

# Data Fabric for Governance and Privacy

## Tutorial 1: Trust Your Data

### Background

Data assets can be enriched with information that helps users find data faster, to decide whether the data is appropriate for the task at hand, whether they can trust the data, and how to work with the data. Such information includes, for example, terms that define the meaning of the data, rules that document ownership or determine quality standards, or reviews.

Data stewards create asset profiles to understand the meaning of data and to assess its quality. Also, they add business context to data by assigning terms. Metadata enrichment automates this process thus increasing the data steward's productivity.

Data is useful only if its context, content, and quality are trusted. To keep it that way, data must continuously be evaluated and appropriate remediation be taken if required. Data stewards can configure recurring jobs to continuously track changes to the content and structure of data and then analyze only data that changed.

The information that is added to assets through metadata enrichment also helps to protect data because it can be used in data protection policies to mask data or to restrict access.

### Pre-requirements

#### Demo environment

To demo MDE you need one of the below environments:

- IBM Cloud SaaS account with WKC Professional or Enterprise plan (for now until timed trial is in place).
- IBM Cloud SaaS Dev account
- CPD Software (starting from v 4.5 CPD Software will support Metadata Enrichment)

You need to have access to the relevant Category in glossary and be able to create projects or at least have access to the project.

#### Relevant content in glossary

Relevant content in glossary is required to demo automatic term assignment.

For demo purposes, you can follow the next steps:

- Create a category called **“Banking”** and give yourself appropriate access permissions to it. Require “Manage governance categories” platform permission.
- Download the file below and import these terms into your environment (see IBM documentation if you do not know how to import terms):  
<https://ibm.box.com/s/1hhhqiemvbogfraksfc31o6i9s6x9omd> **“banking.csv”**. Require “Access governance artifacts” permission.
- You can also use Excel demo script:  
<https://ibm.box.com/s/1hhhqiemvbogfraksfc31o6i9s6x9omd> **“Import export business terms script.xlsx”** if you want to create your own set of business terms. This need to be then exported to csv and semicolon needs to be replaced by comas in text editor.

## Create a catalog

You can create a new catalog or use an existing catalog that you have access to.

To create a new catalog, open left hand side menu and click **View all catalogs** under **Catalogs** menu group.

Then click **Create Catalog**

IBM Cloud Pak for Data

All

Search

Buy

Michał Szylar's Account

MS

### New catalog

Name

Demo catalog – Data Fabric

Description

Description text

IBM Cloud Object Storage ⓘ

This service stores the files associated with assets in the catalog.

Object storage instance

CloudObjectStorage

Or click [here](#) to create an additional instance. Click here to [refresh](#) the instances.

Can't be undone.

You can't disable policy enforcement for a catalog after you enabled it.

X

Permanently enforce data policies

☒ Enforce data policies

Duplicate asset handling ⓘ

Change this setting at any time on the Settings page.

☐ Update original assets ⓘ

☐ Overwrite original assets ⓘ

☒ Allow duplicates ⓘ

☐ Preserve original assets and reject duplicates ⓘ

Cancel

Create

For demo purposes, use “**Demo catalog – Data Fabric**” as a name. Description is optional.

Set **Enforce data policies** option. Click **Create**



## Creating Demo catalog - Data Fabric

This might take a minute...

Catalogs

Add to catalog +



## Demo catalog - Data Fabric

Assets

Access control

Settings



You haven't added any assets yet.

To add a new asset, click [Add to catalog](#)

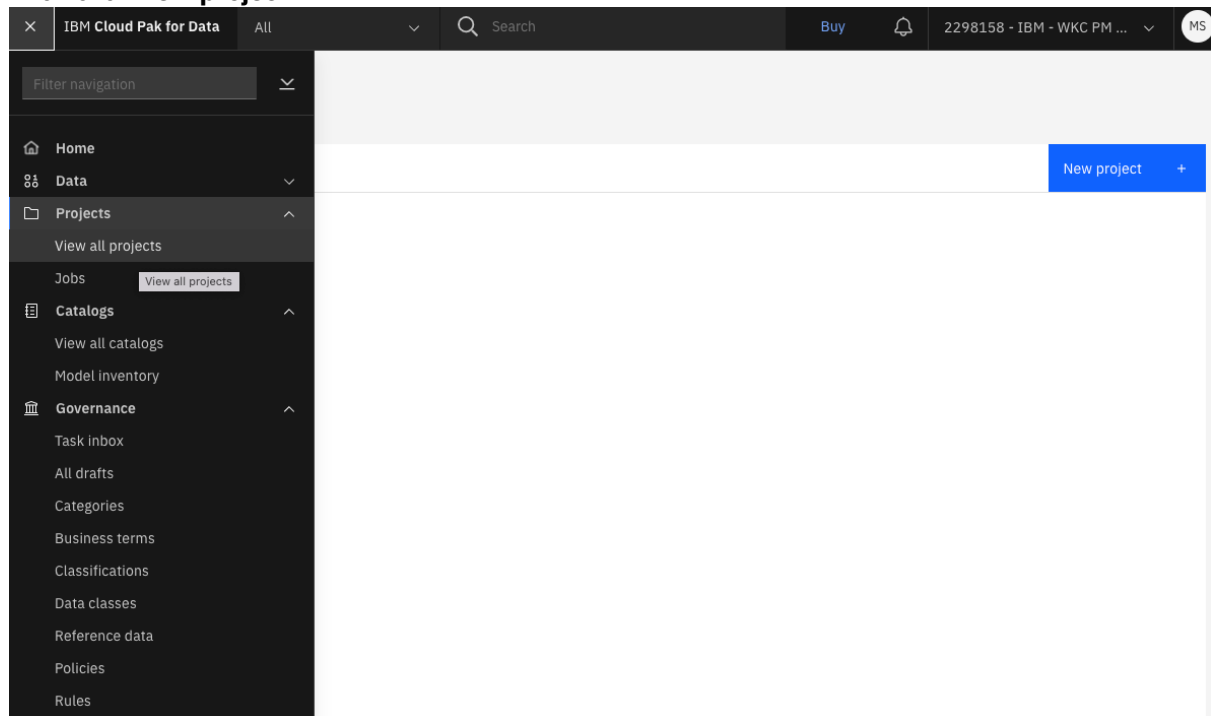
We will return to the catalog in one of the next steps.

## Create project

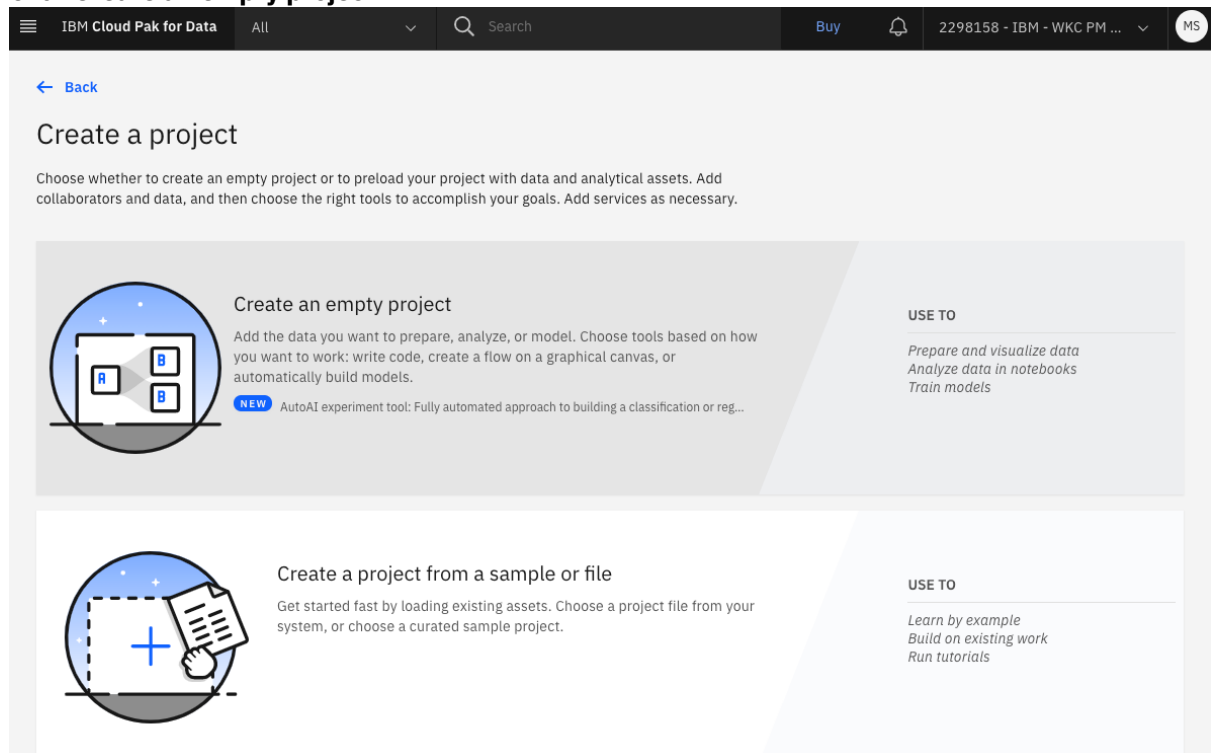
You can create a new project or use an existing project that you have access to.

