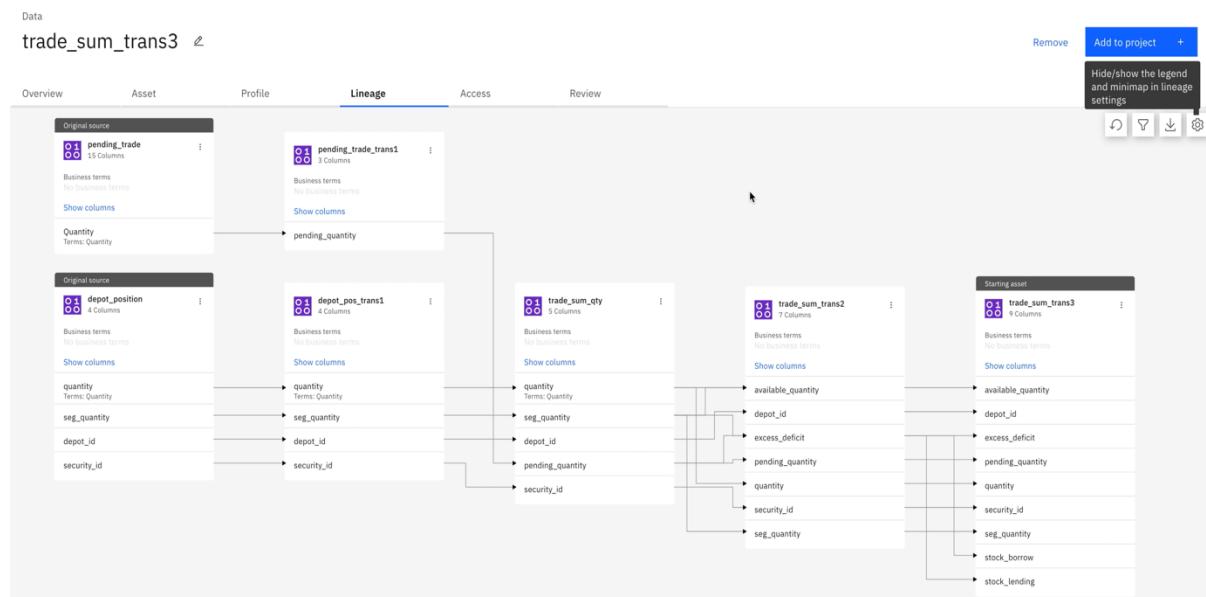


8. We can see the **trade\_sum\_trans3** table is derived from the **depot\_position** and **pending\_trade** transformation script.

9. Click **show columns** in the **trade\_sum\_trans3** table. Select all columns then click **Show columns**.

### Show columns

Select columns to show in the lineage graph.



10. We can see the derivation path of the **quantity** and **pending quantity** columns from the **pending trade script** and the **depot\_position** script to their **transformed table**, and finally to the target tables: **trade\_sum\_trans3**.

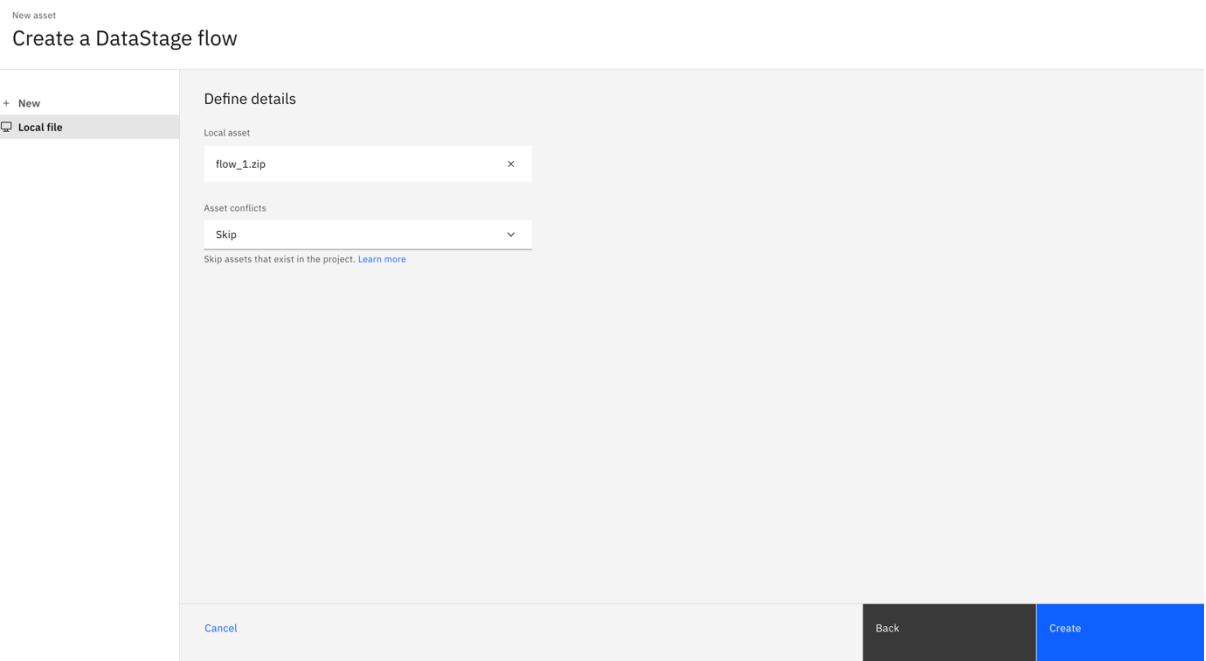
## DATA STAGE FLOWS IMPORT.

YOU WILL FIND THE DATASTAGE FLOWS IN ZIP FORMAT:

1. <https://github.ibm.com/poholkar/Banking-PRIME-BROKERAGE>
2. Click the **Assets** tab, then **New Asset**.
3. Click **New asset à DataStage**.

