



Cloud Pak for Data

Tutorial – Mortgage

Contents

| | | |
|--------|---|----|
| 1. | Prerequisites..... | 4 |
| 3. | Access Credentials | 5 |
| 3.1. | Access credential for Db2 database..... | 5 |
| 3.2. | Sign in to Cloud Pak for Data web console as Administrator..... | 5 |
| 4. | Create Connection | 6 |
| 4.1. | Navigate to Connections..... | 6 |
| 4.2. | Add connection..... | 7 |
| 5. | Discover Assets | 7 |
| 5.1. | Navigate to discover assets | 8 |
| 7. | Implement Policies and Rules..... | 12 |
| 8. | Access data as a Data Scientist | 15 |
| 10. | Navigate to data catalog..... | 21 |
| 11. | Analyze Data | 23 |
| 11.1. | Navigate to Create Dashboard..... | 23 |
| 11.2. | Add a data source | 23 |
| 11.3. | Explore Visualizations and Widgets | 24 |
| 11.4. | Add a graph..... | 24 |
| 11.5. | Modify the graphs..... | 25 |
| 12. | Transform Data | 26 |
| 12.1. | Navigate to transform data | 26 |
| 12.2. | Create a Project | 26 |
| 12.3. | Create a job..... | 27 |
| 12.4. | Add tables from asset browser | 28 |
| 12.5. | Join tables | 30 |
| 12.6. | Preview output from join..... | 31 |
| 12.7. | Store output from join | 32 |
| 12.8. | Transform output data | 33 |
| 12.9. | Apply governance rule | 34 |
| 12.10. | Execute job | 34 |
| 12.11. | Preview output data | 35 |
| 13. | Build Model..... | 37 |
| 13.1. | Navigate to analytics project | 38 |

13.2. Create a model 38

13.3. Review and Run notebook..... 39

13.4. Test the model..... 40

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

2.1. 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.

2.2. From your home directory, clone the tutorial sample files:

```
git clone https://github.com/sanjitc/ICP4XTutorial.git
```

2.3. 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.

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

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

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

```
./load_samples.sh -t mortgage-001
```

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

Sign up



Username

admin

PASSWORD

•••••

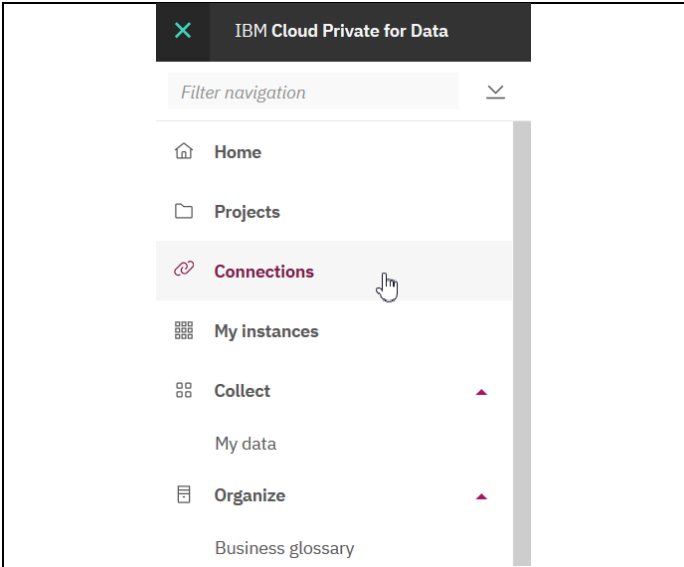

Sign In

Sign in to the Cloud Pak for Data web console as user 'admin' and password is 'password'.

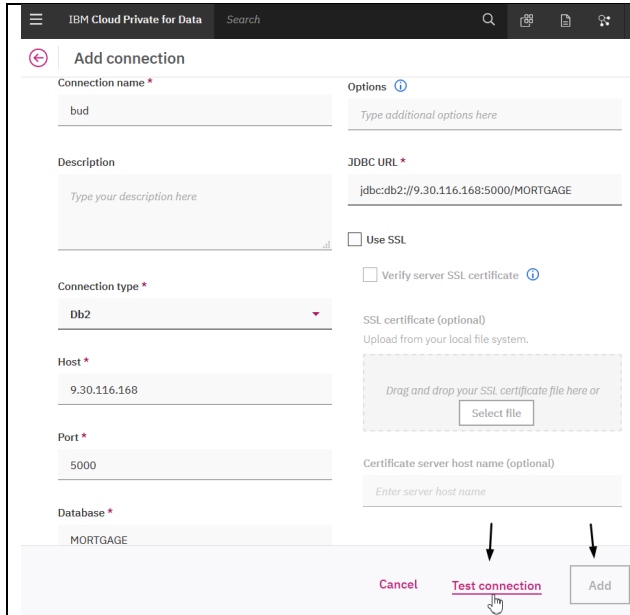
4. Create Connection

Create a connection to the data source for Db2 database.