To create a new project, open left hand side menu and click **View all projects** under **Projects** menu group.

Then click **New project**



## Click **Create an empty project**



For demo purposes, use “**Metadata Enrichment demo project**” as a name. Description is optional. Click **Create**.

IBM Cloud Pak for Data

All

Search

Buy

2298158 - IBM - WKC PM ...

MS

## New project

**Define details**

Name

Metadata Enrichment demo project

Description

Project description

**Storage**

cloud-object-storage-oj

**Choose project options**

☒ Restrict who can be a collaborator ⓘ

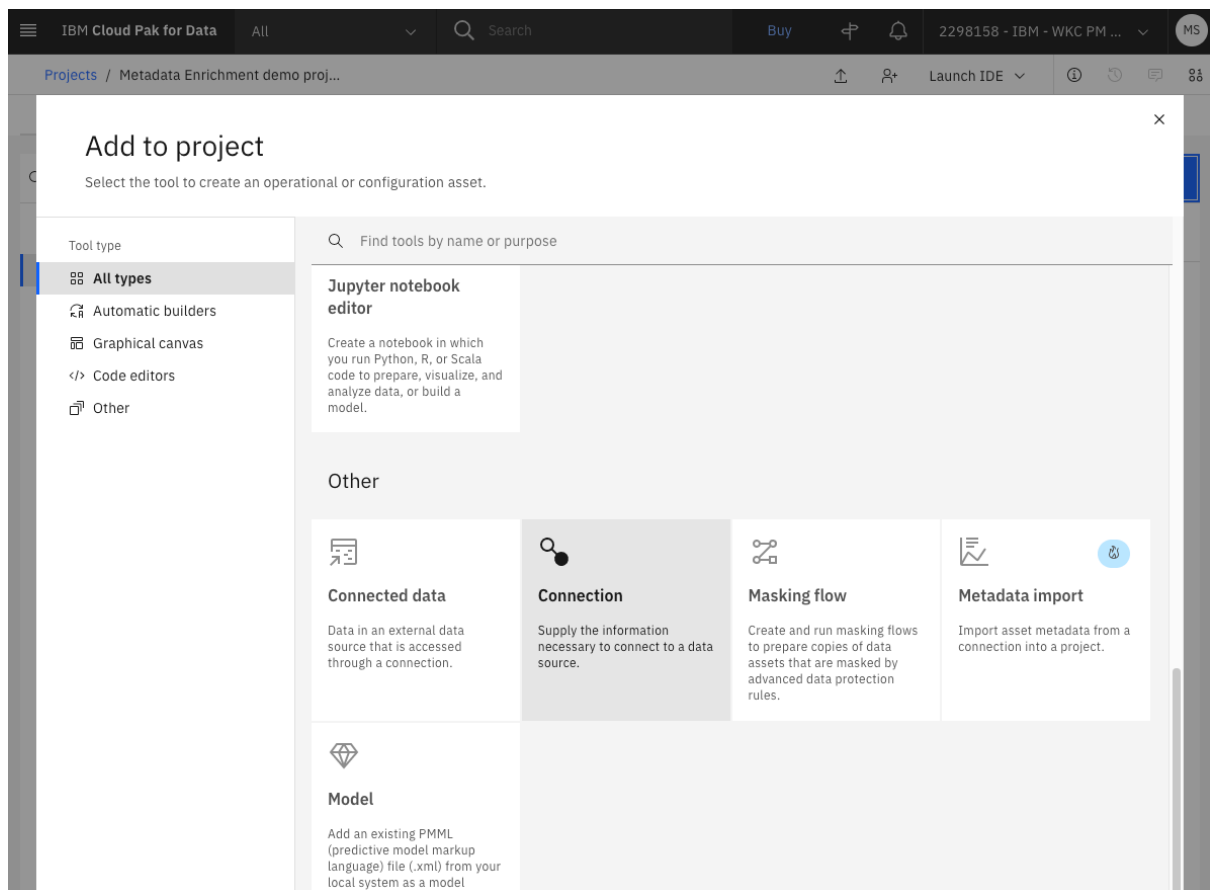
Project includes integration with [Cloud Object Storage](#) for storing project assets.

Cancel

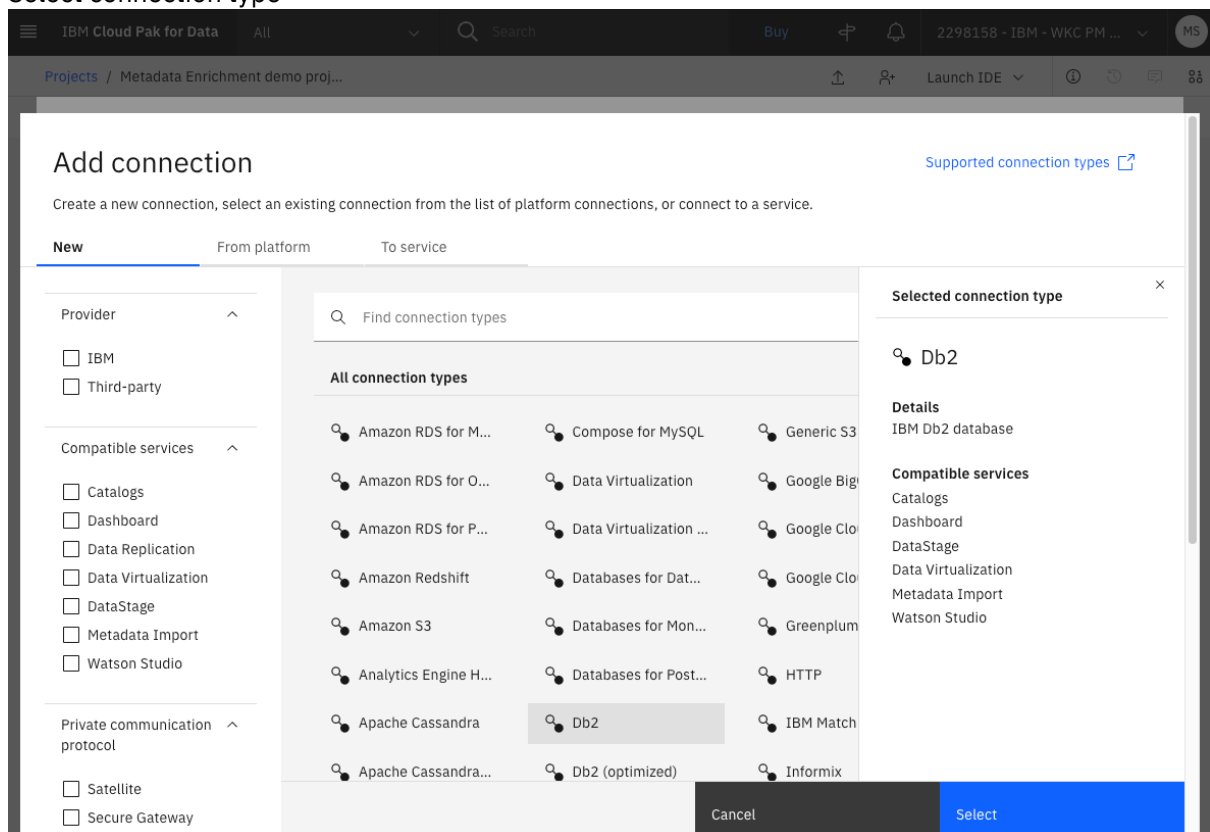
Create

## Create a connection

Go to **Assets** tab in Project view. Click **Add to project** button. Select **Connection**.



## Select connection type



IBM Cloud Pak for DataAllSearchBuy2298158 - IBM - WKC PM...MS

Projects / Metadata Enrichment demo proj...

Create connection: Db2Test connection

Enter the connection information.

Connection overview

Connection details

Credentials

Certificates

Private connectivity

Connection overviewShow more

NameData Fabric Trial - Db2 Warehouse

DescriptionConnection description

Connection detailsDatabase\*

CancelBackCreate

For demo purposes, use **“Data Fabric Trial - Db2 Warehouse”** as a connection name. Description is optional.