The screenshot shows the 'New asset' dialog in Watson Studio. On the left, a sidebar lists 'Tool type' categories: All types, Data access tools, Automated builders, Graphical builders, Code editors, and Component editors. The 'All types' category is selected. In the main area, a message says 'Connected data moved. To add connected data, close this dialog and click Import assets.' Below this is a search bar with placeholder text 'Find tools by name or description'. A red arrow points to the 'DataStage' tool icon, which is part of the 'Graphical builders' section. Other tools shown include 'Dashboard editor', 'Data Refinery', and 'Pipelines'. At the bottom, there's a 'Code editors' section.

4. Upload the DataStage 'Flow\_1' from the local file in zip format you download from the box link.



5. Repeat the procedure and upload the DataStage ‘Flow\_final’ from the local file in zip format you download from the box link.
6. Once uploaded, go back to assets tab and the flows should get reflected.

Projects / Prime Brokerage

Overview Assets Jobs Manage

Find assets Import assets +

23 assets

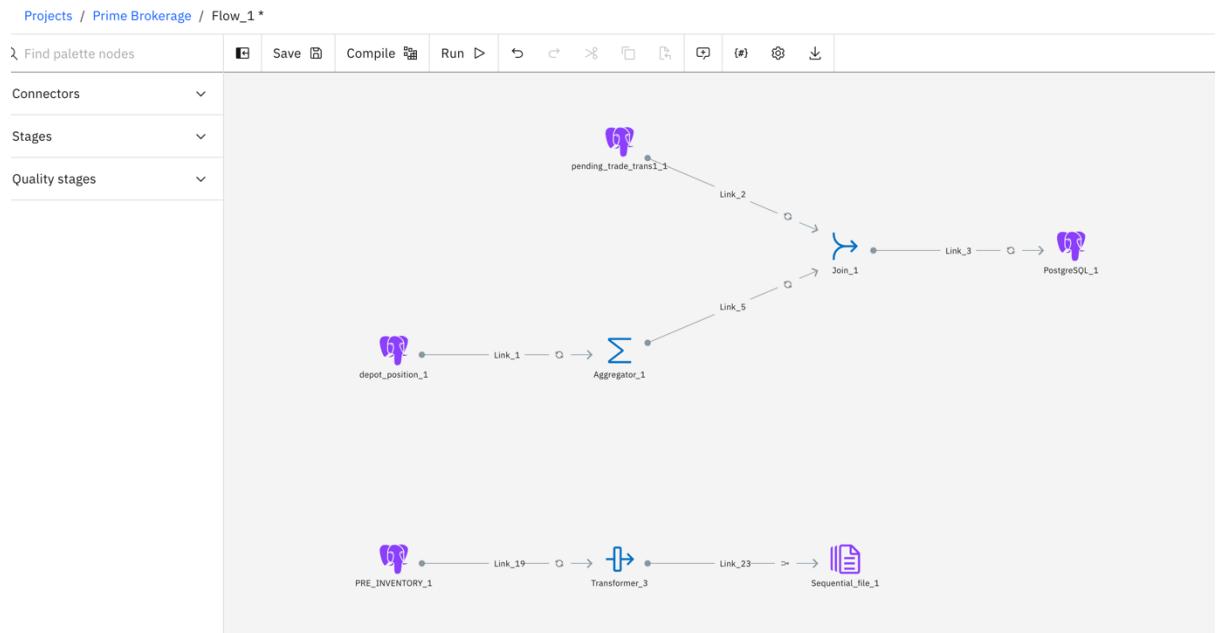
All assets

Asset types

- > Data access 6
- > Data 13
- < Metadata enrichment 2
  - Metadata Enrichments 2
- > Flows 2

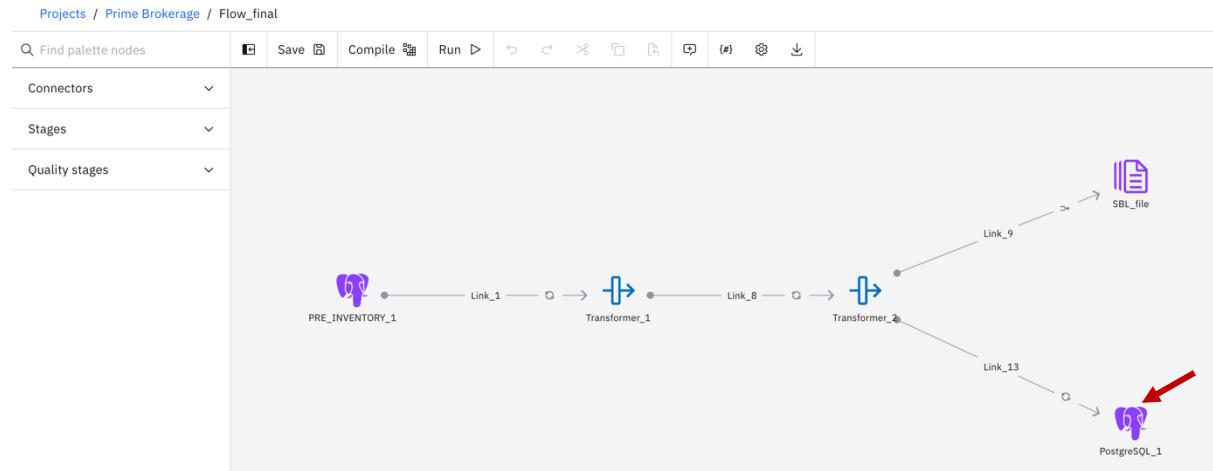
Name	Last modified	⋮
Flow_1 DataStage flow	1 month ago Rabin德拉 Panda	⋮
Flow_final DataStage flow	1 month ago Rabin德拉 Panda	⋮