4.1. Navigate to Connections

| | |
|---|--|
|  <p>The screenshot shows the left-hand navigation pane of the IBM Cloud Private for Data interface. At the top, there is a dark header with the text 'IBM Cloud Private for Data' and a close icon. Below this is a search bar labeled 'Filter navigation'. The main list of navigation items includes: 'Home' (house icon), 'Projects' (folder icon), 'Connections' (link icon, highlighted with a mouse cursor), 'My instances' (grid icon), 'Collect' (grid icon), 'My data' (document icon), 'Organize' (document icon), and 'Business glossary' (document icon). The 'Connections' item is currently selected and highlighted in a light gray color.</p> | <p>On the left pane choose Connections.</p> <p>Next, on the Data Connections window click on the  Add connection icon.</p> |
|---|--|

4.2. 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.

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

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

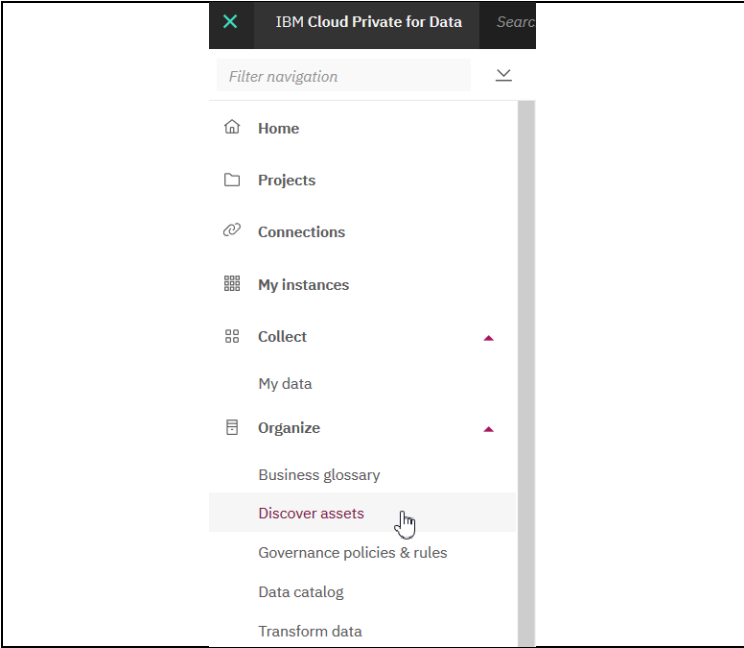


Success The test connection was successful. Click Add to save the connection information.

5. Discover Assets

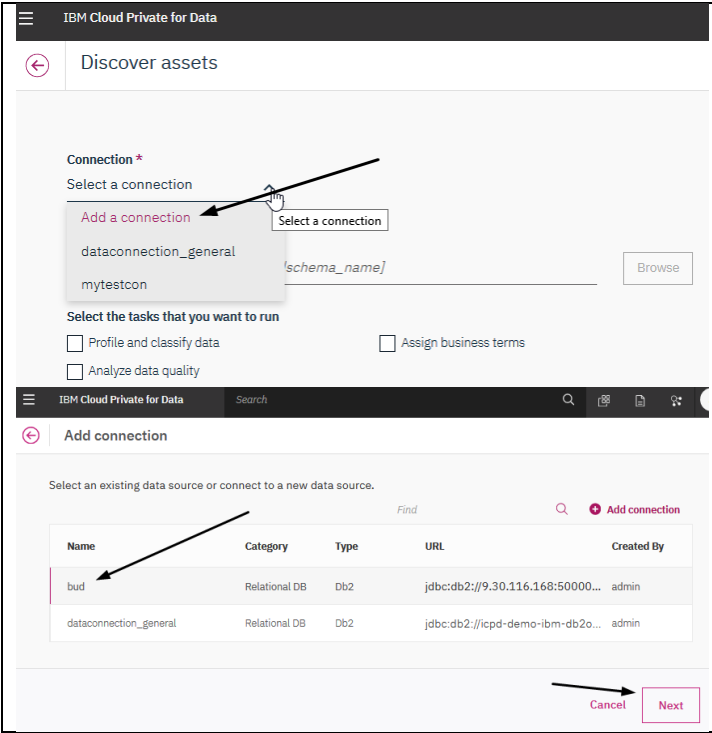
Use the data source created above discover all data assets from Db2 database on IBM Bluemix.

5.1. Navigate to discover assets



The screenshot shows the left-hand navigation pane of the IBM Cloud Private for Data interface. The menu items are: Home, Projects, Connections, My instances, Collect, My data, Organize, Business glossary, Discover assets (highlighted with a mouse cursor), Governance policies & rules, Data catalog, and Transform data. The 'Discover assets' option is highlighted in a light blue color.

From **Organize** option on the left pane, choose **Discover assets**.



