



Cloud Pak for Data

Tutorial – Mortgage

Contents

1.	Prerequisites	5
2.	Setting up database and sample data.....	5
3.	Access Credentials	6
3.1.	Access credential for Db2 database.....	6
3.2.	Sign in to Cloud Pak for Data web console as Administrator.....	6
4.	Create Connection	7
4.2.	Navigate to Connections	7
4.3.	Add connection	7
5.	Discover Assets	8
5.1.	Navigate to discover assets	8
6.	Add users	10
7.	Implement Policies and Rules.....	13
7.1.	Import Governance Assets.....	12
7.2.	Create a policy	13
7.3.	Create a rule	13
7.4.	Add rule to metadata.....	14
8.	Access data as a Data Scientist	15
8.1.	Create analytic project.....	16
8.2.	Assets from Glossary	16
10.	Navigate to data catalog.....	22
11.	Data Virtualization	24
11.1.	Adding a new data source for Db2	24
11.5.	Add virtual table to catalog	26
11.6.	Publish virtualized table.....	26
11.7.	Access information for virtual table	27
11.8.	Deliver Dataset	28
12.	Build Model	30
12.1.	Navigate to analytics project	30
12.2.	Create deployment space	30
12.3.	Create notebook	31
12.4.	Review and run notebook.....	32
12.5.	Test the model.....	33

Cloud Pak for Data – Tutorial

Cloud Pak for Data is a single end to end platform for data management, governance and data science analytics. It provides a one stop shop for data scientists, data engineer and data stewards to collaborate on the platform to acquire, govern and extract best insights from the data in the least amount of time.

In this demo, user will use a set of a fictitious mortgage data that available in Db2 database on IBM Bluemix Cloud. User will perform following tasks to predict if a prospective customer may default on their mortgage.

- Create connection from Cloud Pak for Data to Db2 database on cloud
- Discover Db2 assets from Cloud Pak for Data
- Transform the Db2 data on Cloud Pak for Data
- Use analytics dashboard to build visualizations
- Build a simple machine learning model from prediction

1. Prerequisites

- Access to an operational Cloud Pak for Data Instance
- Install Git on the machine that you will use for the tutorial.

2. Setting up database and sample data

Log in to the cluster where Cloud Pak for Data is deployed or log in to a Linux-based system (RedHat or Ubuntu) that can access the cluster over your network.

From your home directory, clone the tutorial sample files:

```
git clone git@github.com:IBM-ICP4D/ICP4DTutorial.git
```

Change to the tutorials directory:

```
cd ICP4XTutorial/tutorials/
```

The sample data-loading utility, `load_samples.sh`, provides an easy way to host a Db2 server and load it with sample data.

Run the following command to view the list of sample data that is provided in the `load_samples.sh` utility:

```
./load_samples.sh -l
```

Run the following command to load the sample data into a Db2 database:

```
./load_samples.sh -t mortgage-002
```

After the loading process completes, an instance of Db2 is hosted on your cluster as a Docker container.

3. Access Credentials

To work through the tutorial, you need access a Db2 database.

3.1. Access credential for Db2 database

For this tutorial you need JDBC connection to access to a Db2 database that hosted locally on Cloud Pak for Data. Following are JDBC connection credential for Db2:

JDBC Host name	<Same IP address as your web console>
Port number	50000
Database name	MORTGAGE
User ID	db2inst1
Password	password
Db2	Version 11.1
JDBC connection string	jdbc:db2://<same IP as Web Console>:50000/MORTGAGE

3.2. Sign in to Cloud Pak for Data web console as Administrator

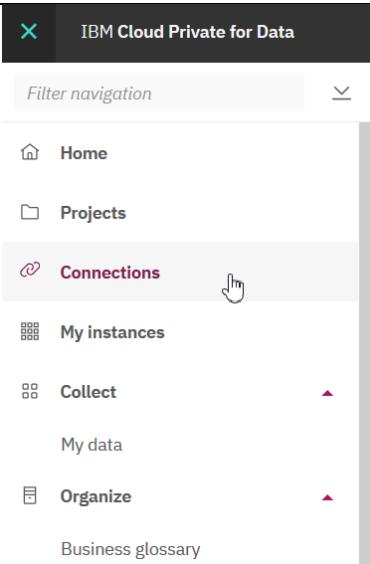
You should have an operational Cloud Pak for Data Instance. Use latest version of Firefox or Google Chrome browser to access the Cloud Pak for Data web console. Starting from here all instruction need to execute on Cloud Pak for Data web console only. You need to login as admin who has administrator privileges.

Sign in  Username <input type="text" value="admin"/> PASSWORD <input type="password" value="*****"/> <input type="button" value="Sign In"/>	Sigh in to the Cloud Pak for Data web console as user 'admin' and password is 'password'.
--	---

4. Create Connection

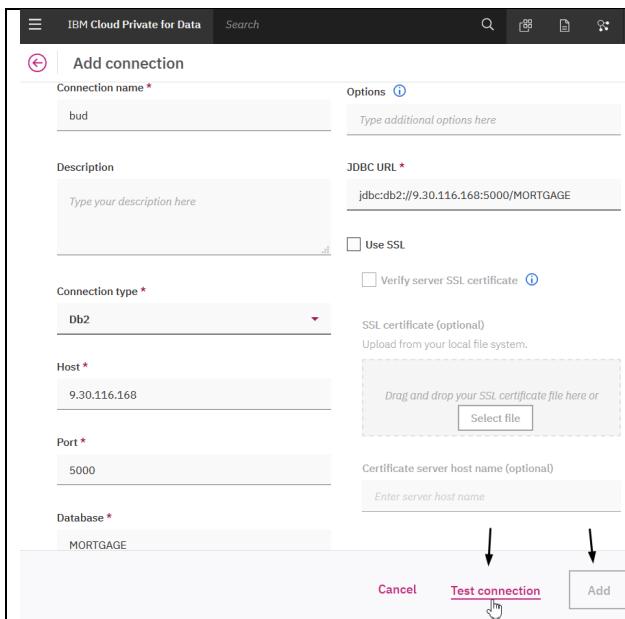
Create a connection to the data source for Db2 database.

4.2. Navigate to Connections



On the left pane choose **Connections**. Next, on the **Data Connections** window click on the **+ Add connection** icon.

4.3. Add connection



Fill out the **Add Connection** information according to the information provided in step ‘2.1. Access credential for DB2. Credential used in following step is just an example.

- For **Choose connection** use the drop-down menu and select ‘Db2’.
- Use ‘Bud’ as the **Name**
- JDBC URL** is ‘`jdbc:db2://172.16.171.29:50000/MORTGAGE`’
- Username** is ‘`db2inst1`’ and **Password** is ‘`password`’.

Next click on **Test Connection**, once it successful click on **Save Connection**.

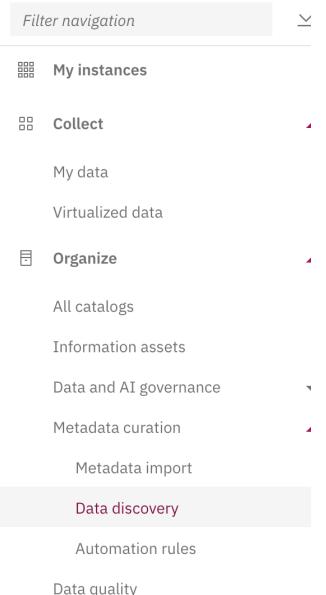


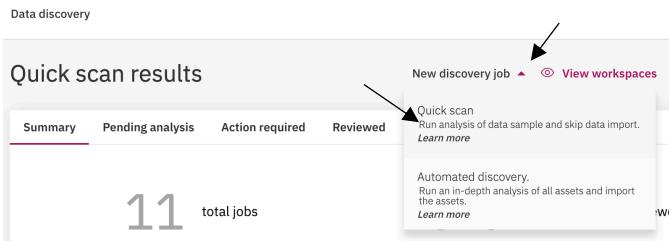
Success The test connection was successful. Click Add to save the connection information.