7. Click Flow\_1 -> Run

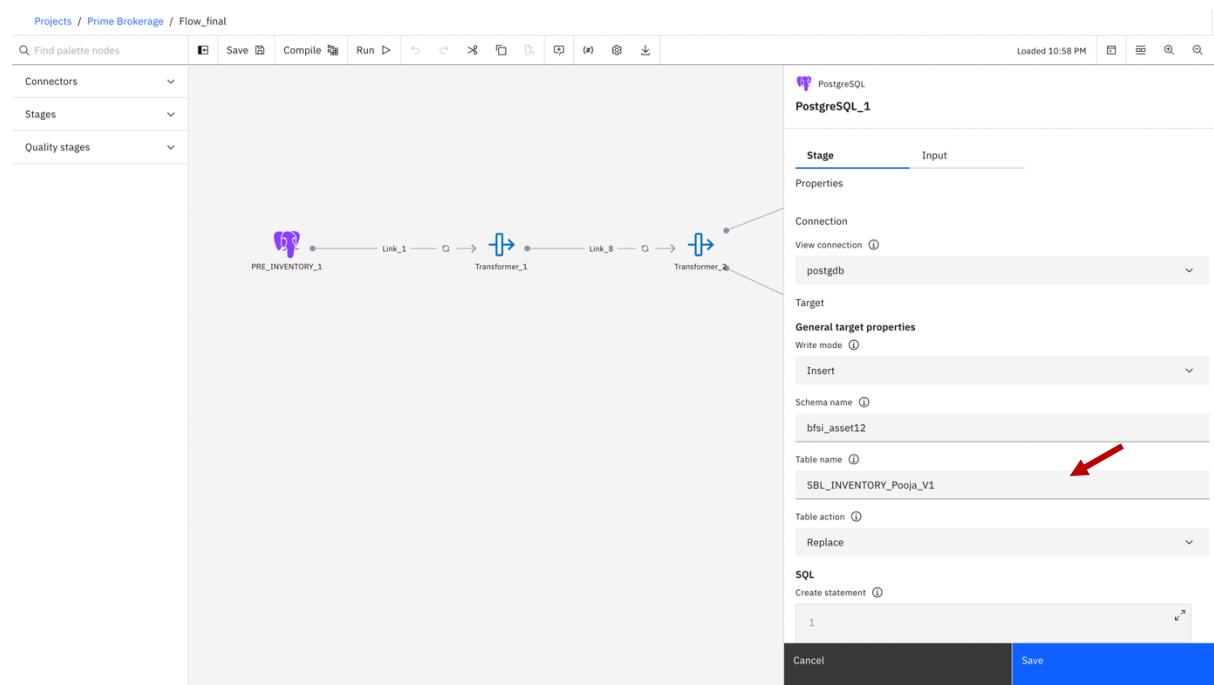


8. Go back to assets page, click Flow\_final

9. Now, we will modify the flow to have each candidate create their separate SBL file. Double-click “PostgreSQL\_1” task as highlighted below.



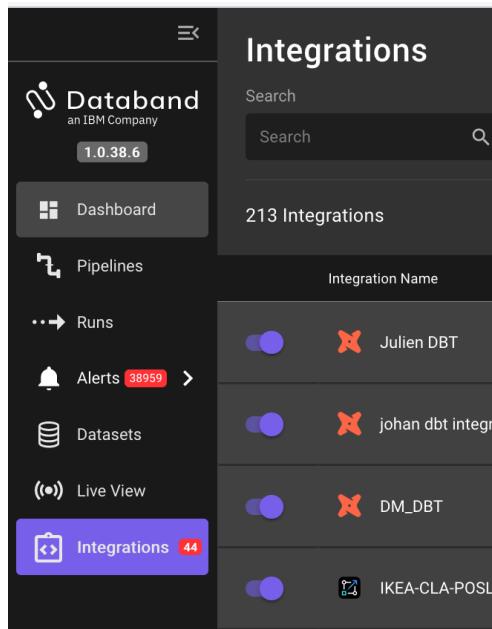
10. Under the Table name create new SBL file with name as SBL\_INVENTORY\_(Yourinitials)\_V1 and save.



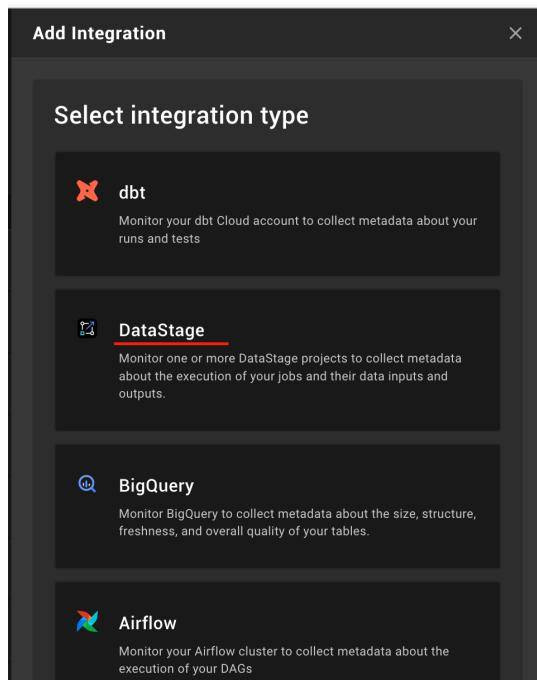
The Final Flow Creates the Stock Brokerage and Lending file (SBL) which we will push in Postgres database for further Business analysis.

## CONFIGURE DATASTAGE WITH DATABAND

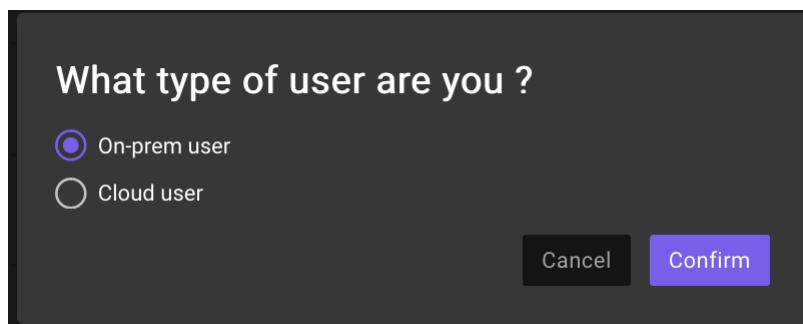
1. Login to <https://ibm-bp-demo.databand.ai/> with provided userids
2. Go to Integrations tab.