The screenshot shows two screens from the IBM Cloud Private for Data interface. The top screen is 'Discover assets', which has a 'Connection *' section with a 'Select a connection' dropdown and an 'Add a connection' button. The bottom screen is 'Add connection', which shows a table of existing data sources. The 'bud' connection is highlighted in the table. Arrows indicate the flow from the 'Discover assets' screen to the 'Add connection' screen, and then to the 'Next' button.

| Name | Category | Type | URL | Created By |
|------------------------|---------------|------|----------------------------------|------------|
| bud | Relational DB | Db2 | jdbc:db2://9.30.116.168:50000... | admin |
| dataconnection_general | Relational DB | Db2 | jdbc:db2://icpd-demo-ibm-db2o... | admin |

To discover assets

1. Add a connection
2. Choose the connection named 'bud' that you created previously, click Next

The screenshot shows two parts of the IBM Cloud Private for Data interface. The top part is the 'Add connection' screen, where a connection named 'bud' is being created with a JDBC URL. The bottom part is the 'Discover assets' screen, where the 'bud' connection is selected, and the 'Discover' button is highlighted. Arrows indicate the sequence of steps: adding the connection, selecting it, and then clicking 'Discover'.

Add connection

This connection can be used to discover assets.

Connection name *

Options ⓘ

Description

JDBC URL *

☐ Use SSL

Cancel Test connection **Add**

Discover assets

Discovery results ⓘ

Connection * **Bud**

Discovery root ⓘ **Browse**

☐ Preview scan ⓘ

Discovery options

☒ Analyze columns ☒ Assign terms

☒ Analyze data quality

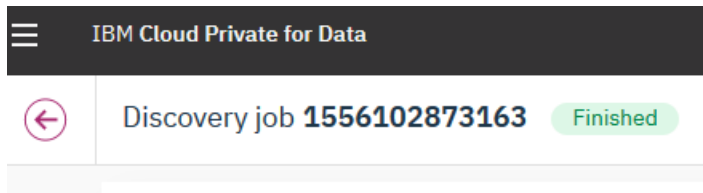
☐ Use data sampling

Set the maximum number of records that you want to include in your data set sample:

Discover

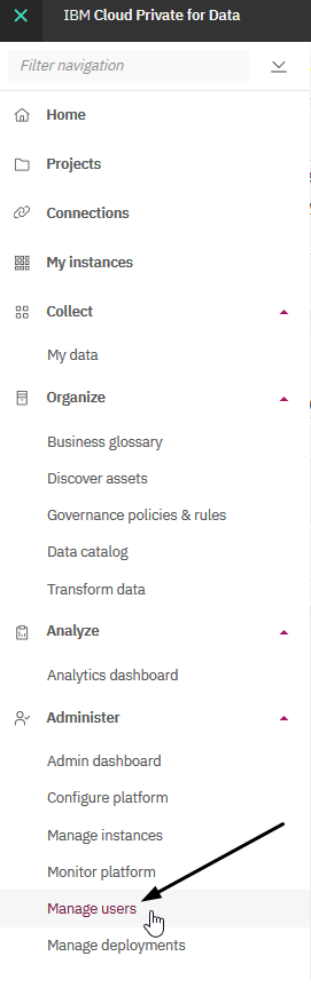
3. Add the bud connection
4. Choose the connection named 'bud' that you created previously.
5. Check 'Analyze data quality' under the 'Select the tasks that you want to run'.
6. Click on **Discover**

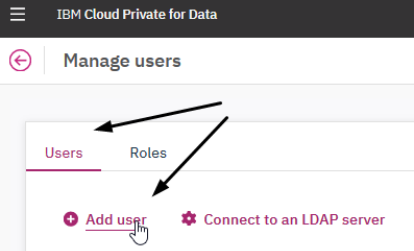
It may take few minutes to complete.



6. Add users

Create users with different roles.

| | |
|--|--|
|  | <p>From Administer option on the left pane, choose Manage users.</p> |
|--|--|

| | |
|---|--|
|  | <p>Switch tab to 'Users' and click on 'Add user'</p> |
|---|--|

Add user

Name
dst1

Username
dst1

Email
dst1@abc.com

User roles
☐ Administrator
☐ Business Analyst
☐ Data Engineer
☒ Data Scientist
☐ Data Steward

Cancel Add

Fill out Add User information for a data scientist

1. 'Name' as **dst1**
2. Username is **dst1**
3. Use a valid email address
4. Chose the user roles as Data Scientist

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


Roles Users

Add user Connect to an LDAP server Filter users

| NAME | STATUS | USERNAME | DATE ADDED | USER ID | ROLES |
|-------|----------|----------|---------------------|---------|------------------------|
| admin | Approved | admin | 12/31/1969, 6:00 PM | 999 | Administrator + 4 more |
| deng1 | Approved | deng1 | 08/06/2018, 8:46 AM | 1008 | Data Engineer |
| dst1 | Approved | dst1 | 08/11/2018, 4:31 PM | 1009 | Data Scientist |

View user Edit user Delete user

Before hand over user, change the password.