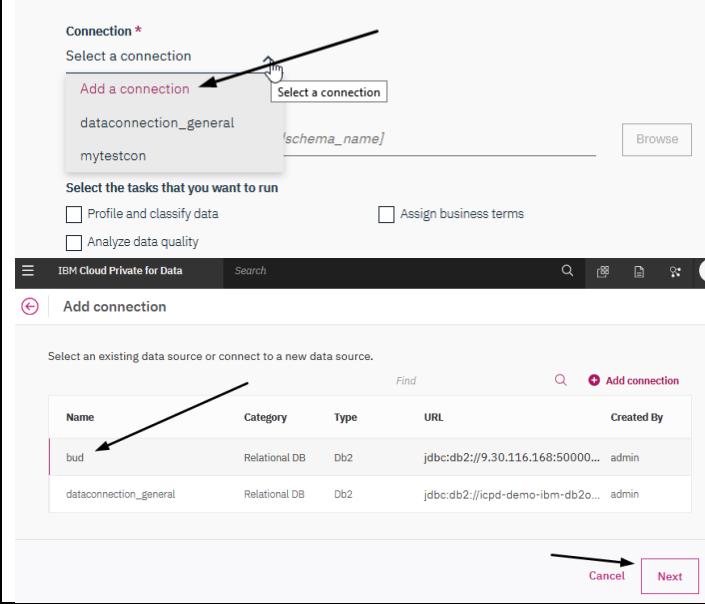
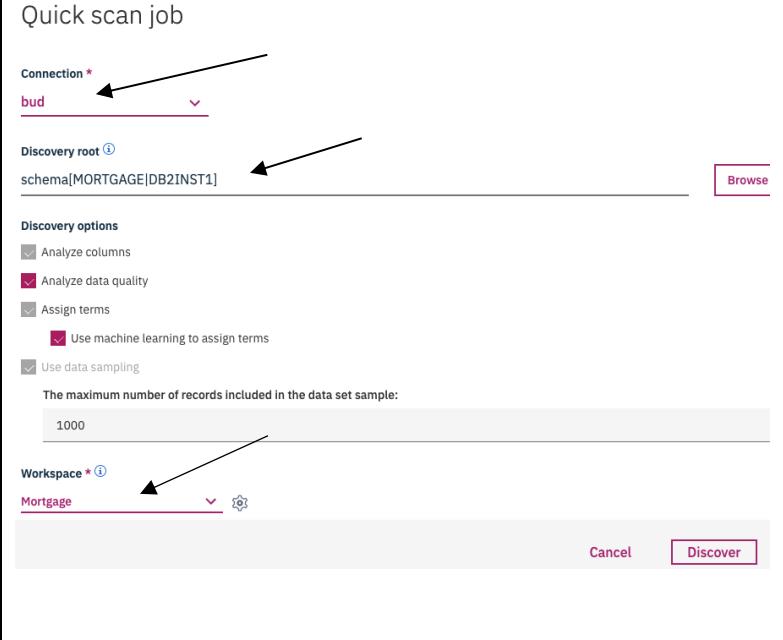
5. Discover Assets

Use the data source created above discover all data assets from Db2 database.

5.1. Navigate to discover assets

 6.	<p>From Organize option on the left pane, choose Metadata Curation > Data discovery.</p>
--	---

	<p>To select discover job</p> <p>Navigate to New discover job > Quick scan</p>
---	--

	<p>To discover assets</p> <ol style="list-style-type: none"> Click on Add a connection Choose the connection named bud that you created previously, click Next
	<ol style="list-style-type: none"> Choose the connection named bud that you created previously. Select Discover root as MORTGAGE > DB2INST1 Check necessary Discover options Click on Add a workspace under Workspace and named it as Mortgage. Click Create. Click on Discover It may take few minutes to complete.

<p>Click on View results or View workspaces to explore the discover assets.</p>																						
<p>Quick scan results New discovery job ▾ </p>																						
<table border="1"> <thead> <tr> <th>Summary</th> <th>Pending analysis</th> <th>Action required</th> <th>Reviewed</th> </tr> </thead> <tbody> <tr> <td>Status</td> <td> <input checked="" type="radio"/> All jobs pending analysis <input type="radio"/> Analyzing <input type="radio"/> In queue for analysis </td> <td colspan="2"> <input type="button" value="Pause"/> <input type="button" value="View results"/> <input type="button" value="1 item selected (select up to 15)"/> <input type="button" value="Cancel"/> </td> </tr> <tr> <td></td> <td></td> <td> <input checked="" type="checkbox"/> Job ID <input type="checkbox"/> Data assets <input type="checkbox"/> Connection <input type="checkbox"/> Started by <input type="checkbox"/> Processing time <input type="checkbox"/> Status <input type="checkbox"/> Status updated </td> <td></td> </tr> <tr> <td></td> <td></td> <td colspan="2"> <table border="1"> <tr> <td><input checked="" type="checkbox"/> qs_1571071613091 -</td> <td>bud</td> <td>admin</td> <td>2 minutes 15 seconds</td> <td>Analyzing -</td> </tr> </table> </td> </tr> </tbody> </table>	Summary	Pending analysis	Action required	Reviewed	Status	<input checked="" type="radio"/> All jobs pending analysis <input type="radio"/> Analyzing <input type="radio"/> In queue for analysis	<input type="button" value="Pause"/> <input type="button" value="View results"/> <input type="button" value="1 item selected (select up to 15)"/> <input type="button" value="Cancel"/>				<input checked="" type="checkbox"/> Job ID <input type="checkbox"/> Data assets <input type="checkbox"/> Connection <input type="checkbox"/> Started by <input type="checkbox"/> Processing time <input type="checkbox"/> Status <input type="checkbox"/> Status updated				<table border="1"> <tr> <td><input checked="" type="checkbox"/> qs_1571071613091 -</td> <td>bud</td> <td>admin</td> <td>2 minutes 15 seconds</td> <td>Analyzing -</td> </tr> </table>		<input checked="" type="checkbox"/> qs_1571071613091 -	bud	admin	2 minutes 15 seconds	Analyzing -	
Summary	Pending analysis	Action required	Reviewed																			
Status	<input checked="" type="radio"/> All jobs pending analysis <input type="radio"/> Analyzing <input type="radio"/> In queue for analysis	<input type="button" value="Pause"/> <input type="button" value="View results"/> <input type="button" value="1 item selected (select up to 15)"/> <input type="button" value="Cancel"/>																				
		<input checked="" type="checkbox"/> Job ID <input type="checkbox"/> Data assets <input type="checkbox"/> Connection <input type="checkbox"/> Started by <input type="checkbox"/> Processing time <input type="checkbox"/> Status <input type="checkbox"/> Status updated																				
		<table border="1"> <tr> <td><input checked="" type="checkbox"/> qs_1571071613091 -</td> <td>bud</td> <td>admin</td> <td>2 minutes 15 seconds</td> <td>Analyzing -</td> </tr> </table>		<input checked="" type="checkbox"/> qs_1571071613091 -	bud	admin	2 minutes 15 seconds	Analyzing -														
<input checked="" type="checkbox"/> qs_1571071613091 -	bud	admin	2 minutes 15 seconds	Analyzing -																		

6. Add users

Create users with different roles.

From **Administer** option on the left pane, choose **Manage users**.

Switch tab to 'Users' and click on 'Add user'

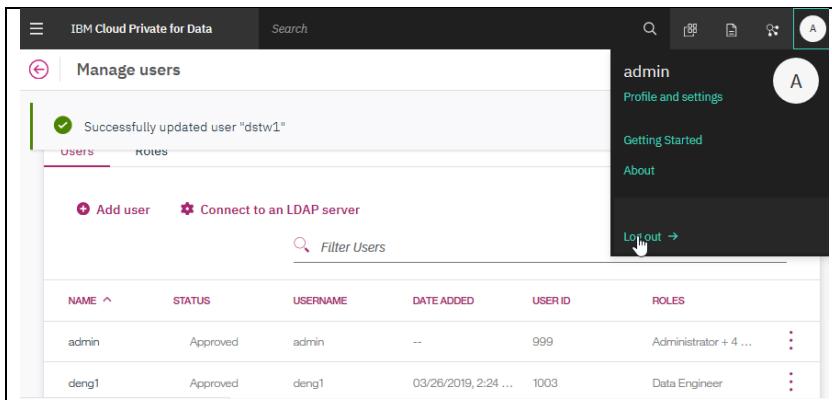
Fill out Add User information for a data scientist

1. Name as **dst1**
2. Username is **dst1**
3. Use a valid email address
4. Set Password as **dst1**
5. Choose the user roles as Data Scientist

Click on **Add** to confirm the add user

Follow same steps in Add User section (above) and two more account. Create **deng1** for Data Engineer and **dstw1** a data steward.