IBM Cloud Pak for DataAllSearchBuy2298158 - IBM - WKC PM...MS

Projects / Metadata Enrichment demo proj...

Create connection: Db2Test connection

Enter the connection information.

Connection overview

Connection details

Credentials

Certificates

Private connectivity

Connection details

Database\*BLUDB

Hostname or IP address\*db2w-ruggyab.us-south.db2w.cloud.ibm.com

Port\*50001

Credentials

Username\*CPDEMO

Password\*

Username and Password Security MechanismDefault

CancelBackCreate

For demo purposes, we are using following credentials:

- Name: Data Fabric Trial - Db2 Warehouse
- Database: BLUDB
- Hostname or IP address: db2w-ruggyab.us-south.db2w.cloud.ibm.com
- Port: 50001
- Username: CPDEMO
- Password: DataFabric@2022IBM
- Check the checkbox for “Port is SSL-Enabled”

Click the “Test Connection” button to make sure your entries are valid, and the connection is working. If the test connection is successful, click ‘Create’.

## Create Metadata Import asset - Import data to the project

In **Project -> Assets** click **Add to project**. Select ‘Metadata Import’. For demo purposes, use **“Metadata import demo – Data Fabric”** as the name. Optional: enter a description. Optional: use existing tags or create new ones. Select connection created in the previous step.

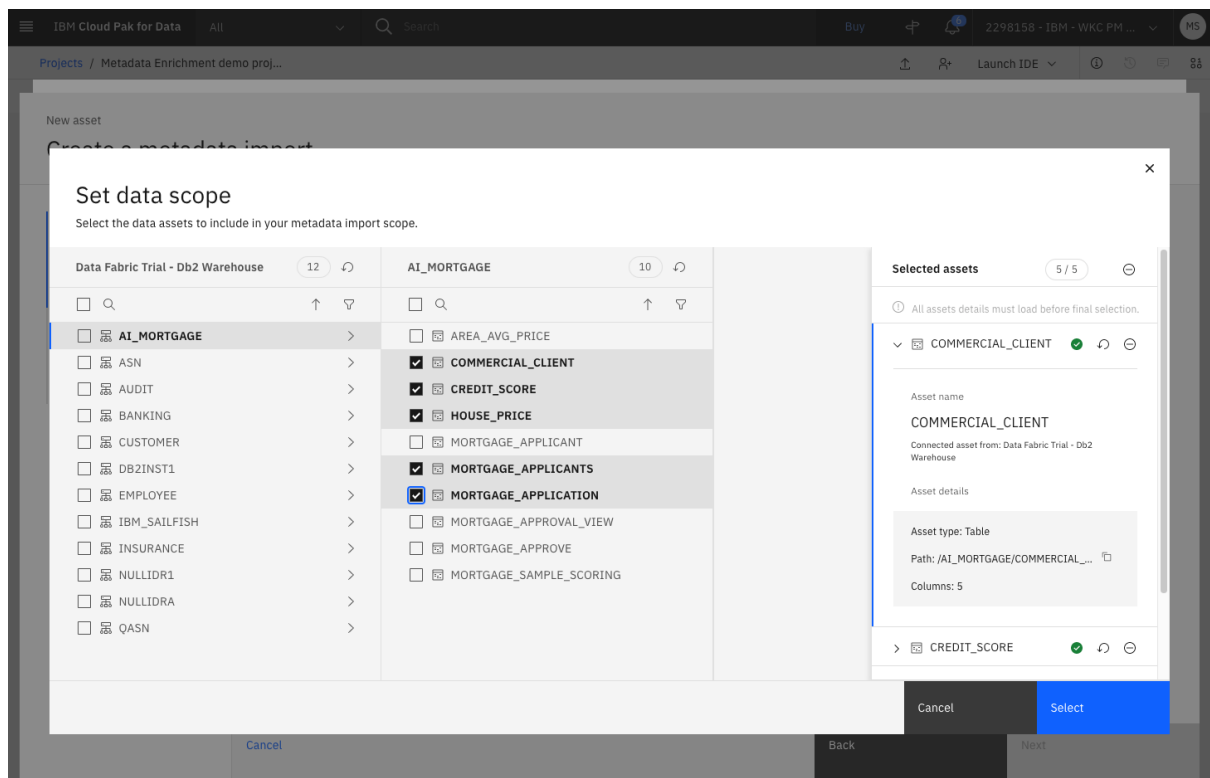
The screenshot shows the 'Create a metadata import' form in the IBM Cloud Pak for Data interface. The form is titled 'New asset' and 'Create a metadata import'. It has a sidebar with steps: Details (selected), Data scope, Schedule (Optional), and Review. The 'Details' section is titled 'Define details' and includes a 'Name' field with the value 'Metadata import demo - Data Fabric', a 'Description (optional)' text area, and a 'Tags (optional)' section with the text 'No tags added yet'. On the right, there is a dropdown menu for 'Data Fabric Trial - Db2 Warehouse' with a '+ Create a new connection' link and a 'Close menu' button. Below this, the 'Import target' section has two radio buttons: 'This project (Metadata Enrichment demo project)' (selected) and 'A catalog'. At the bottom, there are three buttons: 'Cancel', 'Back', and 'Next'.

Chose “This Project” for the import target. Click ‘Next’.

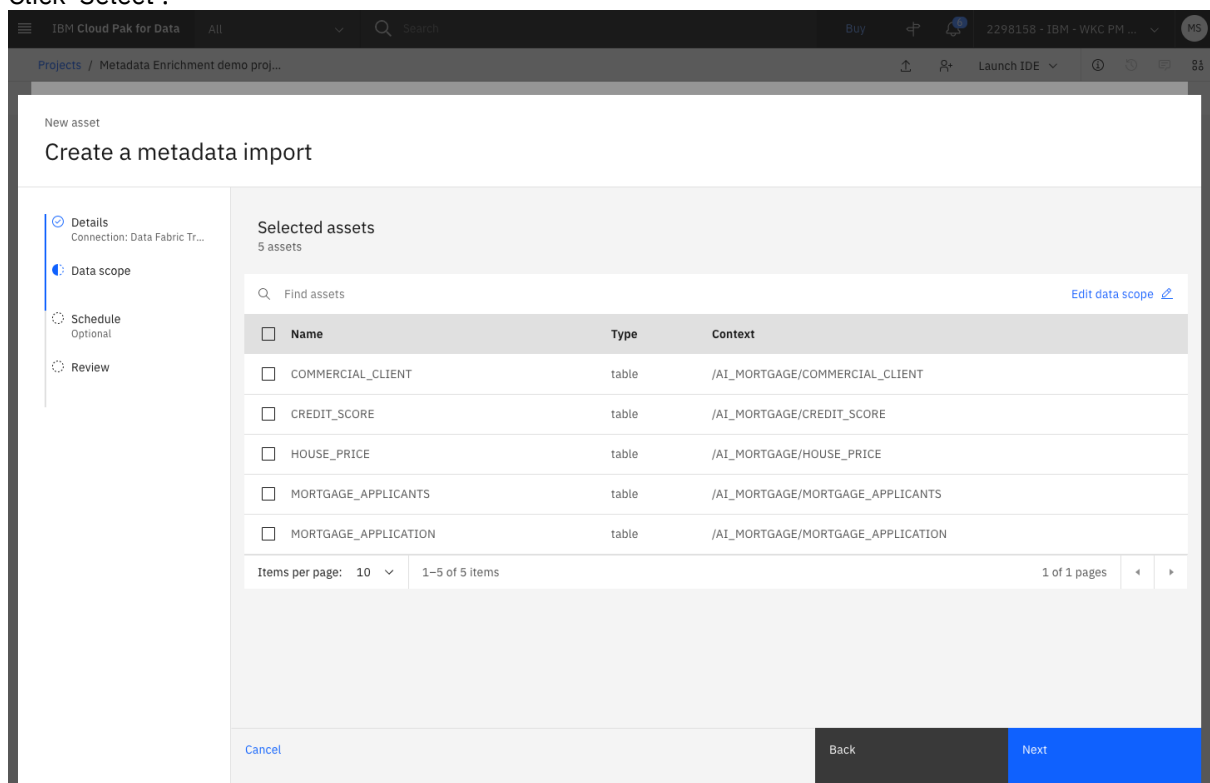
Select data scope. For the demo, from AI\_MORTGAGE schema, select:

- COMMERCIAL\_CLIENT
- CREDIT\_SCORE
- HOUSE\_PRICE
- MORTGAGE\_APPLICANTS
- MORTGAGE\_APPLICATION

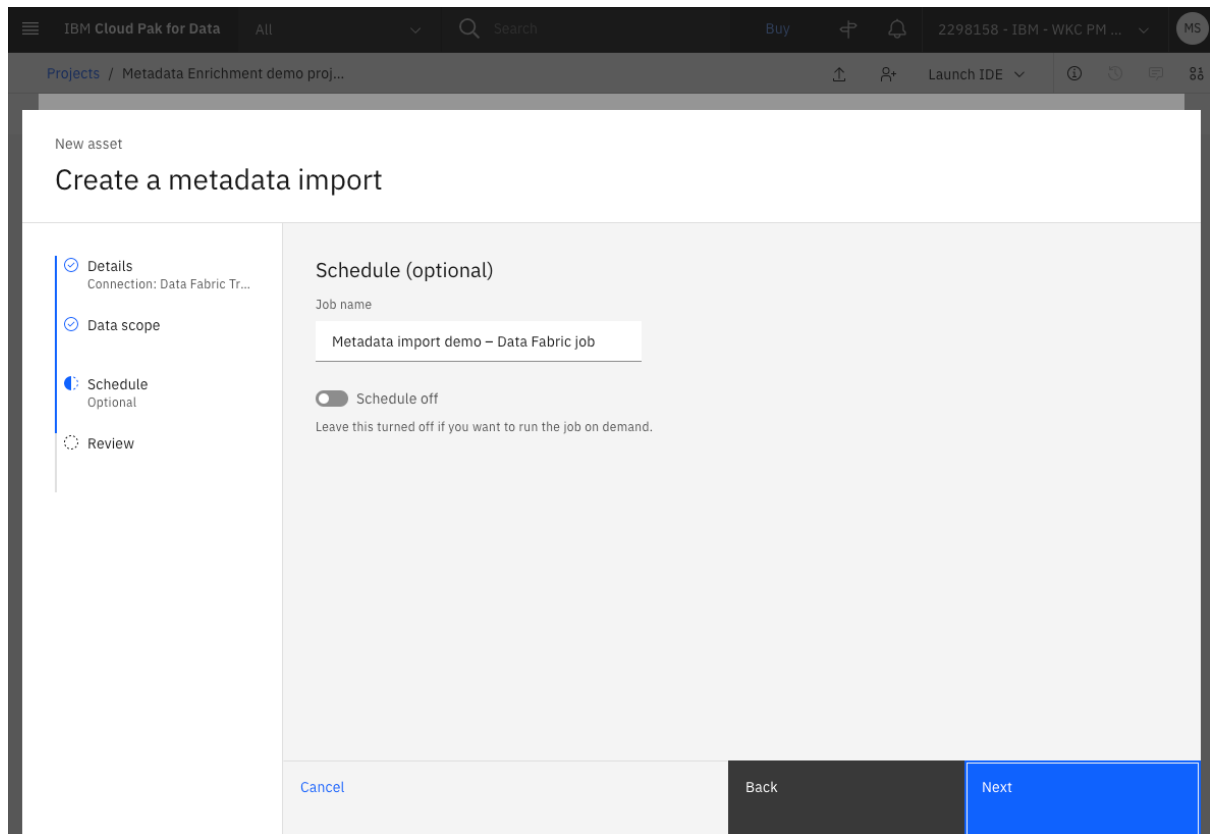




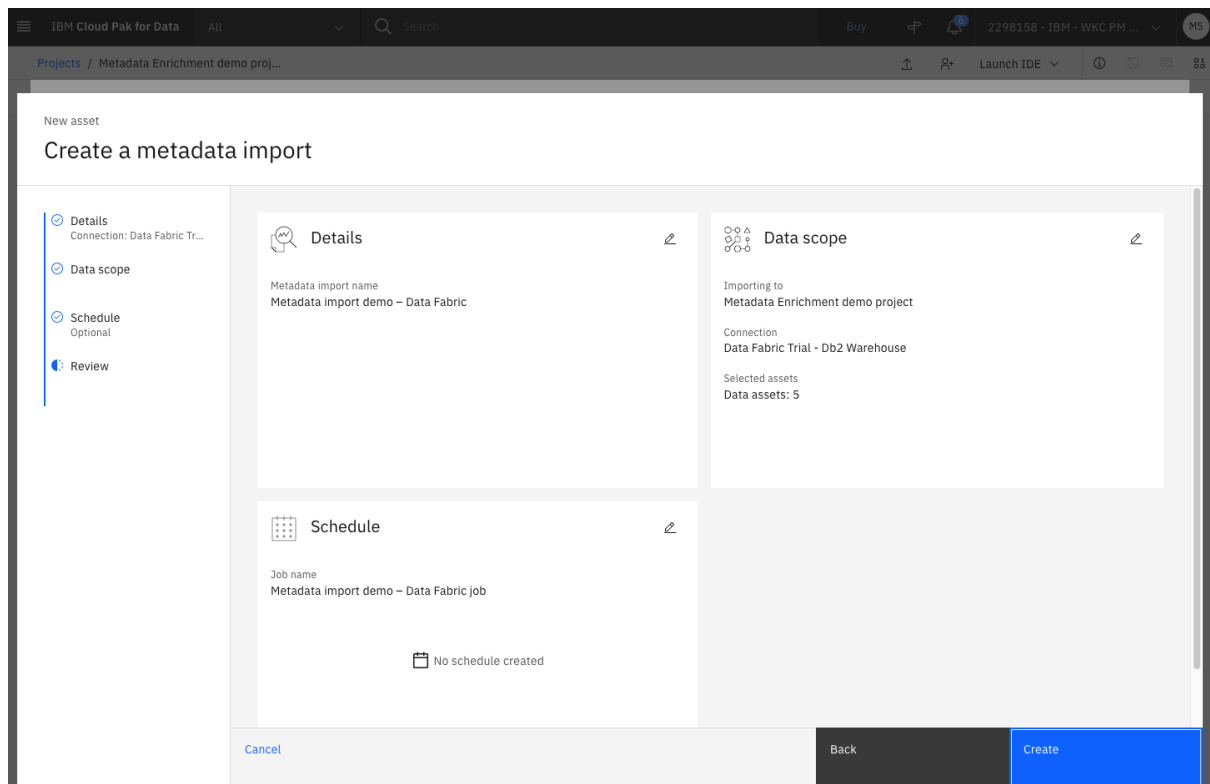
Click 'Select'.



Click "Next"



Click 'Next' on the Schedule optional step. Click 'Create'.



Wait until it finishes - Proper UI notification is displayed.

IBM Cloud Pak for Data

All

Search

Buy

2298158 - IBM - WKC PM ...

MS

Projects / Metadata Enrichment demo proj... / Metadata import demo – Data Fa...

Metadata import complete. 5 assets were imported successfully.

X

Metadata import demo – Data Fabric

Edit metadata import

Reimport assets

Metadata import

Imported assets

5 assets

Find assets

<input type="checkbox"/>	Name	Type	Context	Last imported	Status
<input type="checkbox"/>	COMMERCIAL_CLIENT	text/csv	AI_MORTGAGE/COMMERCIAL_CLIENT	Mar 07, 2022, 09:34 AM	Imported
<input type="checkbox"/>	MORTGAGE_APPLICATION	text/csv	AI_MORTGAGE/MORTGAGE_APPLICATION	Mar 07, 2022, 09:34 AM	Imported
<input type="checkbox"/>	CREDIT_SCORE	text/csv	AI_MORTGAGE/CREDIT_SCORE	Mar 07, 2022, 09:34 AM	Imported
<input type="checkbox"/>	HOUSE_PRICE	text/csv	AI_MORTGAGE/HOUSE_PRICE	Mar 07, 2022, 09:34 AM	Imported
<input type="checkbox"/>	MORTGAGE_APPLICANTS	text/csv	AI_MORTGAGE/MORTGAGE_APPLICANTS	Mar 07, 2022, 09:34 AM	Imported