1. Access dst1 user setting by click on  icon
2. Choose 'Edit user'

Edit user

Name
dst1

Email
-

New password

Re-enter new password

User roles
☐ Administrator
☐ Business Analyst
☐ Data Engineer
☒ Data Scientist
☐ Data Steward

Account approval
☒ Approved user

Cancel Save

1. Type password as **dst1** in 'New password' and 'Re-enter new password' fields.
2. Click on Save

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 |

Log out from user **admin**

7. Implement Policies and Rules

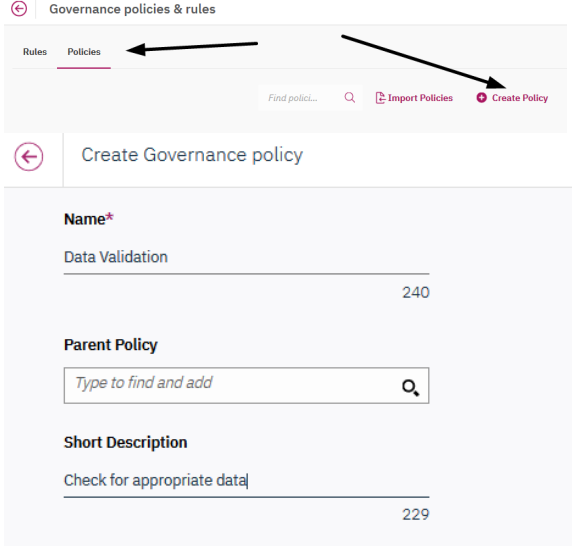
Create governance policies and rules for the entire organization to ensure clarity and compatibility among departments, projects, or products.

Sign in to the Cloud Pak for Data web console as user 'dstw1' and password is 'dstw1' that you created earlier.

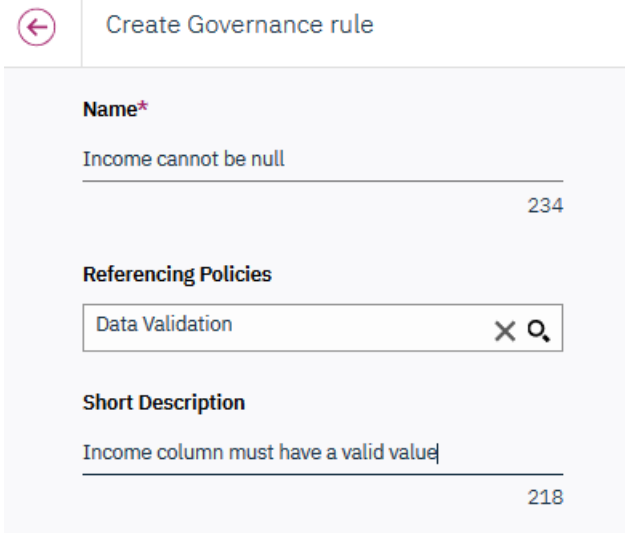
7.1. Create a policy

Choose **Organize** from the left pane, then select **Governance policies and rules**.

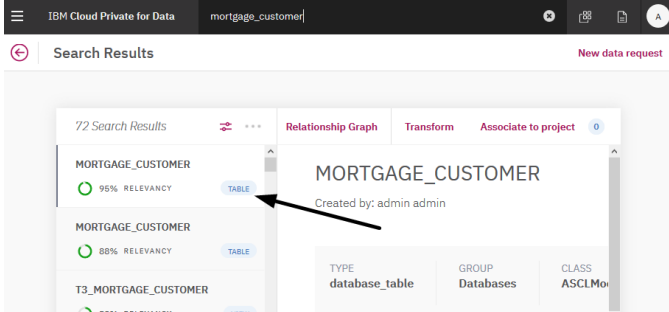
Select **Policies** tab and click on **Create Policy**

| | |
|---|---|
|  | <p>On the Create Information Governance Policy window create a polity with following information and click on Save:</p> <p>Name: Data Validation Short Description: Check for appropriate data</p> <p>It will take few minutes to appear under list of available policies.</p> |
|---|---|

7.2. Create a rule

| | |
|--|---|
| <p>Select Rules tab and click on Create Rule</p> | |
|  | <p>On the Create Information Governance Policy window create a rule with following information and click on Save:</p> <p>Name: Income cannot be null Referencing policies: Data Validation Short Description: Income column must have a valid value</p> <p>It will take few minutes to appear under list of available rules.</p> |