User	Role	Password
• deng1	Data Engineer	deng1
• dstw1	Data Stewards	dstw1



The screenshot shows the 'Manage users' page of the IBM Cloud Private for Data interface. At the top, there is a success message: 'Successfully updated user "dstw1"'. Below this, there are two tabs: 'USERS' (which is selected) and 'ROLES'. There are also links for 'Add user' and 'Connect to an LDAP server'. A search bar labeled 'Filter Users' is present. On the right side, a sidebar for the user 'admin' includes links for 'Profile and settings', 'Getting Started', 'About', and 'Logout' (with a mouse cursor hovering over it). The main table lists two users:

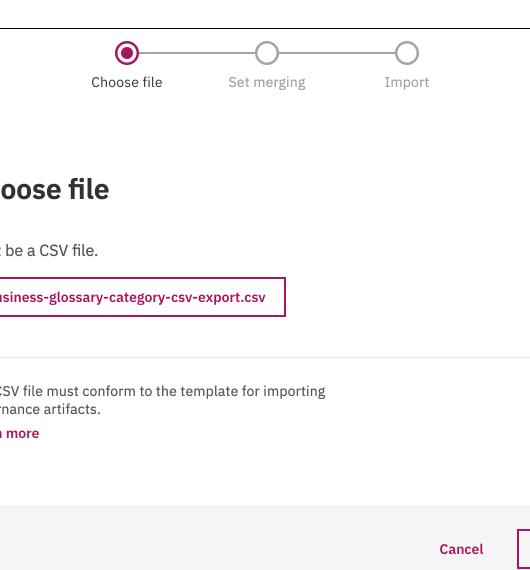
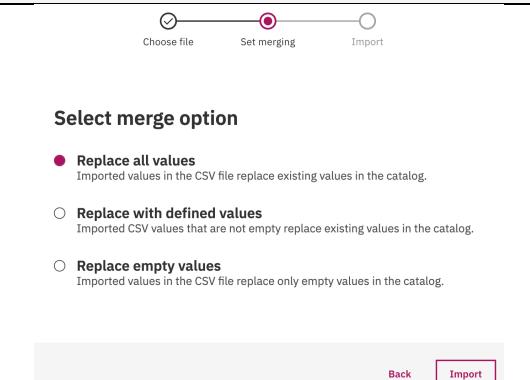
NAME ^	STATUS	USERNAME	DATE ADDED	USER ID	ROLES
admin	Approved	admin	--	999	Administrator + 4 ...
deng1	Approved	deng1	03/26/2019, 2:24 ...	1003	Data Engineer

Log out from user **admin**

7. Implement Categories and Terms

Cloud Pak for Data enables you to structure your enterprise information in a logical way, discover relationships between assets, and keep your data always up-to-date. You can import existing glossary with categories, terms, information governance policies and rules.

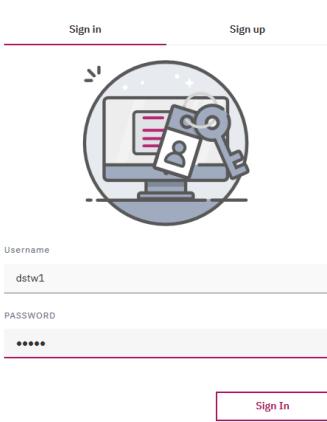
7.1. Import Categories

<p>Choose Organize > Data and AI governance > Categories from the left pane.</p> 		<p>Click on Import to import the CSV file contains category information that you downloaded from Git.</p>
 <p>Choose file</p> <p>Must be a CSV file.</p> <div style="border: 1px solid #ccc; padding: 2px; display: inline-block;">business-glossary-category-csv-export.csv</div> <p>The CSV file must conform to the template for importing governance artifacts. Learn more</p>		<p>Choose the CSV file location</p> <p>Click Next</p>
 <p>Select merge option</p> <ul style="list-style-type: none"> <input checked="" type="radio"/> Replace all values Imported values in the CSV file replace existing values in the catalog. <input type="radio"/> Replace with defined values Imported CSV values that are not empty replace existing values in the catalog. <input type="radio"/> Replace empty values Imported values in the CSV file replace only empty values in the catalog. 		<p>Select merge option as Replace all values</p> <p>Click Import</p>

7.2. Import Terms

8. Implement Policies and Rules

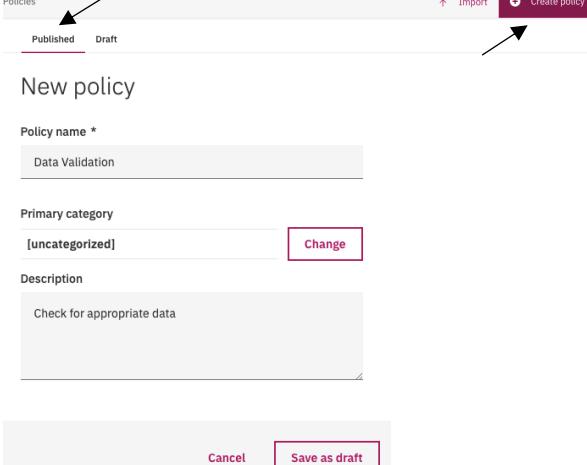
Create governance policies and rules for the entire organization to ensure clarity and compatibility among departments, projects, or products.

	<p>Sigh in to the Cloud Pak for Data web console as user 'dstw1' and password is 'dstw1' that you created earlier.</p>
---	--

7.2. Create a policy

Choose **Organize > Data and AI governance > Policy** from the left pane

Select **Published** tab and click on **Create Policy**

	<p>On the New policy window create a policy with following information and click on Save as draft:</p> <p>Name: Data Validation Description: Check for appropriate data</p> <p>It will take few minutes to appear under list of available policies.</p>
---	--

7.3. Create a rule

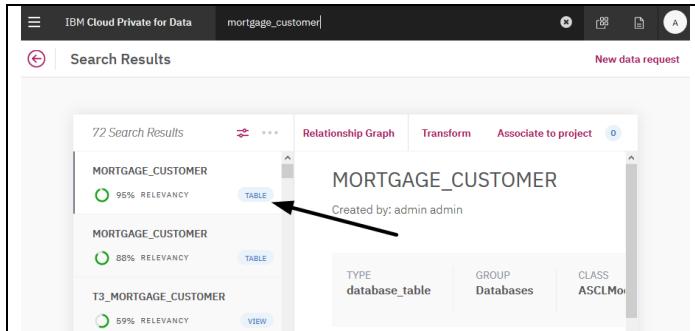
Choose **Organize > Data and AI governance > Rule** from the left pane

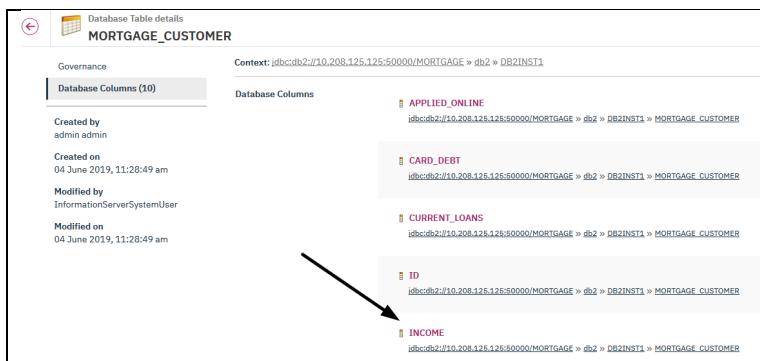
Select **Published** tab and click on **Create Rule**
 Choose **Governance rule**

New governance rule Governance rule name <input type="text" value="Income cannot be null"/> Description <input type="text" value="Income column must have a valid value"/> Cancel Save as draft	On the New governance rule window create a rule with following information and click on Save as draft : Name: Income cannot be null Referencing policies: Data Validation Short Description: Income column must have a valid value It will take few minutes to appear under list of available rules.
---	---

	Click on Add policy under Parent policies to assign the rule to it.
---	---

7.4. Add rule to metadata

	Click on the enterprise search, Search for 'mortgage_customer' and hit enter From the search results select table 'mortgage_customer' Click on Details tab at the top
---	--

	On Database Table Details window choose Database Columns from left Select INCOME column Next click on  icon (right top corner) and choose Edit
---	---

Scroll down to **Implement Rules** section

Search and select the rule **Income cannot be null** that you created earlier.

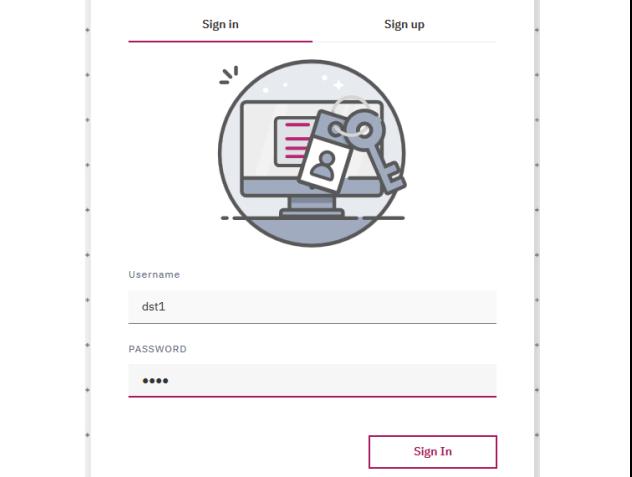
Click on **Save**

The screenshot shows the 'Database Column details' page for the column 'INCOME'. On the left, there's a sidebar with sections like 'Header (1)', 'General Information', 'Quality Analysis', 'Suggested Term Assignments', and 'Notes'. The main area has two tabs: 'Assigned to Terms' and 'Implements Rules'. Under 'Assigned to Terms', there's a section to 'Add to list' with a search bar and a 'Remove all' button. Under 'Implements Rules', there's a search bar with 'inco' typed in, a 'Remove all' button, and a list containing 'Income cannot be null'. At the top right of the page, there are 'Cancel' and 'Save' buttons. A large black arrow points from the 'Save' button to the 'Income cannot be null' rule in the 'Implements Rules' list.