3. Click on Add Integration (top right corner) to add Cloud Pak for Data integration and choose DataStage from the list.



4. Choose On-prem user.

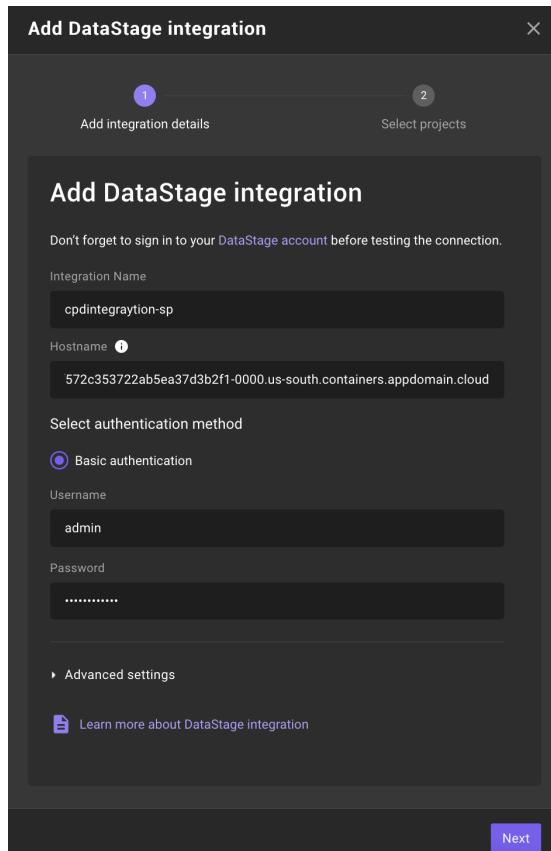


5. Provide input details and click Next.

Integration name – please add your name in the integration name to uniquely identify your integration.

Hostname – Cloud Pak for Data URL

Username – Username of Cloud Pak for Data console



- Select your project from the list that you created in the above step (Ex: Prime brokerage) and Click on Save

Select DataStage projects			
1 project selected			
	Project	Creator	Source
<input type="checkbox"/>	aws-s3-wkc-project	admin	CP4D-BP
<input type="checkbox"/>	MLops	admin	CP4D-BP
<input type="checkbox"/>	Demo Project	admin	CP4D-BP
<input type="checkbox"/>	Banking Project Discover	admin	CP4D-BP
<input type="checkbox"/>	git_project	admin	CP4D-BP
<input type="checkbox"/>	StreamingData	admin	CP4D-BP
<input type="checkbox"/>	PyDSSample	admin	CP4D-BP
<input checked="" type="checkbox"/>	Prime Brokerage Old	admin	CP4D-BP
<input type="checkbox"/>	BANKING	admin	CP4D-BP
<input type="checkbox"/>	MyProject	admin	CP4D-BP

7. Go to Cloud Pak for Data console and run DataStage flows at least 5 times as mentioned in step (mention exact step number). This will help Databand to create baseline for the pipeline that we want to observe.
8. Once DataStage flow runs are successful, come back to Databand UI. Navigate to Pipelines tab.

The screenshot shows the Databand UI interface. The left sidebar has a purple header 'Pipelines'. Other menu items include 'Dashboard', 'Runs', 'Alerts (38973)', 'Datasets', 'Live View', and 'Integrations (45)'. The main area is titled 'Pipelines' and displays a list of 1187 pipelines. A search bar is at the top. The list includes entries like 'ETL - ServiceNow Incident table.DataStage job' and 'cdp-demo\_dev\_prepare\_cu...'. Each entry has a checkbox, a preview icon, a name, a source, a project, and a run status.

9. From the Projects dropdown, filter the project that you created for DataStage flows. This is list only flows which are part of that project. Observe the information that it brings for all the pipelines.

This screenshot shows the same Databand UI as above, but with a dropdown menu open under 'Project'. The 'Prime' project is selected. The list of pipelines is now filtered to show only those belonging to the 'Prime Brokerage' project. The columns remain the same: Name, Source, Project, and Run Status.

Name	Source	Project	Run Status
cdp-demo_dev_custom_dag	airflow	Prime Brokerage	19 Failed, 1 Running, 65 Successful
ETL - ServiceNow Incident table.DataStage job	1000331010	CIO Operational DataHub - Development	1 Failed, 14 Failed, 1319 Successful
ETL - ServiceNow sys_update_group table.DataStage job	1000331010	CIO Operational DataHub - Development	1242 Successful

- Click on the [Flow\\_final.DataStage job](#) pipeline to view its details. It will list all the previous runs of that pipelines that you executed in the step above along with their status. Observe the information that it brings for all the executions.

Name	Owner	Documentation	Project	Run States	Alert Definitions	Alerts	Teams	Last Run
'l_ Flow_final.DataStage job	1000330999	<a href="#">Add link</a>	Prime Brokerage	3 Failed 10 Success	<a href="#">3</a>	<span style="color: red;">12</span> <span style="color: yellow;">0</span>	No teams	Today at 05:57:11 PM >
'l_ Flow_1.DataStage job	1000330999	<a href="#">Add link</a>	Prime Brokerage	3 Success	<a href="#">1</a>		No teams	Mar 14, 2023 03:08:27 PM >
'l_ Flow_2.DataStage job	1000330999	<a href="#">Add link</a>	Prime Brokerage	1 Success	<a href="#">0</a>		No teams	Mar 10, 2023 01:16:59 PM >
'l_ InvFlow_p.DataStage job	1000330999	<a href="#">Add link</a>	Prime Brokerage	4 Success	<a href="#">0</a>		No teams	Mar 10, 2023 01:12:23 PM >

- Open any one of the successful runs to view more details about it.