7.3. Add rule to metadata

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



On Database Table Details window choose **Database Columns** from left

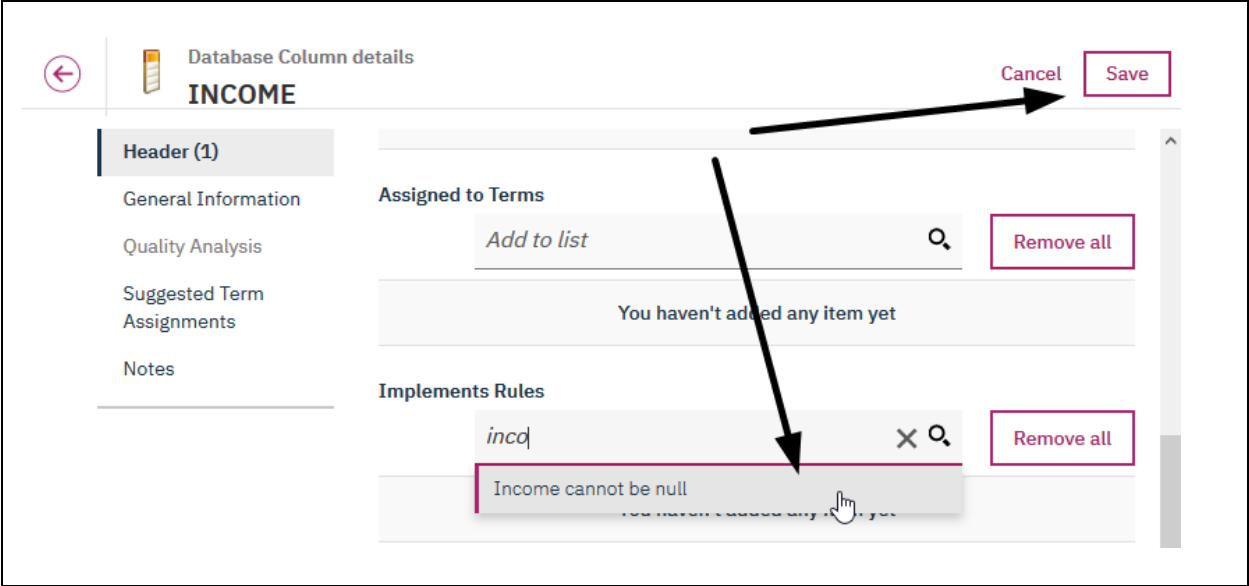
Select **INCOME** column

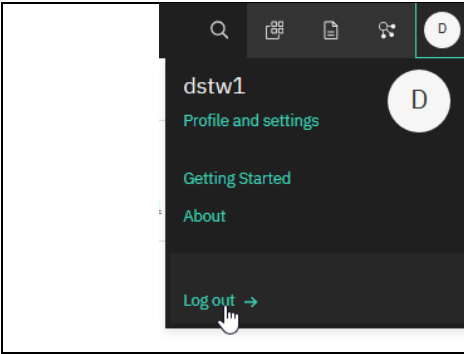
Next click on  icon (right top corner) and choose **Edit**

Scroll down to **Implement Rules** section

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

Click on **Save**

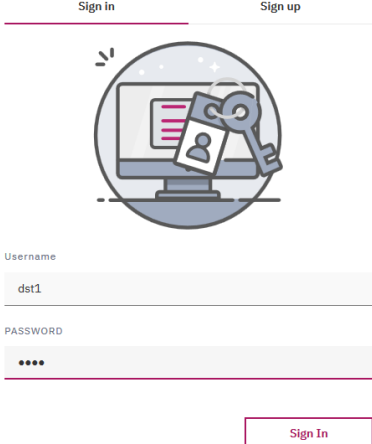




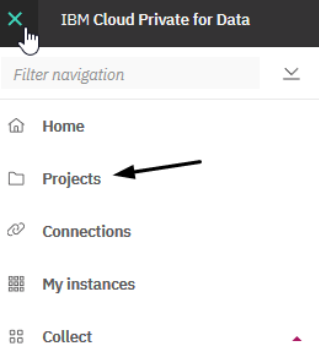
Log out from user 'dstw1'

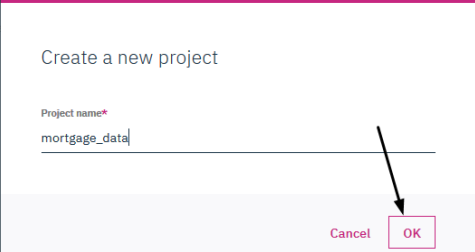
8. Access data as a Data Scientist

Explore the data require for build a model

| | |
|---|--|
|  | <p>Sign 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  New project icon</p> |
|--|--|

| | |
|---|--|
|  | <p>Provide a project name and click OK</p> <p>On the next 'Create project' window, click on 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, the 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 DataSearch

←Query Details: Glossary Categories and Terms

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**

8.3. 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)

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

Customer

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

[Industry Accelerators » Domains](#)

Terms

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.

8.4. Enterprise search

IBM Cloud Private for Data

Click on the enterprise search

Search

mortgage

IBM Cloud Private for Data

Search for 'mortgage' and hit enter

MORTGAGE_PROPERTY

MORTGAGE_PROPERTY

Database Table

[dbc:db2://9.30.116.168:50000](#)

[/MORTGAGE](#)

[db2 » DB2INST1](#)

★★★★★

0 Ratings

None

Quality score

Description

Select your rating:

★★★★★

New Comment:

Write a comment

Submit

All Comments (0)

Relationships

LOCATION Database Column

Context

+

Database Column (3)