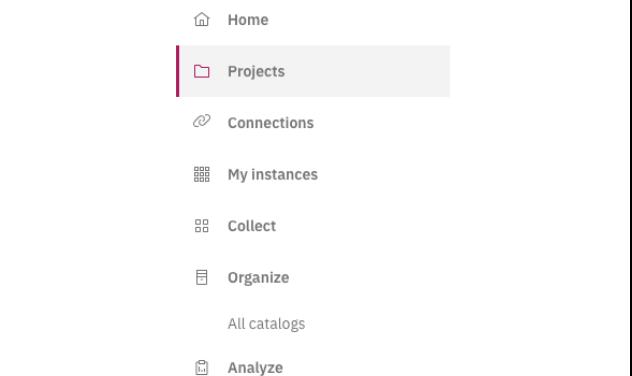
The screenshot shows the navigation bar of the Cloud Pak for Data interface. It features a dark sidebar with the user name 'dstw1' and links for 'Profile and settings', 'Getting Started', and 'About'. At the bottom of the sidebar is a 'Log out' link with a right-pointing arrow. To the right of the sidebar is a white panel with the text 'Log out from user 'dstw1''. A mouse cursor is hovering over the 'Log out' link in the sidebar.

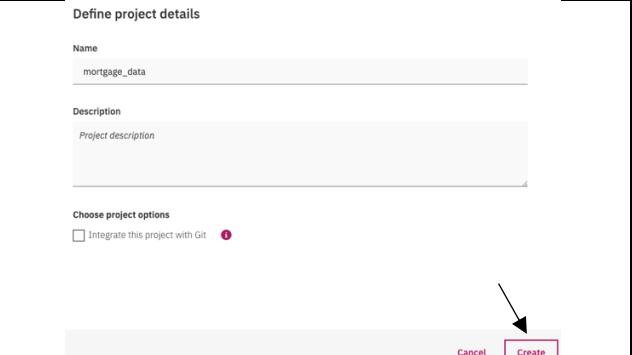
8. Access data as a Data Scientist

Explore the data require for build a model

	<p>Sigh in to the Cloud Pak for Data web console as user 'dst1' and password is 'dst1' that you created earlier.</p>
---	--

8.1. Create analytic project

	<p>Create a new analytical project by 'Projects' from right pane.</p> <p>Click on the  icon</p> <p>Select Create an empty project</p>
--	--

	<p>Provide a project name and click Create</p>
---	---

8.2. Assets from Glossary

Let's look for mortgage related terms in glossary to get an idea about different data assets available on the system.

Choose **Organize** from the left pane, select **Data Catalog -> Queries -> Glossary Categories and Terms.**

You should have all mortgage related information as follows. Click on each **ASSET NAME, TERMS** for additional information. The TERM DESCRIPTION provides a basic information about each term.

IBM Cloud Private for Data		Search	
Query Results: 10			
Category			
ASSET NAME	CATEGORY DESCRIPTION	TERMS	TERM DESCRIPTION
 Address Information	Location related glossary for a JK insurance customer	 Customer Zipcode  Continuity Of Address Segment  Address part 1  Customer City  Address part 2  Customer Street Suffix  Customer Street Name  Customer State  Customer House Label  Country Of Residence	Current zip code for customer's address Customer City Current suffix for street for customer's address Current street name for customer's address Current state of residence for a customer House number with optional suffix
 Crown Jewels	All data that is sensitive customer info per regulatory obligations	 Sensitive Personal Data	Any data deemed to be sensitive personal info for a customer
 Insurance Customer Details	Category for individual insurance customers	 Gender  Market Segment  Summary	Customer's gender, if known Customer Market Segment Summary information about a JKLV insured customer

For example, click on ASSET NAME **Customer**

a. Check Asset Details

Go through each item related to mortgage in glossary to have better idea about data you need for your project.

IBM Cloud Private for Data Search

Category Details: Customer

Terms (36)

Customer

Includes core customer concepts that support customer related areas of interest and analysis.

Industry Accelerators > Domains

Created by admin admin

Created on 12 April 2019, 9:25:03 am

Modified by admin admin

Modified on 12 April 2019, 9:25:03 am

Terms

Account Closed Date
Account Name
Account Number
Account Opened Date
Account Status
Account Status Date
Account Type

The asset **Customer** shows different terms associated with it.

Check each **Terms** for additional information.

b. Enterprise search

IBM Cloud Private for Data

Click on the enterprise search

mortgage

Search for ‘mortgage’ and hit enter

MORTGAGE_PROPERTY

MORTGAGE_PROPERTY

Database Table
jdbc:db2://9.30.116.168:50000 /MORTGAGE db2 > DB2INST1

★★★★★ 0 Ratings None Quality score

Description

Select your rating:

New Comment:
Write a comment

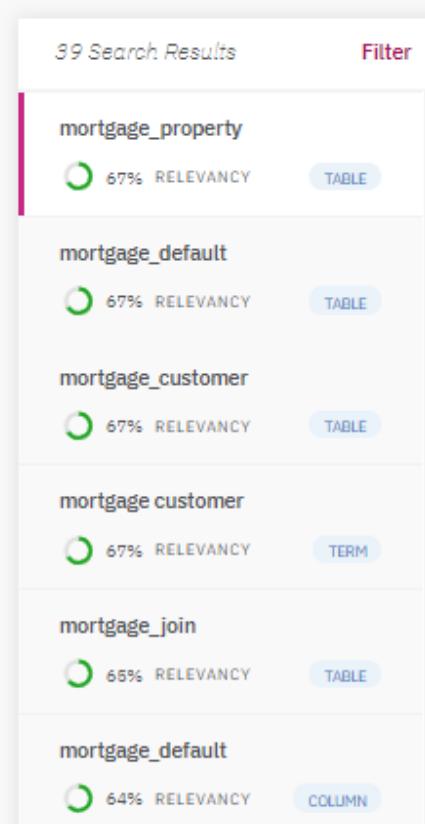
All Comments (0) Submit

Relationships

Choose the **mortgage_property** table and click on **Relationship Graph** to see details about the table.

Click on the ‘+’ next to **Database Column** to expand list of columns in the table.

Same way you can view other mortgage related tables.



39 Search Results

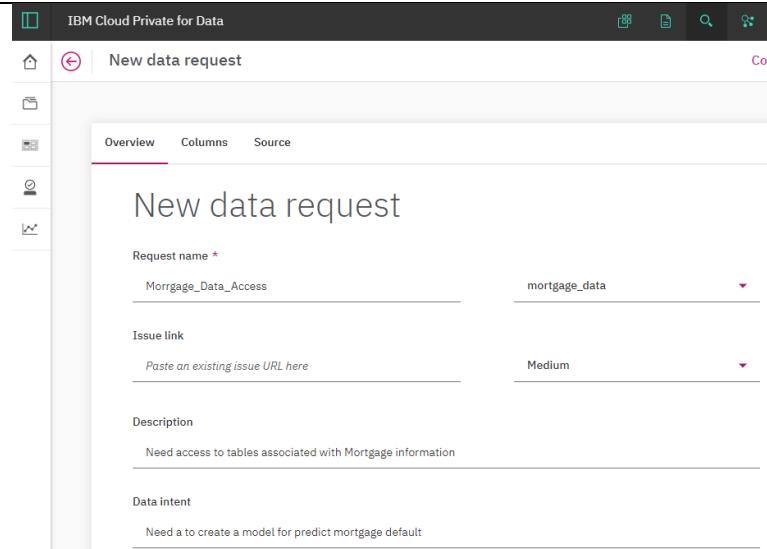
Filter

- mortgage_property
67% RELEVANCY TABLE
- mortgage_default
67% RELEVANCY TABLE
- mortgage_customer
67% RELEVANCY TABLE
- mortgage customer
67% RELEVANCY TERM
- mortgage_join
65% RELEVANCY TABLE
- mortgage_default
64% RELEVANCY COLUMN

Go back to the enterprise **Search Result**

The enterprise search will return all objects that mentioned word mortgage but as a data scientist you don't have access to any of those objects.

Click on the **New Data Request** on top right corner for request access to mortgage related datasets.