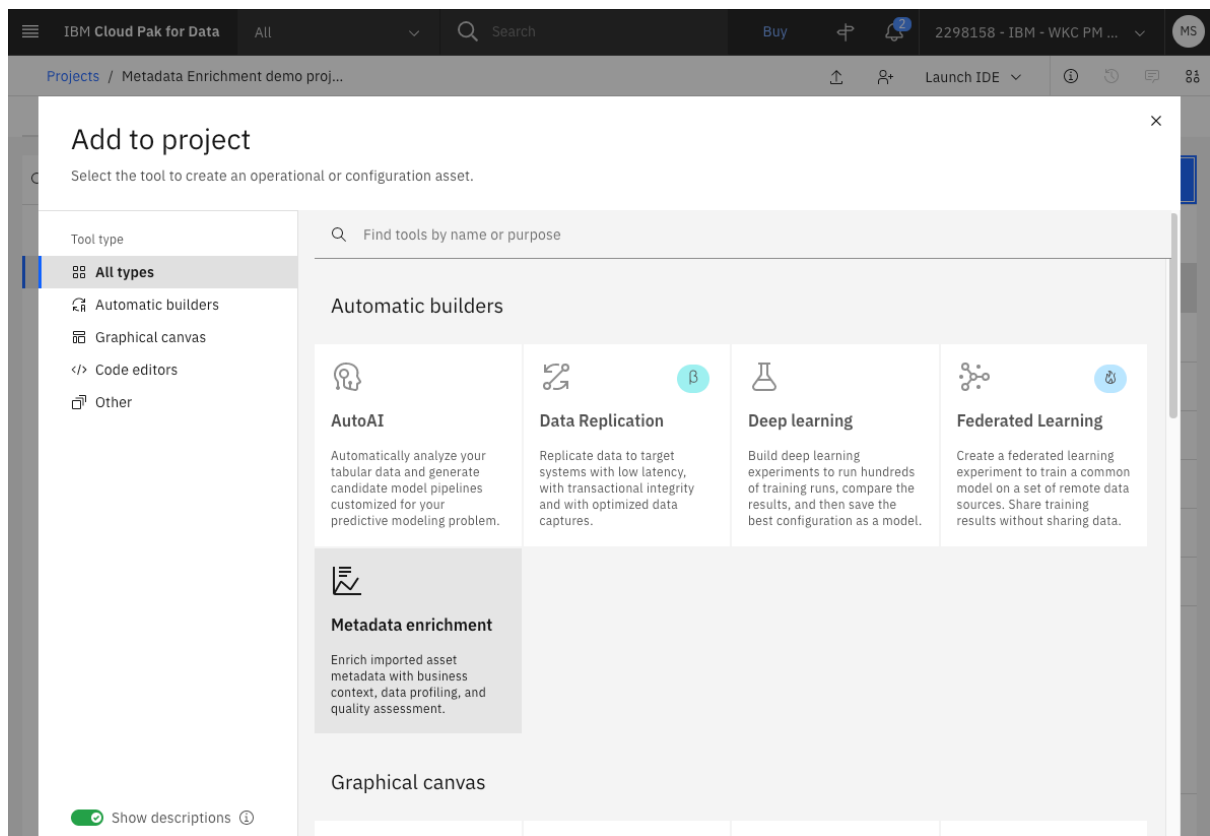
Items per page: 20 1–5 of 5 items 1 of 1 pages

## Metadata Enrichment

### Create Metadata Enrichment asset

*Metadata enrichment automatically providing context, content, and quality information in a unified framework.*

In Project -> Assets click 'Add to project'. Select 'Metadata Enrichment'.



For demo purposes use **Metadata enrichment demo - Data Fabric** as the name or whatever you would like to use.

Click 'Next'.

Select data from project. Click on data assets and select data assets you want run MDE. I select all six available tables.

IBM Cloud Pak for Data

All

Search

Buy

2298158 - IBM - WKC PM ...

MS

Projects / Metadata Enrichment demo proj... / Metadata import demo – Data Fa...

Metadata import complete. 5 assets were imported successfully.

Metadata import demo – Data Fabric

Metadata import

Edit metadata import

Reimport assets

Imported assets

5 assets

Find assets

<input type="checkbox"/>	Name	Type	Context	Last imported	Status
<input type="checkbox"/>	COMMERCIAL_CLIENT	text/csv	AI_MORTGAGE/COMMERCIAL_CLIENT	Mar 07, 2022, 09:34 AM	Imported
<input type="checkbox"/>	MORTGAGE_APPLICATION	text/csv	AI_MORTGAGE/MORTGAGE_APPLICATION	Mar 07, 2022, 09:34 AM	Imported
<input type="checkbox"/>	CREDIT_SCORE	text/csv	AI_MORTGAGE/CREDIT_SCORE	Mar 07, 2022, 09:34 AM	Imported
<input type="checkbox"/>	HOUSE_PRICE	text/csv	AI_MORTGAGE/HOUSE_PRICE	Mar 07, 2022, 09:34 AM	Imported
<input type="checkbox"/>	MORTGAGE_APPLICANTS	text/csv	AI_MORTGAGE/MORTGAGE_APPLICANTS	Mar 07, 2022, 09:34 AM	Imported

Items per page: 20 1–5 of 5 items 1 of 1 pages

Click ‘Select’.

IBM Cloud Pak for Data

All

Search

Buy

2298158 - IBM - WKC PM ...

MS

Projects / Metadata Enrichment demo proj...

New asset

Create a metadata enrichment

Details

Data scope

Objective

Schedule Optional

Review

Selected assets

5 assets

Find assets

Add assets to data scope

<input type="checkbox"/>	Asset name	Last modified	Created on
<input type="checkbox"/>	COMMERCIAL_CLIENT	4 min ago	4 min ago by you
<input type="checkbox"/>	CREDIT_SCORE	4 min ago	4 min ago by you
<input type="checkbox"/>	HOUSE_PRICE	4 min ago	4 min ago by you
<input type="checkbox"/>	MORTGAGE_APPLICANTS	4 min ago	4 min ago by you
<input type="checkbox"/>	MORTGAGE_APPLICATION	4 min ago	4 min ago by you

Items per page: 20 1–5 of 5 items 1 of 1 page

Cancel

Back

Next

Click ‘Next’. Select all available options for Enrichment objective.

## Enrichment objective

### Profile data

Provides statistics about the asset content and assigns and suggests data classes



### Analyze quality

Provides a data quality score for tables and columns



### Assign terms

Assigns and suggests business terms to tables and columns



Click 'Select categories' - !!! **Important!!!**. Selection in this step will impact the outcome of the metadata enrichment. For demo purposes, select uncategorized and Banking categories. The Uncategorized category contains data classes required by profiling. As well, in the Banking category there are some Terms created specifically for the BANKDEMO assets in this demo.

### Select categories

×

The selected categories determine the business terms and data classes to be used to enrich the data.

3 selected

Find categories

<

Categories

☒ [uncategorized]

☐ [uncategorized']

☐ Auto Insurance Claims

☒ Banking >

☐ Customer

☐ Customer Information

☐ Employee >

☐ Finance

Cancel

Select

Click 'Select'.

IBM Cloud Pak for Data

All

Search

Buy

2298158 - IBM - WKC PM ...

MS

Projects / Metadata Enrichment demo proj...

1

2

3

4

5

New asset

Create a metadata enrichment

Details

Data scope

Objective

Schedule Optional

Review

Enrichment objective

Profile data

Provides statistics about the asset content and assigns and suggests data classes

Analyze quality

Provides a data quality score for tables and columns

Assign terms

Assigns and suggests business terms to tables and columns

Categories

Selected categories

[uncategorized]

Banking

Sampling

Basic

Minimum sample size to optimize for speed

Analyze: 1,000 rows per table

Classify based on most

Moderate

Serves as a trade-off between speed and accuracy

Analyze: 10,000 rows per table

Classify based on most

Comprehensive

Large sample size to optimize for accuracy

Analyze: 100,000 rows per table

Classify all values per

Custom

Rows per table to analyze

1000

Classify column based on:

Cancel

Back

Next

You can leave sampling as it is (Basic). Click 'Next'.

IBM Cloud Pak for Data

All

Search

Buy

2298158 - IBM - WKC PM ...

MS

Projects / Metadata Enrichment demo proj...

1

2

3

4

5

New asset

Create a metadata enrichment

Details

Data scope

Objective

Schedule Optional

Review

Schedule enrichment job

Job name

Metadata enrichment demo - Data Fabric

Schedule off

Data scope of reruns

New or modified assets

All data assets

Cancel

Back

Next

Click 'Next'.

IBM Cloud Pak for DataAllSearchBuy2298158 - IBM - WKC PM ...MS

Projects / Metadata Enrichment demo proj...

New asset

Create a metadata enrichment

Details

Data scope

Objective

ScheduleOptional

Review

Details

Metadata enrichment name  
Metadata enrichment demo - Data Fabric

Description  
—

Tags  
—

Objective

Enrichment options  
Profile data | Analyze quality | Assign terms

Categories  
[uncategorized] | Banking

Sampling  
From top: Basic

Data scope

Structured data  
5 data assets

Schedule

Job name  
Metadata enrichment demo - Data Fabric

Starts  
—Last run  
—

Repeats  
—

CancelBackCreate

Click 'Create'. After the last step MDE process starts and relevant job is created. The process can take ~2-3 minutes.