Context

MORTGAGE_PROPERTY Database Table

Context

ID Database Column

Context

SALE_PRICE Database Column

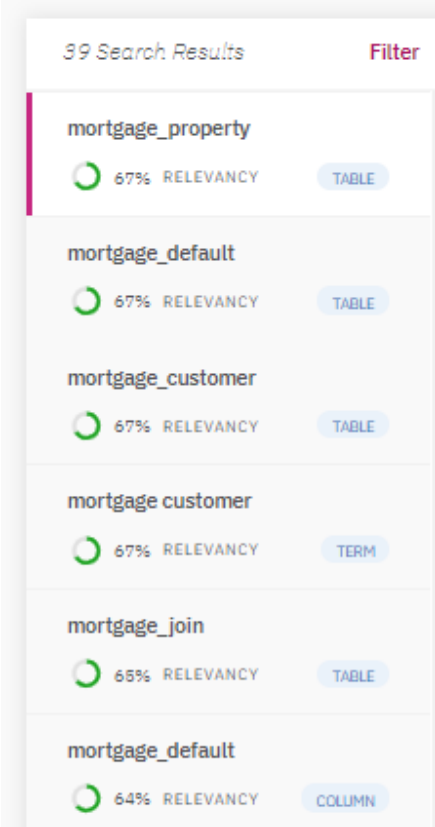
Context

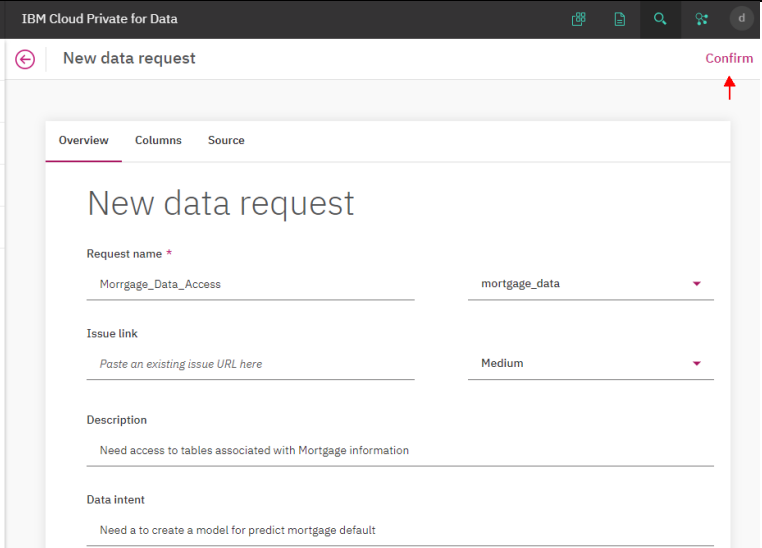
Choose the **mortgage_property** table and click on **Relationship Graph** to see details about the table.

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

Same way you can view other mortgage related tables.

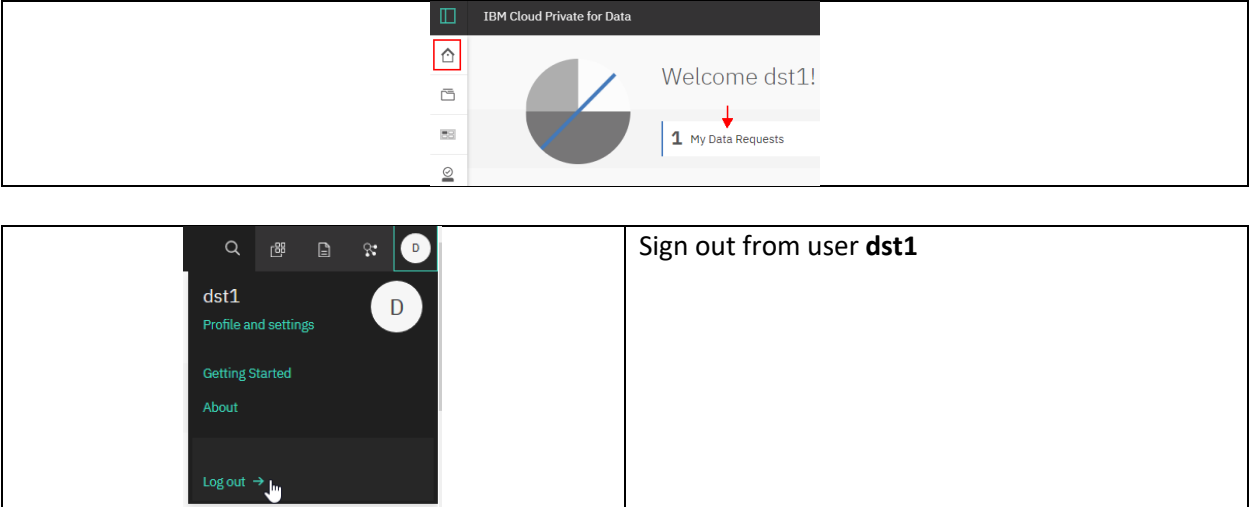
17

| | |
|--|---|
|  | <p>Go back to the enterprise Search Result</p> <p>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.</p> <p>Click on the New Data Request on top right corner for request access to mortgage related datasets.</p> |
|--|---|

| | |
|---|---|
|  | <p>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.</p> |
|---|---|

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.



9. Review data request

Sign in

Sign up




Username

deng1

PASSWORD

Sign In

Sign in to the Cloud Pak for Data web console as user 'deng1' and password is 'deng1' that you created earlier.



Welcome deng1!