Run Name	Status	Start Time	End Time
'l_ Flow_final.DataStage job_cc64cf43-25a8-4517-8d77-8	<span style="color: green;">✓</span> Runs	Apr 20, 2023 05:04:30 PM >	Apr 20, 2023 05:05:19 PM >
'l_ Flow_final.DataStage job_dfd8635a-89db-4391-8659-7	<span style="color: red;">✗</span>	Apr 20, 2023 05:01:54 PM >	Apr 20, 2023 05:02:23 PM >
'l_ Flow_final.DataStage job_72d168e6-2786-49ca-b38e-1	<span style="color: red;">✗</span>	Apr 20, 2023 03:35:02 PM >	Apr 20, 2023 03:35:31 PM >
'l_ Flow_final.DataStage job_3490585f-891c-437a-bbb9-c	<span style="color: green;">✓</span>	Apr 20, 2023 03:19:34 PM >	Apr 20, 2023 03:20:28 PM >

- Scroll through various tabs and understand the data that each tab shows.

The screenshot shows the Databand interface for a pipeline named "Flow\_final.Data...". The top navigation bar includes tabs for Details, Metrics, Logs, Code, Run Info, Data Interactions, and Histograms. The Metrics tab is currently selected. The left sidebar has sections for Dashboard, Pipelines, Runs (selected), Alerts (999+), and Datasets. The main content area displays "User Params" with "project\_name" set to "Prime Brokerage". Under "Running/Failed Tasks", it says "No data available". A "Tasks" section at the bottom right includes a search bar.

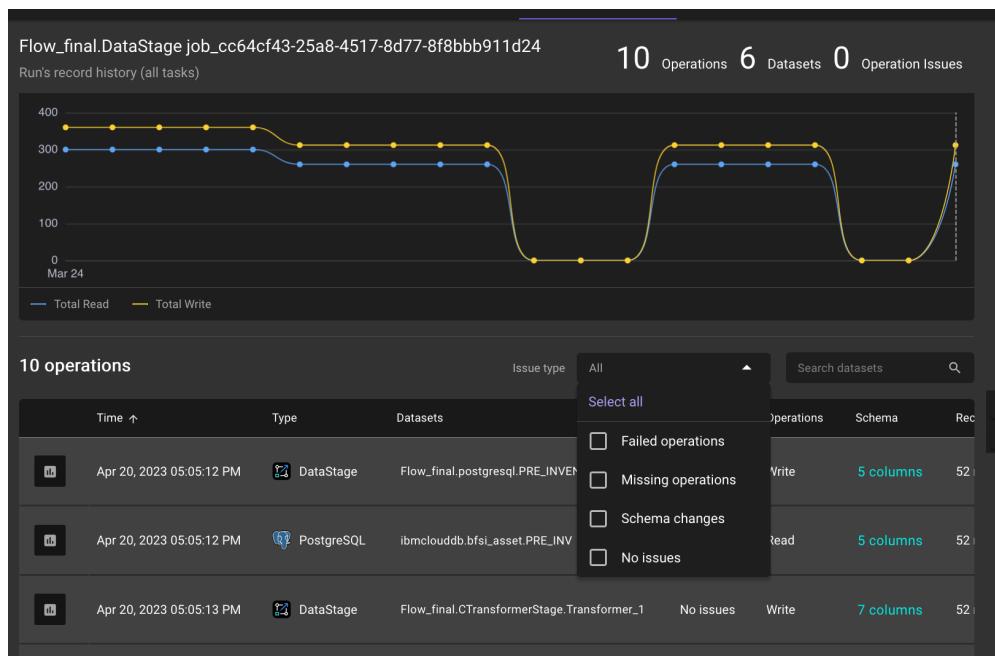
**Metrics** – It displays all the data being processed by each task within the pipeline, allowing for easy filtering and examination of task-specific information.

The screenshot shows the "Metrics" details view. It has filters for Metric type (User), Show (Select sources), and Tasks (PRE\_INVENTORY\_1). A search bar is also present. The main table lists metrics with columns for Key, Value, and Task. Two rows are shown:

Key	Value	Task
Flow_final.postgr...ENTORY_1_write_rows	52	PRE_INVENTORY_1
ibmclouddb.bfsi_...t.PRE_INV_read_rows	52	PRE_INVENTORY_1

At the bottom, it says "Records per page: 25" and "1-2 of 2". Below the table is a "Metrics Trends Chart" section.

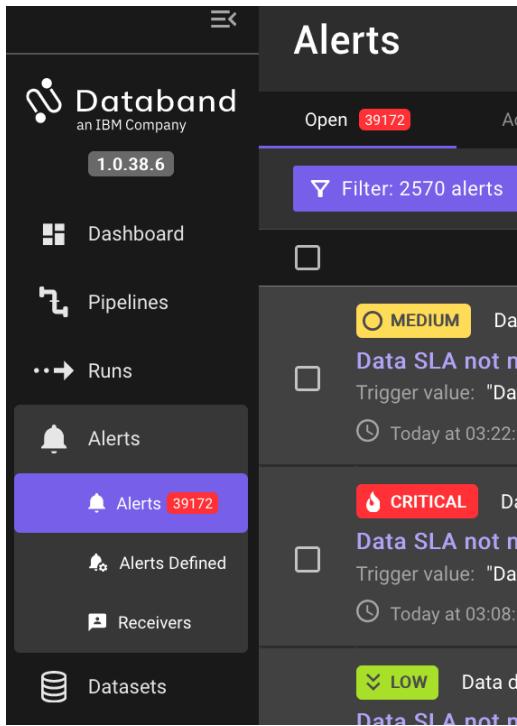
**Data Interactions** – The tab visually presents the data operations performed during pipeline execution, including read-write actions and schema changes. Additionally, users have the ability to filter and sort these operations based on any associated issues or errors.