IBM Cloud Pak for DataAllSearchBuy2298158 - IBM - WKC PM ...MS

Projects / Metadata Enrichment demo proj...

New asset

Create a metadata enrichment

Details

Data scope

Objective

ScheduleOptional

Review

Details

Metadata enrichment name  
Metadata enrichment demo - Data Fabric

Description  
—

Tags  
—

Objective

Enrichment options  
Profile data | Analyze quality | Assign terms

Categories  
[uncategorized] | Banking

Sampling  
From top: Basic

Data scope

Structured data  
5 data assets

Schedule

Job name  
Metadata enrichment demo - Data Fabric

Starts  
—Last run  
—

Repeats  
—

CancelBackCreate

Jobs / runs

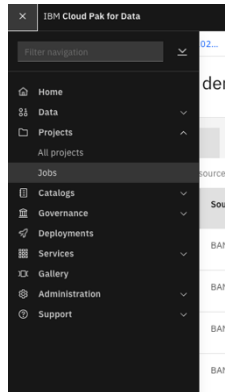


---

*You can use jobs for troubleshooting and to get fine graded insight into all the runs.*

---

You can see job status when you go to hamburger menu->Jobs.

A screenshot of the IBM Cloud Pak for Data 'Job Details' page. The page shows the job overview with 1 run completed and 0 runs failed. A table lists the job runs, showing a single completed run on March 07, 2022, at 9:51:35 AM, with a duration of 00:01:19. The page also includes a search bar, a filter icon, and a 'Last updated' timestamp of 07/03/2022, 09:55. The page is titled 'Job Details' and 'Overview'.

Job Details

Overview

1 Runs Completed 0 Runs Failed

No schedule created.

Runs (1)

Find a job run Last updated: 07/03/2022, 09:55

Start time	Status	Duration	Asset type
Mar 07, 2022 9:51:35 AM Started by you	Completed	00:01:19	Metadata Enrichment

Items per page: 10 1-1 of 1 item 1 of 1 page

To get details on specific run, click on the relevant run.

IBM Cloud Pak for Data

All

Search

Buy

2298158 - IBM - WKC PM ...

MS

Projects / Metadata Enrichment demo proj... / Metadata enrichment demo - Dat... / Job run details

Mar 07, 2022 9:51:35 AM

About this run

Completed

Run details

Duration (seconds): 79

Started by: Michal Szylar

Associated job: [Metadata enrichment demo - Data Fabric](#)

Log

Total 16 lines

Download log

Full metadata enrichment job run (75b71886-165b-46e2-92da-bab66de3ac18) is in state 'Completed'.

Job run asset statistics:  
Total assets: 5  
- Assets with status 'Created': 0  
- Assets with status 'In progress': 0  
- Assets with status 'Not found': 0  
- Assets with status 'Completed': 5  
- Completed assets with errors due to failed profiling operation: 0  
- Completed assets with errors due to failed term assignment operation: 0  
  
Profiling status:  
HB tasks with status 'COMPLETED': 1  
Active/Failed HB tasks: 0

Show more

## Result of MDE – profiling

After Metadata enrichment run is completed, you can navigate back to project “**Metadata Enrichment demo project**” -> **Assets** tab and from the left hand side menu select assets “**Metadata enrichment demo - Data Fabric**”.

IBM Cloud Pak for DataAllSearchBuy2298158 - IBM - WKC PM ...MS

Projects / Metadata Enrichment demo proj...

OverviewAssetsJobsManage

Find assetsAdd to project

8 assets

All assets

Asset types

> Data access2

> Data5

> Metadata enrichment1

Metadata Enrichment1

Metadata Enrichment

Name	Last modified
<div>Metadata enrichment demo - Data Fabric</div> <div>Metadata Enrichment</div>	3 days ago Service

Items per page: 201-1 of 1 items1 of 1 pages

Next click on the **Columns** tab, and on the **City** column. In the right hand side panel, you can find profiling information like: Format, Frequency distribution, Statistics.

IBM Cloud Pak for Data

All

Search

Buy

2298158 - IBM - WKC PM ...

MS

Projects / Metadata Enrichment demo proj...

Metadata enrichment demo - Data Fabric

Default settings

Edit enrichment

Assets (5)

Columns (44)

Find columns (search by column name, asset name or business term or data class)

Columns	Asset	Business terms	Data class	Data quality
<input type="checkbox"/> APPLIEDONLINE	MORTGAGE_APPLICATION Data Fabric Trial - Db2 Warehouse	—	Boolean	100%
<input type="checkbox"/> CITY	CREDIT_SCORE Data Fabric Trial - Db2 Warehouse	City	City	84%
<input type="checkbox"/> CITY	MORTGAGE_APPLICANTS Data Fabric Trial - Db2 Warehouse	City	City	83%
<input type="checkbox"/> COMM_FRAUD_INV	COMMERCIAL_CLIENT Data Fabric Trial - Db2 Warehouse	Commercial frauds...	Boolean	100%
<input type="checkbox"/> COMMERCIAL_CLIENT	COMMERCIAL_CLIENT Data Fabric Trial - Db2 Warehouse	Commercial Client	Boolean	100%
<input type="checkbox"/> CREDIT_SCORE	CREDIT_SCORE Data Fabric Trial - Db2 Warehouse	Credit Score	Code	91%
<input type="checkbox"/> CREDITCARD_DEBT	MORTGAGE_APPLICATION Data Fabric Trial - Db2 Warehouse	Credit Card Debt	NoClassDetected	92%
<input type="checkbox"/> CRM_ID	COMMERCIAL_CLIENT Data Fabric Trial - Db2 Warehouse	Customer Identifier + 1 more	Health Insuranc...	100%
<input type="checkbox"/> EDUCATION	MORTGAGE_APPLICANTS Data Fabric Trial - Db2 Warehouse	Education	Last Name	100%

Items per page: 50

1-44 of 44 items

1 of 1 page

CITY

Column from CREDIT\_SCORE

Go to asset in project

Details

Governance

Source

Data Fabric Trial - Db2 Warehouse / AI\_M

Formats

Inferred format

NA

Other formats

Aaaaaaa

Aaaaaa

Aaaaaaaa

Aaaaaaaa

Aaaaaaaa

Aaaaaaaa

View all (57)

Frequency distribution

New York 23

Portland 21

Denver 18

San Jose 11

Virginia Beach 10

View all (100)

Statistics

Unique 0

Minimum length 4

Maximum length 19

Mean length 8.936

*This analysis (also known as column analysis) literally captures the profile of the data. Structured profiling provides sample values, formats (aka patterns), detection of data properties (lengths, data types, min/max, etc.) and frequency distributions.*

*For columns with unclear names like “CSRidUpdate21”, sample values that look like email address help users understand what is contained in that column. For a column called “ZIP”, a frequency distribution with 80% 5-digit U.S. zip codes and 20% 6 digit-Canadian zip codes helps a user know that this data is from the worldwide database. Profiling gives consistent information to help users understand data fast.*

## Result of MDE – class and term assignment

In the columns view you can also see data class and business term auto assignment. You can navigate to the right-hand side panel -> Governance to adjust auto assigned terms and data classes:

IBM Cloud Pak for Data

All

Search

Buy

2298158 - IBM - WKC PM ...

MS

Projects / Metadata Enrichment demo proj...