2 Ungoverned Analytical A...

1 Data Requests


After sing in Click on [Go to your home page](#)


Check the **Data Request** tab on the home page.

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.

| 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 |

Q





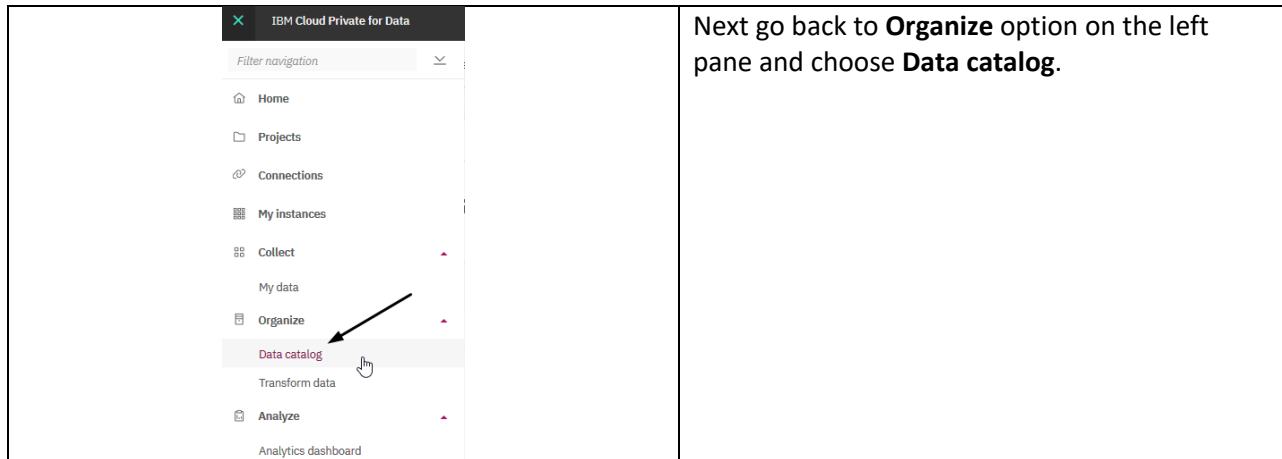
Action

Transform data

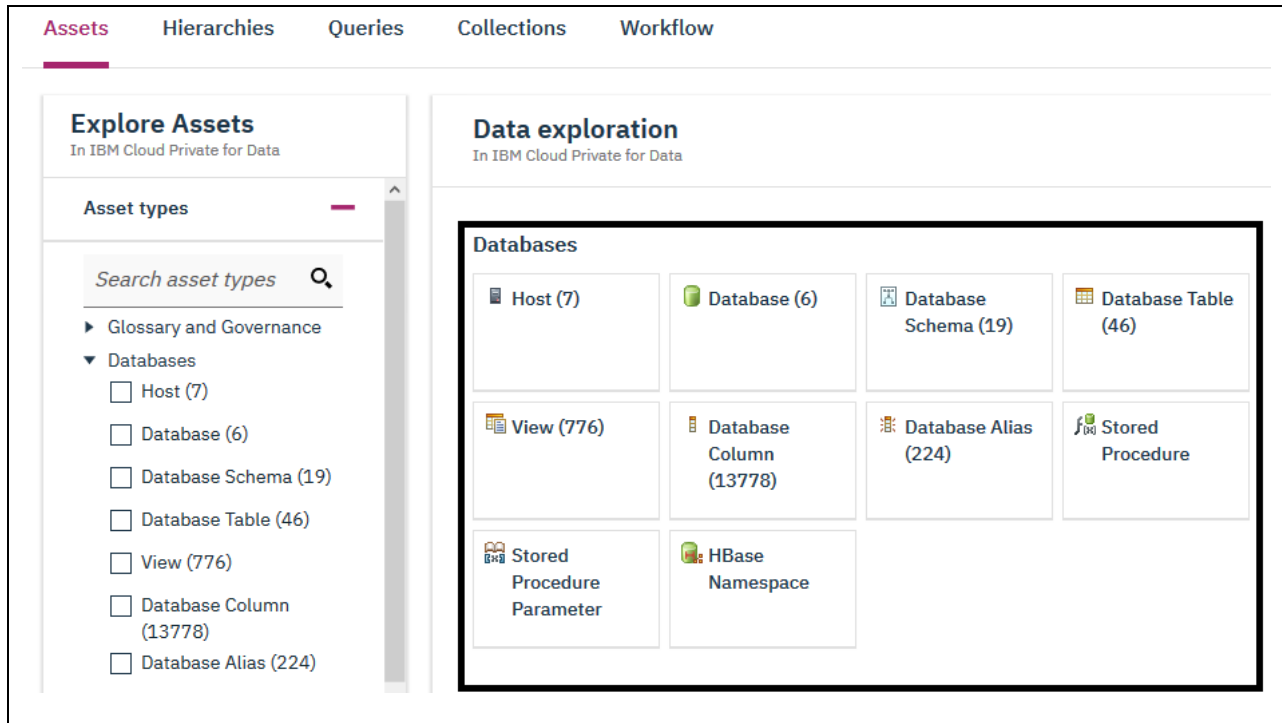
Assign to me

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.