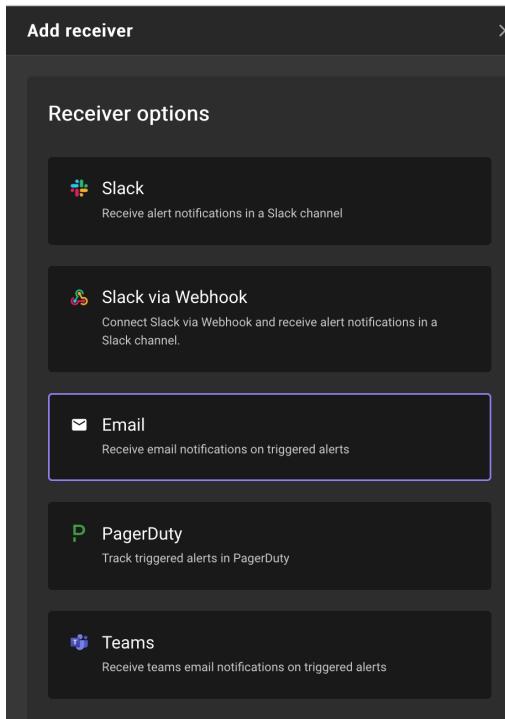
**Logs** – The log tab in the Databand UI presents a comprehensive record of all the logs generated by executed runs from DataStage. This valuable feature facilitates the identification of any potential issues that may have arisen during the execution process .

**Run Info** – Provides a concise summary of the entire execution process.

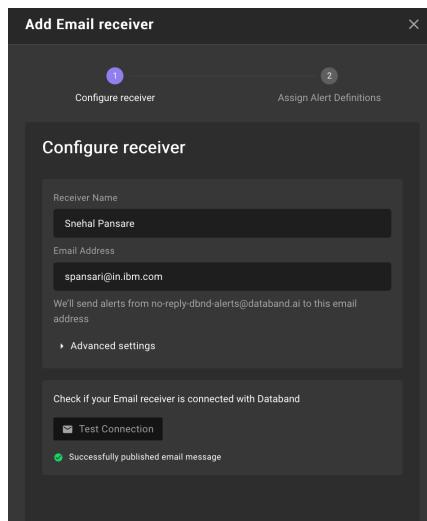
13. Navigate to Alerts tab. Here we will define alerts to generate notification.



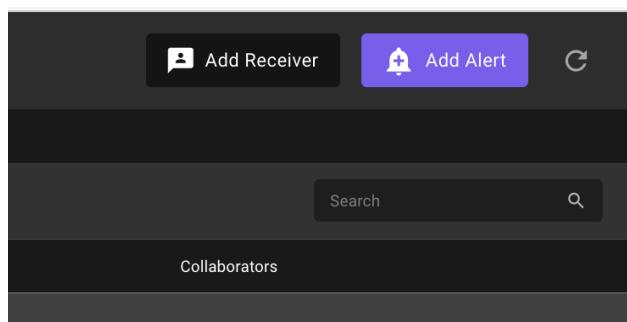
14. Click on Add receiver (top right corner) and choose Email from the list.



15. Give your name as email address to receive alerts when a condition is triggered. Click on “Test Connection” to verify that you are receiving email. Click on Next and Save on next page without selecting any Alert definition.



16. Click on Add Alert (top right corner) and click on Setup for “Pipeline state” from the list.



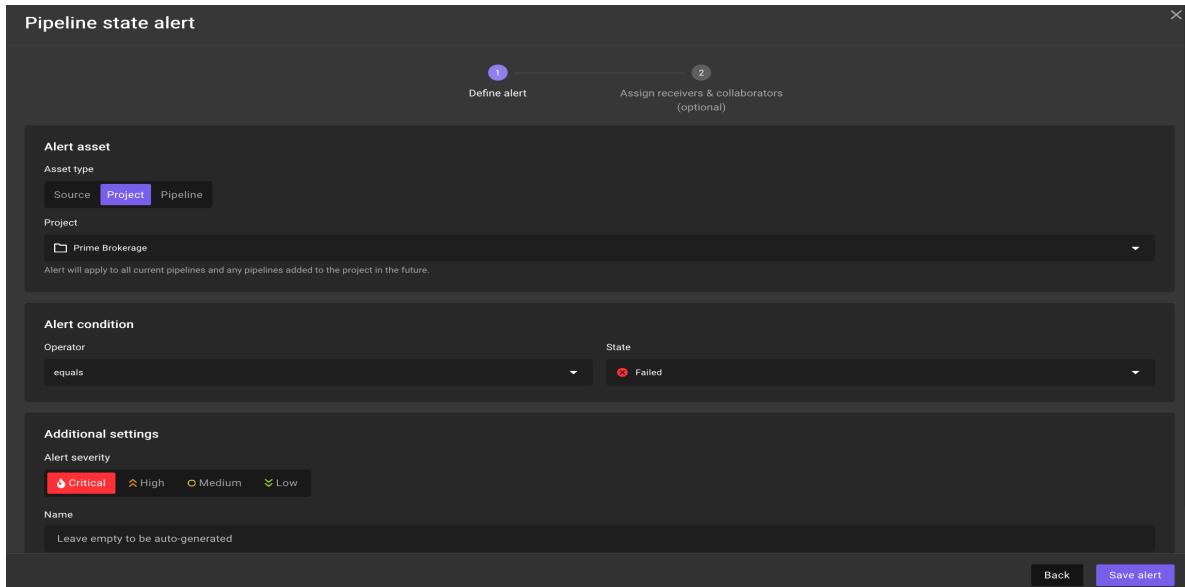
17. Define alert parameters as can be seen in the below screenshot.