Metadata enrichment demo - Data Fabric

Default settings

Edit enrichment

Assets (5)

Columns (44)

Find columns (search by column name, asset name or business term or data class)

Columns	Asset	Business terms	Data class	Data quality
<input type="checkbox"/> APPLIEDONLINE	MORTGAGE_APPLICATION Data Fabric Trial - Db2 Warehouse	—	Boolean	100%
<input type="checkbox"/> CITY	CREDIT_SCORE Data Fabric Trial - Db2 Warehouse	City	City	84%
<input type="checkbox"/> CITY	MORTGAGE_APPLICANTS Data Fabric Trial - Db2 Warehouse	City	City	83%
<input type="checkbox"/> COMM_FRAUD_INV	COMMERCIAL_CLIENT Data Fabric Trial - Db2 Warehouse	Commercial frauds...	Boolean	100%
<input type="checkbox"/> COMMERCIAL_CLIENT	COMMERCIAL_CLIENT Data Fabric Trial - Db2 Warehouse	Commercial Client	Boolean	100%
<input type="checkbox"/> CREDIT_SCORE	CREDIT_SCORE Data Fabric Trial - Db2 Warehouse	Credit Score	Code	91%
<input type="checkbox"/> CREDITCARD_DEBT	MORTGAGE_APPLICATION Data Fabric Trial - Db2 Warehouse	Credit Card Debt	NoClassDetected	92%
<input type="checkbox"/> CRM_ID	COMMERCIAL_CLIENT Data Fabric Trial - Db2 Warehouse	Customer Identifier   + 1 more	Health Insuranc...	100%
<input type="checkbox"/> EDUCATION	MORTGAGE_APPLICANTS Data Fabric Trial - Db2 Warehouse	Education	Last Name	100%

Items per page: 50

1-44 of 44 items

1 of 1 page

CITY

Column from CREDIT\_SCORE

Go to asset in project

Details

Governance

Business terms 1

City Banking  
Confidence score: 100%

Data class

City

Locating, assessing, and managing your most critical data elements is a prerequisite for unlocking business value, complying with regulatory standards, and reducing risk. Watson Knowledge Catalog scales the productivity of data stewards by automating the data curation process with patented technology that mirrors how human beings recognize and curate data.

This automatic business term assignment help connect business meaning to the data — establishing signal from noise. This process of assigning meaning is often referred to using phrases like mapping or auto-tagging. The value of assigning business terms is that data with the right business term is easier to find using search, easier to protect using data protection rules, and easier to analyze using automation rules to trigger specific data quality rules.

Like automatic business term assignment, automatic data class assignment (aka data classification) allows for automatic detection of business meaning and value. Automatic data class assignment analyzes actual data values to assign the best data class, and this result is one of a few factors used in the automatic business term assignment framework.

Automatic data class assignment is run in conjunction with profiling, together known as column analysis. IBM provides over 200 out of the box data classes. These range from a core set (identifier, code, text, quantity, etc.) to specific domains (credit card, email address, etc.). In some instances, these data classes provide validation of the data, like verifying valid credit card number's structure.

## Result of MDE – data quality

To see data quality details, you can click on 3 dots icon -> View data quality:

The screenshot displays the 'Metadata enrichment demo - Data Fabric' interface. It features a table with columns: Columns, Asset, Business terms, Data class, Data quality, and Review status. The table lists 10 columns with their respective data quality scores. A context menu is open for the 'CITY' column, showing options like 'View column details', 'View data quality', and 'Mark as reviewed'. On the right, a sidebar shows details for the 'APPLIEDONLINE' column, including source, formats, frequency distribution, and statistics.

Columns	Asset	Business terms	Data class	Data quality	Review status
APPLIEDONLINE	MORTGAGE_APPLICATION Data Fabric Trial - Db2 Warehouse	—	Boolean	100%	
CITY	CREDIT_SCORE Data Fabric Trial - Db2 Warehouse	City	City	84%	
CITY	MORTGAGE_APPLICANTS Data Fabric Trial - Db2 Warehouse	City	City	83%	
COMM_FRAUD_INV	COMMERCIAL_CLIENT Data Fabric Trial - Db2 Warehouse	Commercial frauds...	Boolean	100%	
COMMERCIAL_CLIENT	COMMERCIAL_CLIENT Data Fabric Trial - Db2 Warehouse	Commercial Client	Boolean	100%	
CREDIT_SCORE	CREDIT_SCORE Data Fabric Trial - Db2 Warehouse	Credit Score	Code	91%	
CREDITCARD_DEBT	MORTGAGE_APPLICATION Data Fabric Trial - Db2 Warehouse	Credit Card Debt	NoClassDetected	92%	
CRM_ID	COMMERCIAL_CLIENT Data Fabric Trial - Db2 Warehouse	Customer Identifier + 1 more	Health Insuranc...	100%	
EDUCATION	MORTGAGE_APPLICANTS Data Fabric Trial - Db2 Warehouse	Education	Last Name	100%	

*Managing and trusting data at enterprise scale depends on extensible framework for analyzing data quality. Using pre-built or customized measures of quality, a user can easily understand the nature of their data from a single number and then take action on their most critical assets.*

*WKC generates a data quality score for each column and data asset right out of the box. The data quality score is calculated for every data asset and column by analyzing every value in every record according to pre-built dimensions. Think about this scoring like a nutrition label for a data asset — it's a standardized set of metrics that captures the overall quality of a data asset. Using this data quality score, customers can quickly locate and value and risk.*

*DQ scores help users locate, diagnose, then prioritize data quality issues. Scores also help identify and monitor business value. For example, if my data quality score is high and few DQ dimension violations are found, I can be confident that a data asset is of quality.*

## Data quality dimensions for CITY

Findings: 200 | Quality score: ● 84%

Dimension	Findings	Percentage of records
Data class violations	198	20%
Inconsistent capitalization	2	0%
Suspect values	0	0%
Inconsistent representation o...	0	0%
Unexpected missing values	0	0%
Format violations	0	0%
Unexpected duplicated values	0	0%
Values out of range	0	0%
Data type violations	0	0%

Under 3dots menu, you can also mark asset or column as reviewed.

IBM Cloud Pak for Data

All

Search

Buy

2298158 - IBM - WKC PM ...

MS

Projects / Metadata Enrichment demo proj...

Metadata enrichment demo - Data Fabric

Default settings

Edit enrichment

Assets (5)

Columns (44)

Find columns (search by column name, asset name or business term or data class)

Columns	Asset	Business terms	Data class	Data quality	Review status
<input type="checkbox"/> APPLIEDONLINE	MORTGAGE_APPLICATION Data Fabric Trial - Db2 Warehouse	—	Boolean	● 100%	⊙
<input type="checkbox"/> CITY	CREDIT_SCORE Data Fabric Trial - Db2 Warehouse	City	City	● 84%	⊙
<input type="checkbox"/> CITY	MORTGAGE_APPLICANTS Data Fabric Trial - Db2 Warehouse	City	City	● 83%	⊙
<input type="checkbox"/> COMM_FRAUD_INV	COMMERCIAL_CLIENT Data Fabric Trial - Db2 Warehouse	Commercial frauds...	Boolean	● 100%	⊙
<input type="checkbox"/> COMMERCIAL_CLIENT	COMMERCIAL_CLIENT Data Fabric Trial - Db2 Warehouse	Commercial Client	Boolean	● 100%	⊙
<input type="checkbox"/> CREDIT_SCORE	CREDIT_SCORE Data Fabric Trial - Db2 Warehouse	Credit Score	Code	● 91%	⊙
<input type="checkbox"/> CREDITCARD_DEBT	MORTGAGE_APPLICATION Data Fabric Trial - Db2 Warehouse	Credit Card Debt	NoClassDetected	● 92%	⊙
<input type="checkbox"/> CRM_ID	COMMERCIAL_CLIENT Data Fabric Trial - Db2 Warehouse	Customer Identifier <a href="#">+ 1 more</a>	Health Insuranc...	● 100%	⊙
<input type="checkbox"/> EDUCATION	MORTGAGE_APPLICANTS Data Fabric Trial - Db2 Warehouse	Education	Last Name	● 100%	⊙

Items per page: 50 1-44 of 44 items 1 of 1 page

CITY

Column from CREDIT\_SCORE

Go to asset in project

Details Governance

Source

Data Fabric Trial - Db2 Warehouse / AI\_...

Formats

Inferred format

NA

Other formats

Aaaaaaa

Aaaaaa

Aaaaaaaa

Aaaaaaaa

Aaaaaaaa

View all (57)

Frequency distribution

New York 23

Portland 21

Denver 18

San Jose 11

Virginia Beach 10

View all (100)