IBM Cloud Private for Data

New data request

Confirm

Overview Columns Source

New data request

Request name *

Mortgage_Data_Access

mortgage_data

Issue link

Paste an existing issue URL here

Medium

Description

Need access to tables associated with Mortgage information

Data intent

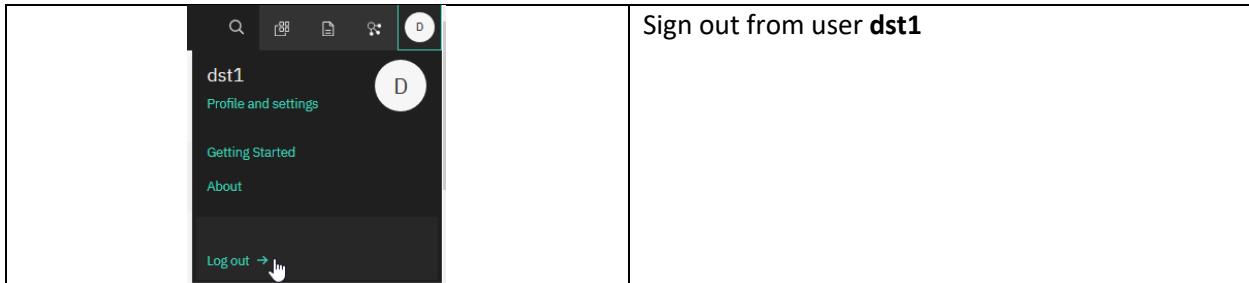
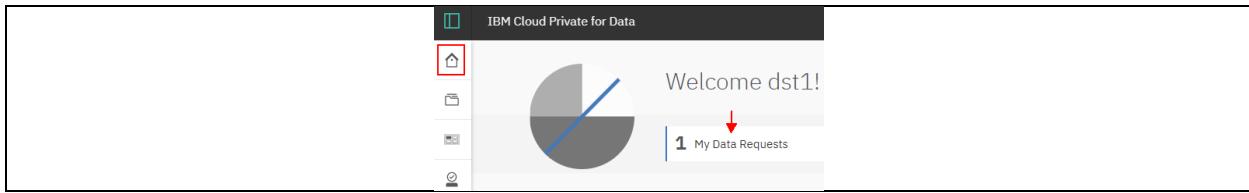
Need a to create a model for predict mortgage default

Fill up the **New Data Request** form with detail information as much possible, so a data engineer can provide accurate dataset. Click Confirm and then Submit request.

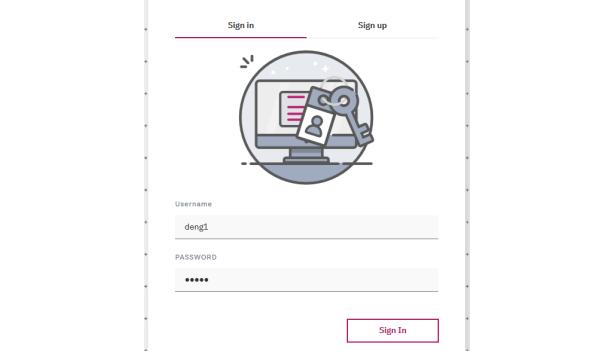
At this point you need to wait for data engineer to address the data request.

You can go to the home page by clicking on  icon from left pane and check the status of the data request.

Cloud Pak for Data – Tutorial



9. Review data request

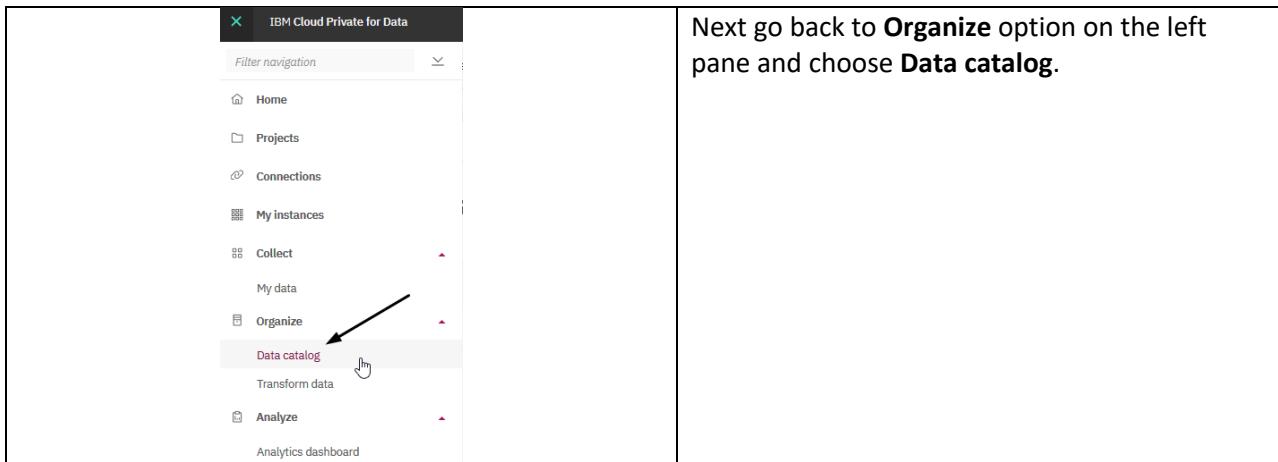
	<p>Sigh in to the Cloud Pak for Data web console as user 'deng1' and password is 'deng1' that you created earlier.</p>
---	--

	<p>After sing in Click on Go to your home page</p> <p>Check the Data Request tab on the home page.</p>
---	--

<p>Click on the new data request that submitted by data scientist earlier for review. After reviewing the request click on Action in top right corner and select assign to me.</p> <div style="border: 1px solid #ccc; padding: 10px; margin-top: 10px;"> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>ID</th> <th>Name</th> <th>Status</th> <th>Requested by</th> <th>Assigned to</th> <th>Priority</th> <th>Last updated</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>Mortgage_Data_Access</td> <td>Claimed</td> <td>dst1</td> <td>deng1</td> <td>Medium</td> <td>3 Jun 2019, 8:15 PM</td> </tr> <tr> <td>2</td> <td>Mortgage_Data_Access_Request</td> <td>Claimed</td> <td>dst1</td> <td>deng1</td> <td>Medium</td> <td>3 Jun 2019, 8:41 PM</td> </tr> <tr> <td>3</td> <td>Mortgage_Data_request1</td> <td>Claimed</td> <td>dst1</td> <td>deng1</td> <td>High</td> <td>3 Jun 2019, 8:39 PM</td> </tr> <tr> <td>4</td> <td>CustData</td> <td>New</td> <td>admin</td> <td>Unassigned</td> <td>High</td> <td>4 Jun 2019, 9:03 AM</td> </tr> </tbody> </table> <div style="margin-top: 10px;"> Action Transform data Assign to </div> </div>							ID	Name	Status	Requested by	Assigned to	Priority	Last updated	1	Mortgage_Data_Access	Claimed	dst1	deng1	Medium	3 Jun 2019, 8:15 PM	2	Mortgage_Data_Access_Request	Claimed	dst1	deng1	Medium	3 Jun 2019, 8:41 PM	3	Mortgage_Data_request1	Claimed	dst1	deng1	High	3 Jun 2019, 8:39 PM	4	CustData	New	admin	Unassigned	High	4 Jun 2019, 9:03 AM
ID	Name	Status	Requested by	Assigned to	Priority	Last updated																																			
1	Mortgage_Data_Access	Claimed	dst1	deng1	Medium	3 Jun 2019, 8:15 PM																																			
2	Mortgage_Data_Access_Request	Claimed	dst1	deng1	Medium	3 Jun 2019, 8:41 PM																																			
3	Mortgage_Data_request1	Claimed	dst1	deng1	High	3 Jun 2019, 8:39 PM																																			
4	CustData	New	admin	Unassigned	High	4 Jun 2019, 9:03 AM																																			

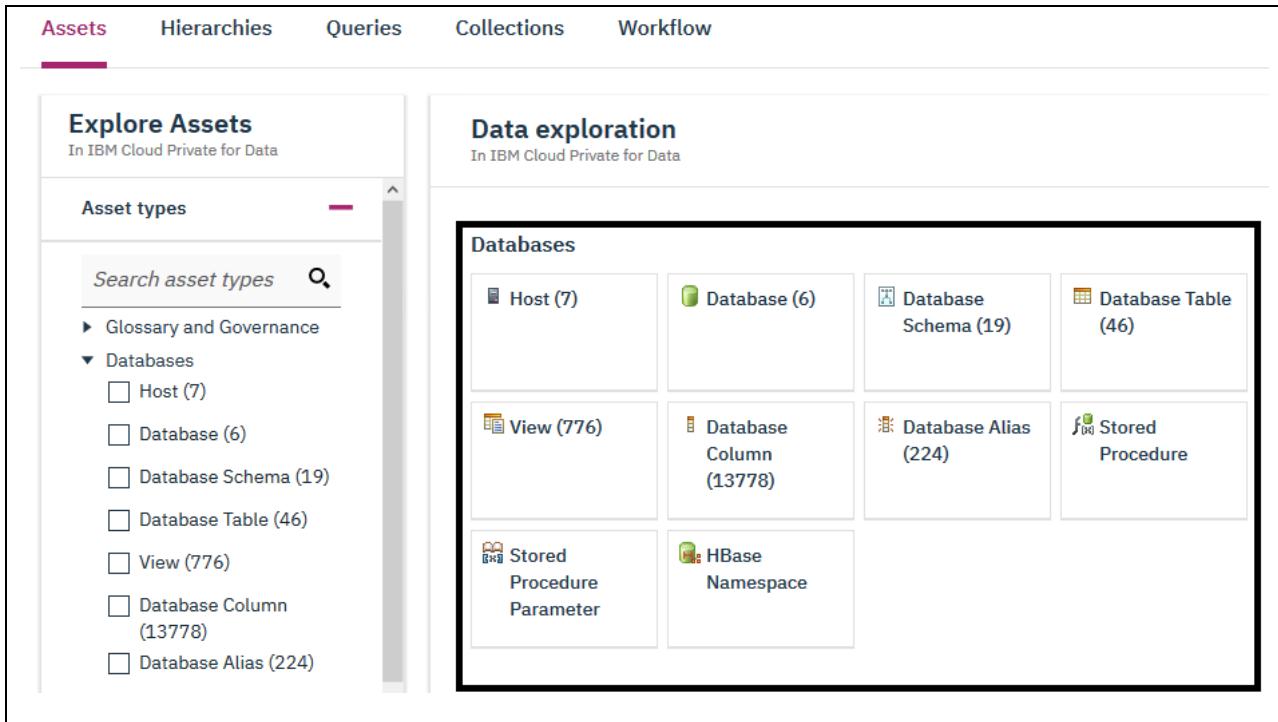
10. Navigate to data catalog

Once discover assets process completed. All database objects automatically cataloged in Cloud Pak for Data. You can review those database object in the catalog.