Asset Type – Select Project

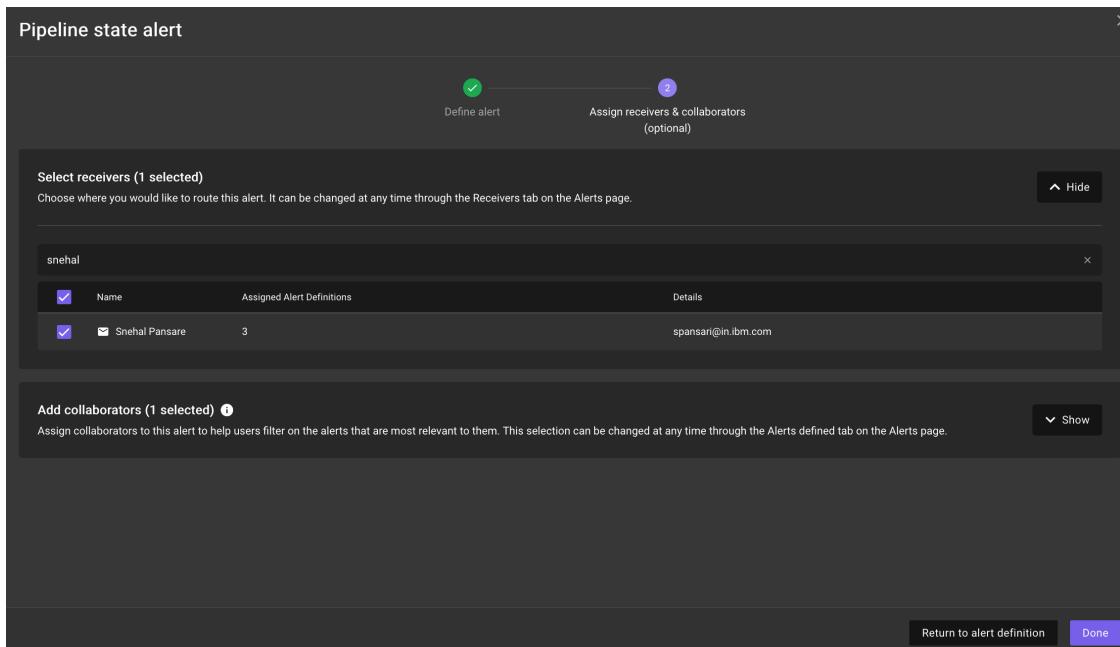
Project – Select Project that you configured in Databand for observing Datastage flows.

Alert Condition – equals to Failed state.

Alert severity - Critical



- Click on Save alert. On next page, click on “Select receivers” and filter by your name. Select checkbox and click on Done.



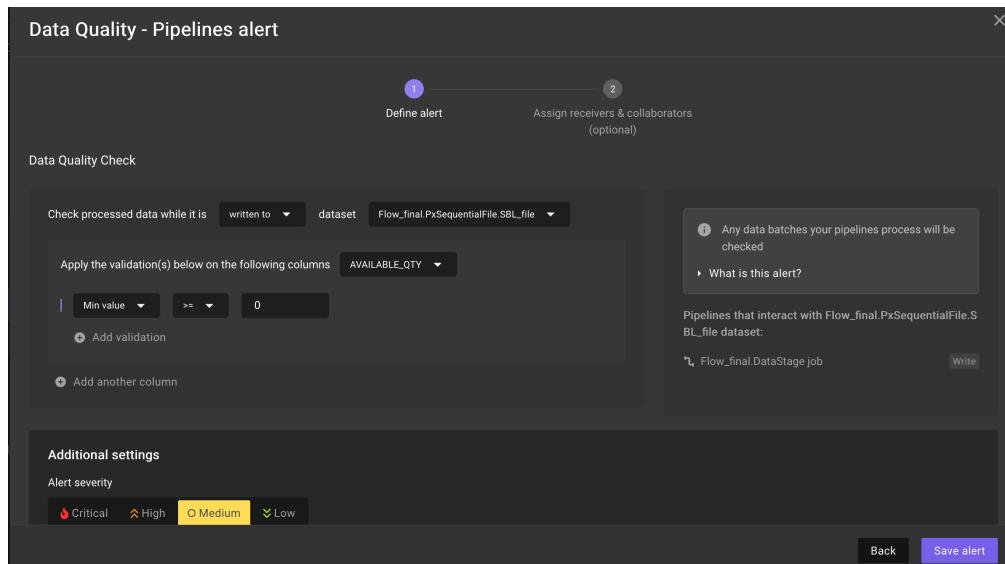
19. We will add one more alert for Custom metric. Click on “Add alert” again from top right corner and choose “Setup” for “Data Quality – Pipelines alert”.

20. Choose parameters as can be seen in below screenshot.

Dataset – choose dataset by filtering with “Flow\_final.PxSequentialFile.SBL\_file”

Columns – choose “AVAILABLE\_QTY” and select Apply

Condition – “Min value”  $\geq 1$



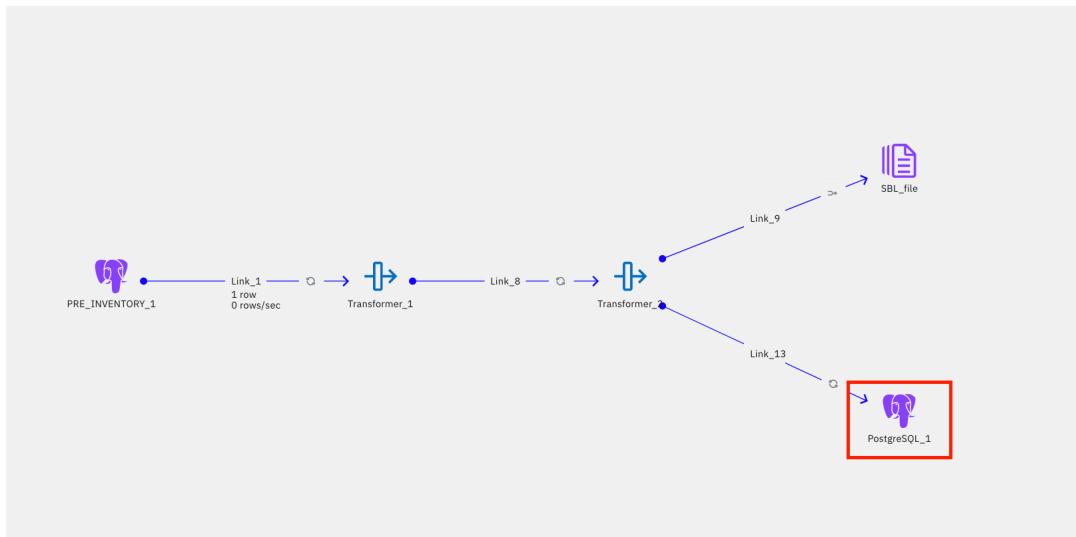
21. Click on Save alert. On next page, click on “Select receivers” and filter by your name. Select checkbox and click on Done.