Statistics

Unique 0

Minimum length 4

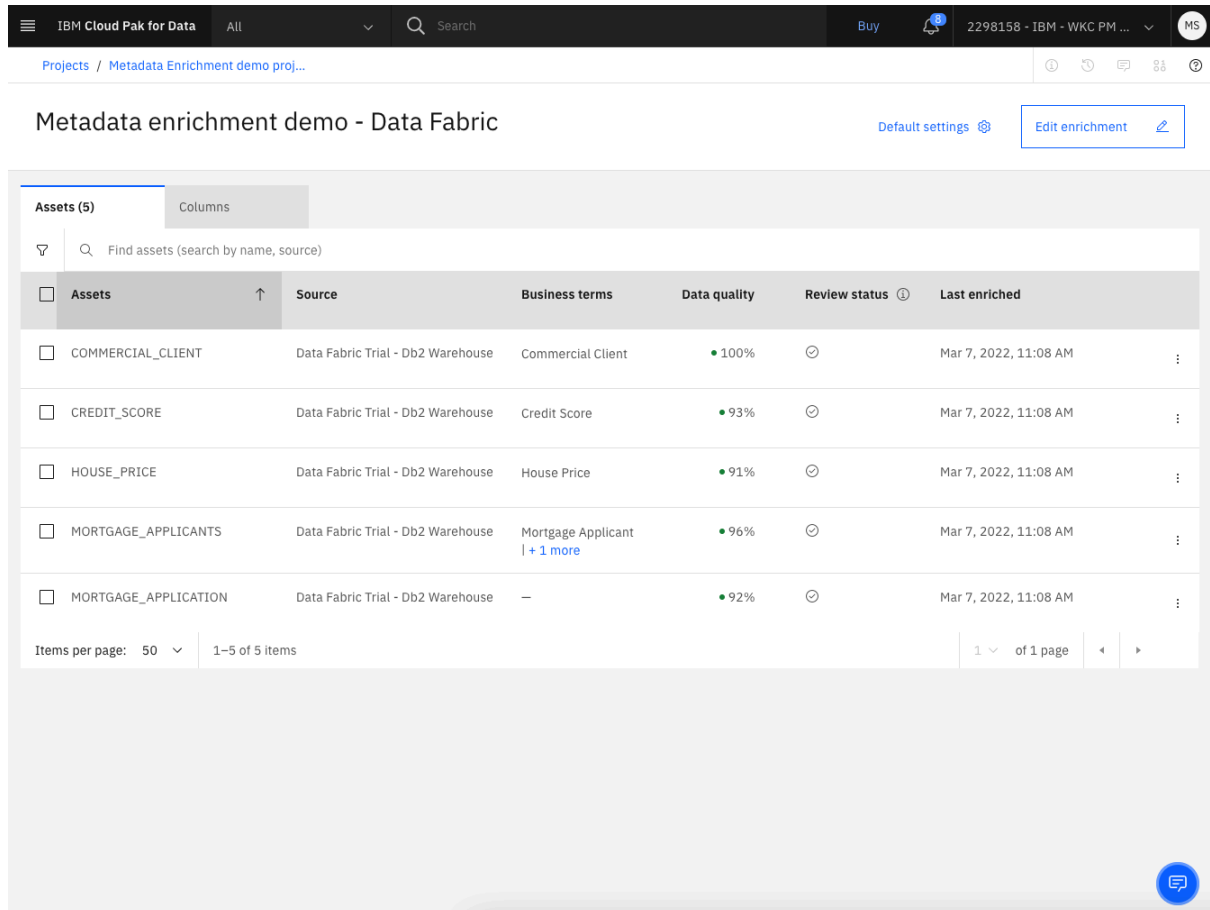
Maximum length 19

Mean length 8.936

## Publishing assets to catalog

After all assets are reviewed, you can publish one or all of them to a catalog.

Select relevant assets and click -> Publish to catalog



IBM Cloud Pak for Data All Search Buy 2298158 - IBM - WKC PM ... MS

Projects / Metadata Enrichment demo proj...

### Metadata enrichment demo - Data Fabric

Default settings Edit enrichment

Assets (5)	Columns				
Assets	Source	Business terms	Data quality	Review status	Last enriched
<input type="checkbox"/> COMMERCIAL_CLIENT	Data Fabric Trial - Db2 Warehouse	Commercial Client	100%	⊙	Mar 7, 2022, 11:08 AM
<input type="checkbox"/> CREDIT_SCORE	Data Fabric Trial - Db2 Warehouse	Credit Score	93%	⊙	Mar 7, 2022, 11:08 AM
<input type="checkbox"/> HOUSE_PRICE	Data Fabric Trial - Db2 Warehouse	House Price	91%	⊙	Mar 7, 2022, 11:08 AM
<input type="checkbox"/> MORTGAGE_APPLICANTS	Data Fabric Trial - Db2 Warehouse	Mortgage Applicant <a href="#">+ 1 more</a>	96%	⊙	Mar 7, 2022, 11:08 AM
<input type="checkbox"/> MORTGAGE_APPLICATION	Data Fabric Trial - Db2 Warehouse	—	92%	⊙	Mar 7, 2022, 11:08 AM

Items per page: 50 1-5 of 5 items 1 of 1 page

Select target catalog created in one of the previous steps.



IBM Cloud Pak for Data

All

Search

Buy

2298158 - IBM - WKC PM ...

MS

## Publish to catalog

Publish the selected assets to the catalog that you select.

Target\*

Demo catalog - Data Fabric

Selected assets (5)

Asset name	Format	Actions
COMMERCIAL_CLIENT	text/csv	
CREDIT_SCORE	text/csv	
HOUSE_PRICE	text/csv	

Description

Description of assets

Duplicate action

☒ Update ☐ Overwrite ☐ Add duplicate ☐ Preserve

Privacy

☒ Public ☐ Private

Connections to add (1)

The following connections will be added to the catalog to successfully publish the assets.

Data Fabric Trial - Db2 Warehouse

No description available for this asset

Tags

Start typing to add tags

Cancel

Publish

Click **Publish**. You can navigate to “**Demo catalog – Data Fabric**” to check results.

IBM Cloud Pak for Data

All

Search

Buy

2298158 - IBM - WKC PM ...

MS

Catalogs

Add to catalog

### Demo catalog - Data Fabric

Assets

Access control

Settings

Find assets

Filter by:

Any asset type

Any tag

Clear all

Hide featured assets

Recently added

Watson recomme...

Highly rated

Data

MORTGAGE\_APPLICATION

Owner: Michal Szylar

Added: Mar 07, 2022 11:27 AM

Tags: AI\_...

☆☆☆☆☆ 0 reviews

Data

CREDIT\_SCORE

Owner: Michal Szylar

Added: Mar 07, 2022 11:27 AM

Tags: AI\_...

☆☆☆☆☆ 0 reviews

Data

HOUSE\_PRICE

Owner: Michal Szylar

Added: Mar 07, 2022 11:27 AM

Tags: AI\_...

☆☆☆☆☆ 0 reviews

Data

COMMERCIAL\_CLIENT

Owner: Michal Szylar

Added: Mar 07, 2022 11:27 AM

Tags: AI\_...

☆☆☆☆☆ 0 reviews

Showing 6 of 6 items

<input type="checkbox"/>	Name	Owner	Tags	Business terms	Asset type	Date added
<input type="checkbox"/>	COMMERCIAL_CLIENT	MS Michal Szylar	AI_MORTGAGE	Commercial Client	Data	Mar 07, 2022
<input type="checkbox"/>	CREDIT_SCORE	MS Michal Szylar	AI_MORTGAGE	Credit Score	Data	Mar 07, 2022
<input type="checkbox"/>	Data Fabric Trial - Db2 Warehouse	MS Michal Szylar			Connection	Mar 07, 2022
<input type="checkbox"/>	HOUSE_PRICE	MS Michal Szylar	AI_MORTGAGE	House Price	Data	Mar 07, 2022
<input type="checkbox"/>	MORTGAGE_APPLICANTS	MS Michal Szylar	AI_MORTGAGE	Mortgage Applicant   + 1 more	Data	Mar 07, 2022
<input type="checkbox"/>	MORTGAGE_APPLICATION	MS Michal Szylar	AI_MORTGAGE		Data	Mar 07, 2022

## Conclusion

Data is like water—best served fresh, well-filtered, and with confidence in its origins and quality. A business’ data-driven decisions are only as good as the data that underlies them. The goal of metadata enrichment is to create value and meaning from data.

Metadata Enrichment enables customers to go from backroom to showroom, from data source to data catalog, faster. It enriches metadata with understanding. How? Automatically providing context, content, and quality information in a unified framework.

Automatic data class and business term assignment help automate the data curation process and enable faster, more scalable data governance. Profiling gives users confidence in the meaning and structure of their data. Data quality scoring and dimensions allow for easy assessment of business value and detection of data quality issues.

With infinite time, people, and a lot of handwritten code, customers could replicate almost all of this functionality. This functionality is valuable because it’s fast, scalable, and pre-built to deliver ROI from enterprise data.

## Resources and documentation

- Metadata Enrichment documentation can be found here:  
<https://dataplatfom.cloud.ibm.com/docs/content/wsj/governance/metadata-enrichment.html#supp-assets>
- You can learn more about Terms and Data Classes in Governance artifacts documentation:  
<https://dataplatfom.cloud.ibm.com/docs/content/wsj/governance/governance.html?audience=wdp>
- Metadata Import documentation:  
<https://dataplatfom.cloud.ibm.com/docs/content/wsj/manage-data/metadata-import.html?audience=wdp>