Next go back to **Organize** option on the left pane and choose **Data catalog**.

At this point Cloud Pak for Data should displays all the database objects. You can click each individual object under **Databases** to explore the catalog generated from discover asset previously. Click on the **Database Table** to check tables discovered from Db2. Take a look into the database named **mortgage**.



Under the **Database Tables** you can see ‘MORTGAGE_CUSTOMER’, ‘MORTGAGE_DEFAULT’ and ‘MORTGAGE_PROPERTY’ tables, cataloged from Db2 database.

The screenshot shows the 'All results' page in the Cloud Pak for Data interface. On the left, there is a 'Filter results' sidebar with a 'Clear all filters' button. Under 'Asset types (1)', 'Database Table (46)' is selected. Other asset types listed include Glossary and Governance, Databases (1), Host, Database, Database Schema, View, Database Column, Database Alias, Stored Procedure, Stored Procedure Parameter, HBase Namespace, and Data Files. A search bar for 'Search asset types' is also present. The main area displays 46 results, with the following entries visible:

- MORTGAGE**
InformationServerSystemUser on Jun 3, 2019, 6:52 PM
- HMON_COLLECTION**
Modified by InformationServerSystemUser on Jun 4, 2019, 11:30 AM
- MONGO_MORTGAGE_DEFAULT**
Created by admin on Jun 3, 2019, 6:34 PM
- MONGO_MORTGAGE_PROPERTY**
Modified by InformationServerSystemUser on Jun 3, 2019, 6:34 PM
- MORTGAGE_CUSTOMER**
Created by admin on Jun 4, 2019, 11:28 AM

No items selected is displayed at the top right of the results list.

11. Data Virtualization

Context: Data virtualization (DV) integrates data sources across multiple types and locations and turns it into one logical data view. In this case, you have data across three different tables. Creating a virtual table you can quickly view data from different tables.

11.1. Adding a new data source for Db2

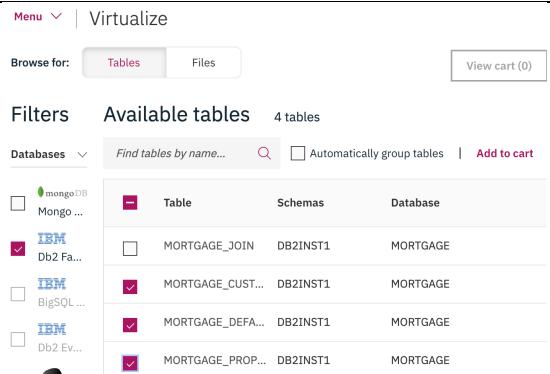
Context: DV supports many relational and non-relational data sources (as well as files that reside on a local disk or network file system) that you can add to your data source ecosystem. After a data source has been added, any user that has virtualize permission can create virtual tables. DV agents connect to relational data sources using JDBC protocol. In this tutorial you will add a data source for Db2 database.

Define a data connection to Db2. Use your existing Db2 database connection for Db2 data source.

1. Go to **Collect > Virtualized data > Menu > Data sources**
2. Click **Add > New data source > Add connection**
3. Select **Db2** that you created earlier and click **Next**

11.3. Select tables for virtualization

Context: the most common mechanism for virtualizing data is to create a "view" or virtual table. Virtual tables can be full or segment of data from one or more tables. You can then run queries against the resulting virtual table.

<ul style="list-style-type: none"> • Click Collect > Virtualized data > Menu > Virtualize • Select tables MORTGAGE_CUSTOMER, MORTGAGE_PROPERTY and MORTGAGE_DEFAULT from MORTGAGE database, then click Add to cart • Click View cart • Click Next 	 <p>The screenshot shows the 'Virtualize' screen with the following details:</p> <ul style="list-style-type: none"> Menu dropdown: Virtualize Browse for: Tables Filters: Databases dropdown set to IBM Db2 FA... Available tables: 4 tables listed: <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">Table</th> <th style="text-align: left;">Schemas</th> <th style="text-align: left;">Database</th> </tr> </thead> <tbody> <tr> <td>MORTGAGE_JOIN</td> <td>DB2INST1</td> <td>MORTGAGE</td> </tr> <tr> <td>MORTGAGE_CUST...</td> <td>DB2INST1</td> <td>MORTGAGE</td> </tr> <tr> <td>MORTGAGE_DEFA...</td> <td>DB2INST1</td> <td>MORTGAGE</td> </tr> <tr> <td>MORTGAGE_PROP...</td> <td>DB2INST1</td> <td>MORTGAGE</td> </tr> </tbody> </table> Databases dropdown: IBM Db2 FA... Find tables by name... input field Automatically group tables checkbox Add to cart button 	Table	Schemas	Database	MORTGAGE_JOIN	DB2INST1	MORTGAGE	MORTGAGE_CUST...	DB2INST1	MORTGAGE	MORTGAGE_DEFA...	DB2INST1	MORTGAGE	MORTGAGE_PROP...	DB2INST1	MORTGAGE
Table	Schemas	Database														
MORTGAGE_JOIN	DB2INST1	MORTGAGE														
MORTGAGE_CUST...	DB2INST1	MORTGAGE														
MORTGAGE_DEFA...	DB2INST1	MORTGAGE														
MORTGAGE_PROP...	DB2INST1	MORTGAGE														

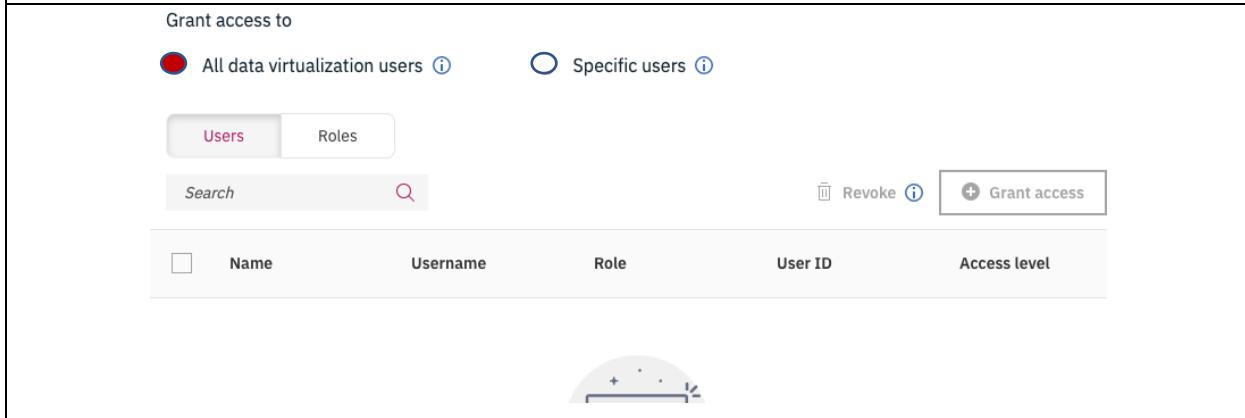
<ul style="list-style-type: none"> • Select Neither • Uncheck the box for Submit to catalog • Click Virtualize to complete the process 	<p style="text-align: right;"><input type="checkbox"/> Submit to catalog Cancel Virtualize</p> <p>Assign to</p> <p><input type="radio"/> Data request ? <input type="radio"/> Project ? <input checked="" type="radio"/> Assign to My virtualized data only</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Table</th> <th>Schema</th> <th>Source schema</th> <th>Host/Database</th> <th>Grouped tables</th> </tr> </thead> <tbody> <tr> <td>MORTGAGE_CUSTOMER</td> <td>USER999</td> <td>X ▾</td> <td>DB2INST1</td> <td>169.46.33.180:MORTGAGE</td> </tr> <tr> <td>MORTGAGE_DEFAULT</td> <td>USER999</td> <td>X ▾</td> <td>DB2INST1</td> <td>169.46.33.180:MORTGAGE</td> </tr> <tr> <td>MORTGAGE_PROPERTY</td> <td>USER999</td> <td>X ▾</td> <td>DB2INST1</td> <td>169.46.33.180:MORTGAGE</td> </tr> </tbody> </table>	Table	Schema	Source schema	Host/Database	Grouped tables	MORTGAGE_CUSTOMER	USER999	X ▾	DB2INST1	169.46.33.180:MORTGAGE	MORTGAGE_DEFAULT	USER999	X ▾	DB2INST1	169.46.33.180:MORTGAGE	MORTGAGE_PROPERTY	USER999	X ▾	DB2INST1	169.46.33.180:MORTGAGE
Table	Schema	Source schema	Host/Database	Grouped tables																	
MORTGAGE_CUSTOMER	USER999	X ▾	DB2INST1	169.46.33.180:MORTGAGE																	
MORTGAGE_DEFAULT	USER999	X ▾	DB2INST1	169.46.33.180:MORTGAGE																	
MORTGAGE_PROPERTY	USER999	X ▾	DB2INST1	169.46.33.180:MORTGAGE																	

11.4. Creating virtual table

You can create a new virtual table based on existing tables under **My data** section. You can use “drag and drop” or write your own SQL to create the view.