22. Go back to Cloud Pak for Data Console and open the DataStage flow that you configured to observe using Databand.

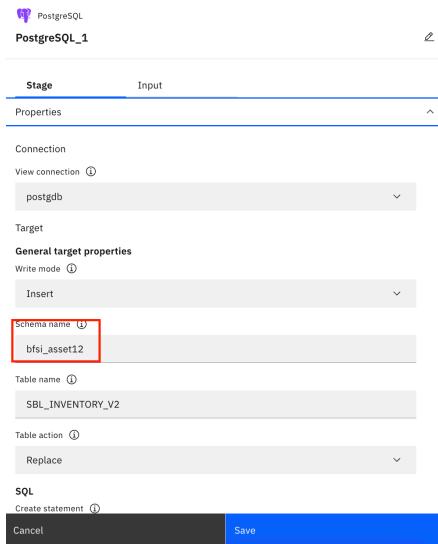
23. Execute the flow for couple of times.

24. Now, we will modify the flow to make it fail to trigger the alert that we just defined.

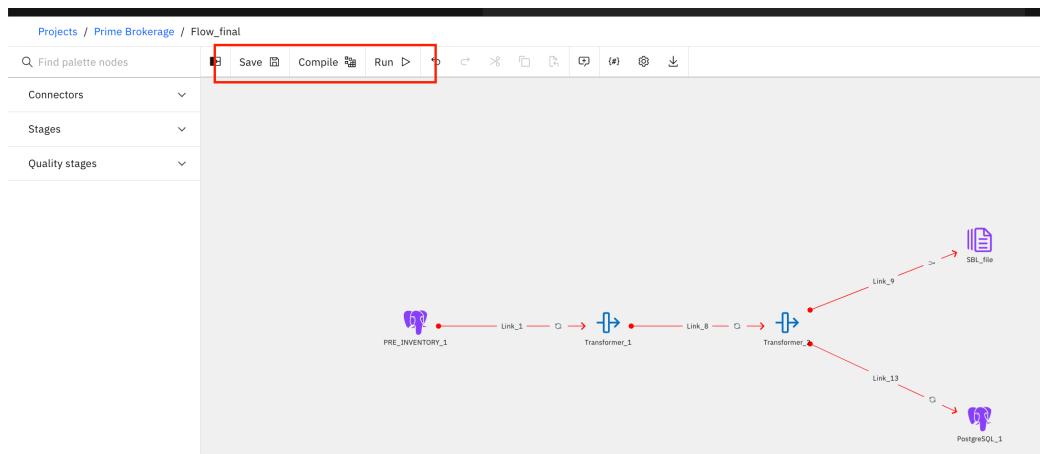
Double-click “PostgreSQL\_1” task as highlighted below.



25. From the panel opened on the right side, update “Schema name” to some random name. (as we want to fail the task). Click on Save.



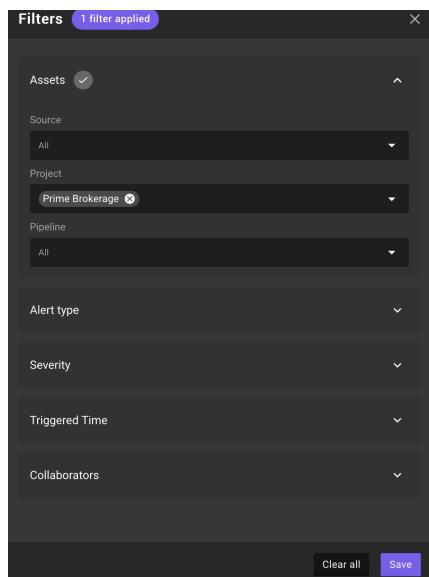
26. Click on “Save”. Click on “Compile” and then “Run” from the top left menu.



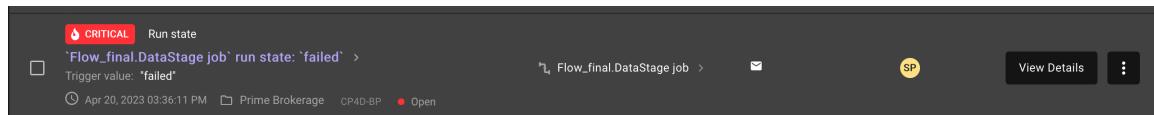
27. Wait until the pipeline fails.

28. Go back to Databand UI and navigate to “Alerts” tab.

29. Click on Filter. Expand Assets and filter Project by your project name as can be seen in below screenshot. Click on “Save”



30. You should see alert generated for the failed DataStage flow.



You should get an email notification.

A screenshot of an email alert. The subject line is 'CRITICAL | Run State Alert'. The body of the email starts with 'Pipeline name: Flow\_final.DataStage job'. Under 'Details:', there is a bulleted list of information: 'Description: 'Flow\_final.DataStage job' run state: 'failed'', 'Alert trigger time: UTC 2023-04-20 11:36:39', 'Alert trigger value: failed', 'Run start time: UTC 2023-04-20 11:31:54', 'Run name: Flow\_final.DataStage job dfd8635a-89db-4391-8659-7f68ccb370b2', 'Run triggered by: 1000330999', 'Project name: Prime Brokerage', and 'Source: ROCKSAng01'. Below the alert details, a message says 'You can see more information and resolve this alert on Databand:' followed by a purple button labeled 'See Alert Details'.

Happy Learning 😊