<ul style="list-style-type: none"> • Click Collect > Virtualized data > Menu > SQL editor to access the editor. • Copy the following SQL statement and paste it on the editor • Click on Run all 	<pre>CREATE VIEW MORTGAGE_JOIN_VIEW AS SELECT A.ID, INCOME, APPLIED_ONLINE, RESIDENCE, YRS_CURRENT_ADD, YRS_CURRENT_EMP, NO_OF_CARDS, CARD_DEBT, CURRENT_LOANS, LOAN_AMOUNT, SALE_PRICE, LOCATION, MORTGAGE_DEFAULT FROM MORTGAGE_CUSTOMER A, MORTGAGE_PROPERTY B, MORTGAGE_DEFAULT C WHERE A.ID = B.ID AND A.ID = C.ID;</pre>
<p style="text-align: center;">Menu ▼ SQL editor</p> <div style="border: 1px solid #ccc; padding: 5px; margin-top: 5px;"> * Untitled - 1 ⊕ Syntax assistant </div> <pre style="font-family: monospace; font-size: small; margin-top: 10px;">1 2 3 CREATE VIEW MORTGAGE_JOIN_VIEW 4 AS 5 SELECT A.ID, INCOME, APPLIED_ONLINE, RESIDENCE, YRS_CURRENT_ 6 YRS_CURRENT_EMP, NO_OF_CARDS, CARD_DEBT, CURRENT_LOAN 7 LOAN_AMOUNT, SALE_PRICE, LOCATION, MORTGAGE_DEFAULT 8 FROM MORTGAGE_CUSTOMER A, 9 MORTGAGE_PROPERTY B, 10 MORTGAGE_DEFAULT C 11 WHERE A.ID = B.ID 12 AND A.ID = C.ID;</pre>	
<ul style="list-style-type: none"> • Click Collect > Virtualized data > Menu > My virtualized data to access the virtual table MORTGAGE_JOIN_VIEW • Check the box associated with MORTGAGE_JOIN_VIEW 	

- Click on the table actions menu 
- Select **Manage access** option
- On grant access window select All data virtualization users
- Click **Continue**



Grant access to

All data virtualization users [\(i\)](#) Specific users [\(i\)](#)

[Users](#) [Roles](#)

Search 

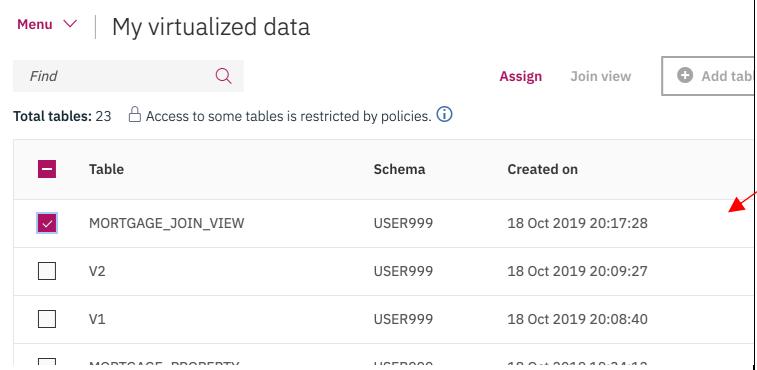
 Revoke [\(i\)](#)  Grant access

<input type="checkbox"/>	Name	Username	Role	User ID	Access level
					

11.5. Add virtual table to catalog

Once you create a virtual table, you can add it to the catalog, making it easily searchable.

- Click **Collect > Virtualized data > Menu > My data** to find the virtual table just created.
- Mark the checkbox associated with virtual table
- Choose **Submit to catalog** from table action
- Click on **Confirm**



Menu  | My virtualized data

Find 

Assign Join view 

Total tables: 23  Access to some tables is restricted by policies. [\(i\)](#)

 Table	Schema	Created on
<input checked="" type="checkbox"/> MORTGAGE_JOIN_VIEW	USER999	18 Oct 2019 20:17:28
<input type="checkbox"/> V2	USER999	18 Oct 2019 20:09:27
<input type="checkbox"/> V1	USER999	18 Oct 2019 20:08:40
<input type="checkbox"/> MORTGAGE_PROPERTY	USER999	18 Oct 2019 20:08:40

11.6. Publish virtualized table

A data steward needs approve the published request before the asset is added to the enterprise data catalog. You signed in as user 'admin', it should allow to publish the virtual table.

Pending Publish to Catalog Requests					
Name	Type	Project	Owner	Date Updated	Status
> USER999.MORTGAGE_JOIN_VIEW	view	-	admin	21 October 2019, 2:49PM	Pending
> USER999.Currency USER999.Country	table	-	admin	17 October 2019, 8:40AM	Pending

- Click on access the **Home** page
- Click on **Pending Publish to Catalog Requests**
- Click on icon on left for virtual table **MORTGAGE_JOIN_VIEW** that you created
- Click on **Approve**

11.7. Access information for virtual table

To access virtual table from external application, you need the JDBC connection information. Click on **Collect > Virtualized data > Menu > Add-on settings** to find out access information. You will use this information later in the building model section.

The screenshot shows the 'Add-on settings' page with the following details:

- Access information** section
- User ID**: user999
- Password**: (with **Show** link)
- JDBC connection URL**: jdbc:db2://dv-server.zen.svc.cluster.local:32051/...

11.8. Deliver Dataset

Go to the home page by clicking on icon from left pane and check the data request tab.

Click on the data request for update that submitted by data scientist earlier.

Data requests + Add new data request

	Name	ID	Status	Last Updated
1	Mortgage_Data_Access	2	New	27 Mar 2019, 11:15 AM

Click on the **Source** and fill out all the necessary information. This information will be picked up by the data scientist later.

Add the **remote data** set information that you created during data transformation. In this case remote data set is MORTGAGE_JOIN_VIEW. Use the **Access information** from the **Add-on settings** information from DV.

New data request

Overview Columns Source

Source

Data source name: mortgage_join

DB2

Username: db2inst1

Password: *****

JDBC URL: 169.45.83.218

[+ Add new dataset](#)

	Remote data set name	Description	Schema	Table
1	mortgage_join		db2inst1	mortage_join

Click on the data request and change the status to **Deliver**.

NAME	ID	STATUS	REQUESTED BY	ACCEPTED BY	LAST UPDATED	ACTIONS
mortagedata1	7	Delivered	dst1	deng1	6 Aug 2018, 12:55 PM	Deliver
Mortgage_Data_Access	9	Accepted	dst1	deng1	15 Aug 2018, 1:17 AM	Deliver

IBM Cloud Private for Data

Welcome deng1

signed in as: deng1

Getting Started

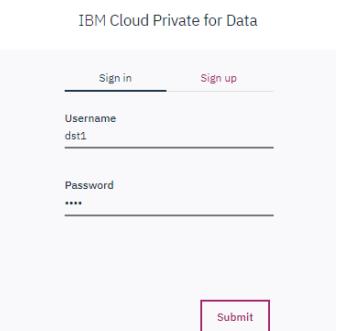
Settings

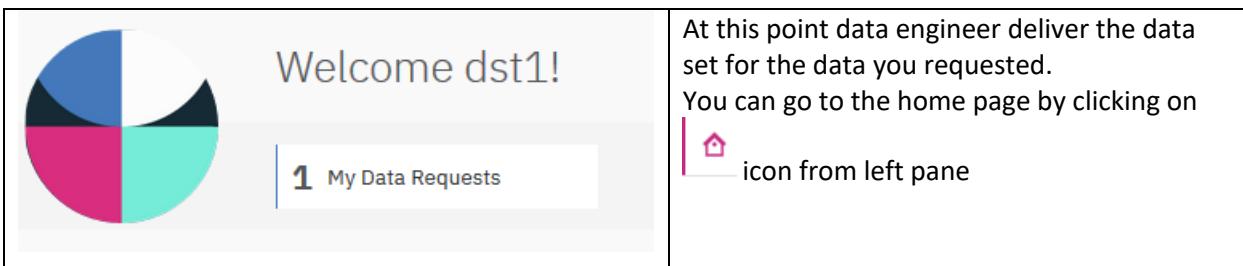
Sign Out

Sign out from user **deng1**

12. Build Model

With Cloud Pak for Data, you can collaborate with other team members on analytic projects to create visualizations and machine learning models with data from your enterprise. In this step you will build a simple model to predict the possibilities of mortgage default by customer. The object of this model is to show the functionality of Cloud Pak for Data, not the prediction accuracy. One can use lot more data and build a complex algorithm to get better accuracy.

	<p>Sign in to the Cloud Pak for Data web console as user 'dst1' and password is 'dst1' that you created earlier.</p>
---	--



12.1. Navigate to analytics project

Select **Projects** option from the left pane and click on the analytics project 'mortgage_data' that you created earlier.

12.2. Create deployment space

Create a separate deployment space for your project 'mortgage_data'.

Choose : My Projects > **mortgage_data** > Settings > Associate a deployment space > New

<p>Connect to a deployment space</p> <p>New Existing</p> <p>Name MortgageDeploymentSpace</p> <p>Description (Optional) <i>Description of deployment space</i></p> <p style="text-align: right;">Cancel Associate</p>	<p>Name new deployment space as 'MortgageDeploymentSpace'</p> <p>Click on Associate</p>
--	--

12.3. Create notebook

Create a notebook from a predefined Jupyter notebook that available on Github.

<ul style="list-style-type: none"> • Go to : My Projects > mortgage_data > Add to project • Choose asset type as Notebook • The new notebook needs to create from URL • Name the notebook as MortgageNotebook • Use notebook URL as https://github.com/IBM-ICP4D/icp4d-tutorials/blob/master/assets/mortgage-002/MortgageNotebook.V25.jupyter-py36.ipynb • Click on Create Notebook 	<p>My Projects > mortgage_data > Add Notebook</p> <p>New notebook</p> <p>Blank From file From URL</p> <p>Name MortgageNotebook 24 characters remaining</p> <p>Description (optional) <i>Type your Description here</i> 500 characters remaining</p> <p>Select runtime Default Python 3.6 (1 vCPU and 2 GB RAM) ▾</p> <p>Notebook URL https://github.com/IBM-ICP4D/icp4d-tutorials/blob/master/assets/mortgage-002/MortgageNotebook.V25</p> <p style="text-align: right;">Cancel Create Notebook</p>
---	---

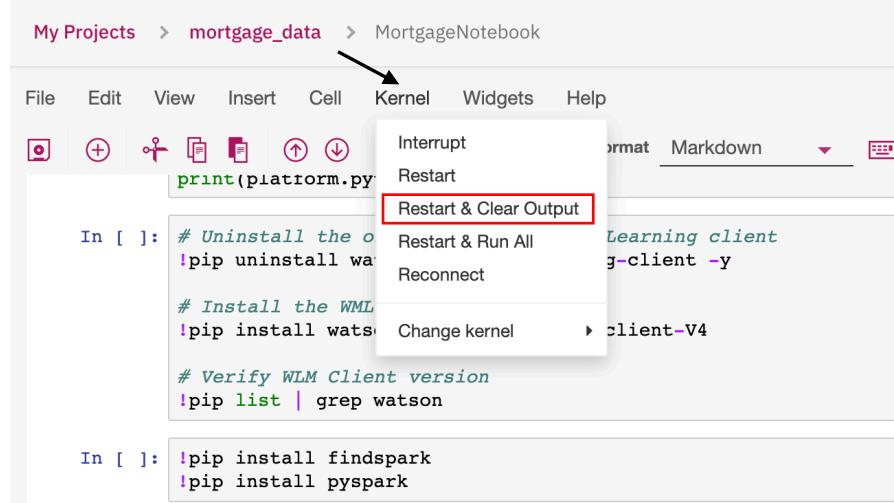
12.4. Review and run notebook

The majority of the code in the notebook is standard open source code that's used for various steps in the predictive analytics process.

Switch to edit mode by clicking on  icon from top of the screen.

Do not run all cells at once. Follow the instruction below to run the notebook.

Run the **Step 1: Intall** section first. Once all package installed make sure restart the Python kernel before move on next step.



The screenshot shows a Jupyter Notebook interface. At the top, there is a navigation bar with 'My Projects' → 'mortgage_data' → 'MortgageNotebook'. Below this is a menu bar with 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', 'Widgets', and 'Help'. An arrow points from the text above to the 'Kernel' menu item. The 'Kernel' menu is open, displaying options: 'Interrupt', 'Restart', 'Restart & Clear Output' (which is highlighted with a red box), 'Restart & Run All', 'Reconnect', 'Change kernel' (with a dropdown menu showing 'client-v4'), and 'Learning client' with a sub-option 'client -y'. In the main notebook area, there are two code cells. The first cell, labeled 'In []:', contains Python commands to uninstall WML, install Watson Machine Learning, verify the version, and install findspark and pyspark. The second cell, also labeled 'In []:', contains similar commands. The code uses green for syntax highlighting.

Action: restart the kernel!

Go the **Step 2: Authenticate** section and update the **url**, **username** and **password** fields with your CPD UI console details and access credential.

Step 2: Authenticate

```
[ ]: WML_CREDENTIALS = {
    "instance_id": "openshift",
    "url" : "https://zen-cpd-zen.apps.testcluster.demo.ibmcloud.com",
    "username": "admin",
    "password": "passw0rd",
    "version": "2.5.0"
}
```

In the next notebook cell, update the **dsn_url**, **dsn_uid** and **dsn_pwd** values with the information available from **Collect > Virtualized data > Menu > Add-on settings**.

```
[ ]: #Enter the values for your database connection found under data virtualization
dsn_url = "jdbc:db2://dv-server.zen.svc.cluster.local:32051/bigsql" # e.g.
dsn_uid = "user1022" # e.g.
dsn_pwd = "sw?#@lt_674MfPI5" # e.g.
```

Run all cells between step 2 and 6.

On **Step 7: Set default space**, run the first cell and find out the **GUID** for space name **MortgageDeploymentSpace**.

On the next cell replaced the **GUID** with one that you found above.

```
In [ ]: # Example: client.set.default_space('b49e13e8-ec68-408d-84a1-957e28c154b1')
client.set.default_space('GUID')
```

Run through remaining cells, so that it generates and deployed the model.

Before exit, save the notebook .

12.5. Test the model

Go to: Analyze > Analytics deployment to access deployed model

Select the **MortgageDeploymentSpace** from the list of analytic deployment space

Click on the **MORTGAGE PREDICTION MODEL**

Choose the **MORTGAGE PREDICTION** model

Click on **Test** tab

<pre>{ "input_data": [{ "fields": ["INCOME", "APPLIED_ONLINE", "RESIDENCE", "YRS_CURRENT_ADD", "YRS_CURRENT_EMP", "NO_OF_CARDS", "CARD_DEBT", "CURRENT_LOANS", "CREDIT_SCORE"] }] }</pre>	<p>Copy this sample data and paste it on the Enter input data box.</p> <p>Click on Predict</p>
---	--

```
        "LOAN_AMOUNT",
        "SALE_PRICE",
        "LOCATION"
    ],
    "values": [
        [
            43151,
            "N",
            "P",
            6,
            9,
            1,
            750,
            1,
            8600,
            320000,
            110
        ]
    ]
}
```

According on input values, model will predict and displays the result.

Analytics deployment spaces > MortgageDeploymentSpace > MORTGAGE PREDICTION MODEL > MORTGAGE PREDICTION

ONLINE

MORTGAGE PREDICTION

API reference Test

Enter input data

Result

Body

```
{  
  "input_data": [  
    {  
      "fields": [  
        "INCOME",  
        "APPLIED_ONLINE",  
        "RESIDENCE",  
        "YRS_CURRENT_ADD",  
        "YRS_CURRENT_EMP",  
        "NO_OF_CARDS",  
      ]  
    }  
  ]  
}
```

Predict

0 {
1 "predictions": [
2 {
3 "fields": [
4 "INCOME",
5 "APPLIED_ONLINE",
6 "RESIDENCE",
7 "YRS_CURRENT_ADD",
8 "YRS_CURRENT_EMP",
9 "NO_OF_CARDS",
10 "CARD_DEBT",
11 "CURRENT_LOANS",
12 "LOAN_AMOUNT",
13 "SALE_PRICE",
14 "LOCATION",
15 "MORTGAGE_DEFAULT",
16 "AppliedOnlineEncoded",
]
 }
]
}

MORTGAGE PREDICTION

Deployed

Created Nov 07, 2019 11:48 PM

Updated Nov 08, 2019 06:32 PM

Deployment ID b7a58231-fd99-4d9f-a760-7d81...

Software /v4/runtimes/spark-mllib_2.3

Description No description provided

Associated asset

Model MORTGAGE PREDICTION M...

Model ID 4809b65e-9cab-4870-b93c-7444...