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.

Filter results

Clear all filters

Asset types (1)

- Glossary and Governance
- ▼ Databases (1)
 - ☐ Host
 - ☐ Database
 - ☐ Database Schema
 - ☒ Database Table (46)
 - ☐ View
 - ☐ Database Column
 - ☐ Database Alias
 - ☐ Stored Procedure
 - ☐ Stored Procedure Parameter
 - ☐ HBase Namespace
- Data Files

All results

46 results

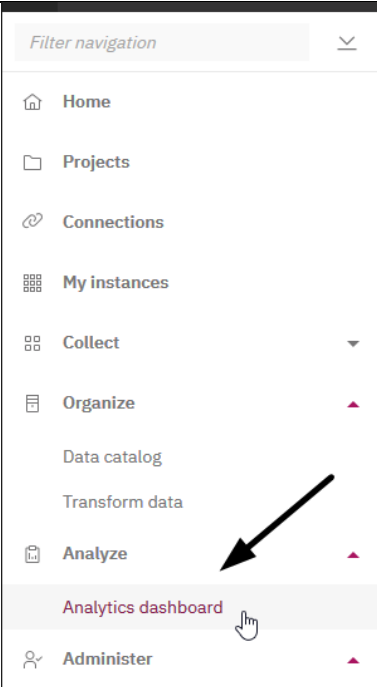
No items selected

| | | |
|--------------------------|--|--|
| <input type="checkbox"/> | /MORTGAGE » db2 » SYSTOOLS | InformationServerSystemUser on Jun 3, 2019, 6:52 PM |
| <input type="checkbox"/> | HMON_COLLECTION idbc:db2://10.208.125.125:50000 /MORTGAGE » db2 » SYSTOOLS | Modified by InformationServerSystemUser on Jun 4, 2019, 11:30 AM |
| <input type="checkbox"/> | MONGO_MORTGAGE_DEFAULT idbc:db2://dv-server.zen.svc.cluster.local:32051 /bissql » db2 » USER999 | Created by admin on Jun 3, 2019, 6:34 PM |
| <input type="checkbox"/> | MONGO_MORTGAGE_PROPERTY idbc:db2://dv-server.zen.svc.cluster.local:32051 /bissql » db2 » USER999 | Modified by InformationServerSystemUser on Jun 3, 2019, 6:34 PM |
| <input type="checkbox"/> | MORTGAGE_CUSTOMER idbc:db2://10.208.125.125:50000 /MORTGAGE » db2 » DB2INST1 | Created by admin on Jun 4, 2019, 11:28 AM |


11. Analyze Data

With the analytics dashboard, you can build sophisticated visualizations of your analytics results and communicate the insights you've discovered in your data on the dashboard. Then, share the dashboard with others. The analytics dashboard tool in Cloud Pak for Data provides a great way for a line-of-business user to begin investigating data for patterns and insights. The dashboard can then be handed off to a data scientist for deeper analysis and predictive modeling.

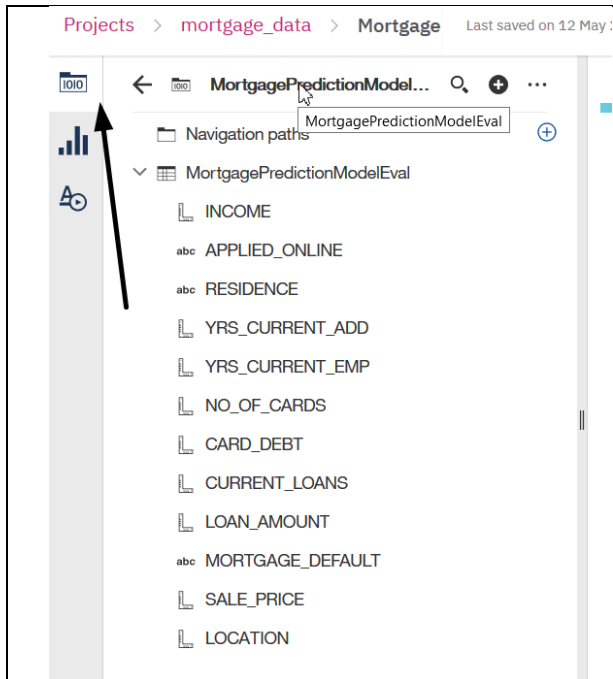
11.1. Navigate to Create Dashboard

| | |
|--|---|
|  | <p>Let's create a new dashboard. From the menu, click Analyze, then click Analytics dashboard. From the Dashboards page, click Create dashboard.</p> <p>Type a name for your dashboard (for example Mortgage), select a project (mortgage_data), and click Create.</p> <p>Select a template, for example the freeform tabbed template. A new empty dashboard opens.</p> |
|--|---|

11.2. Add a data source

| | |
|---|--|
|  | <p>In the newly created dashboard click on Add a source and then, on the right, click on Insert to dashboard under MortgagePredictionModelEval</p> |
|---|--|

11.3. Explore Visualizations and Widgets

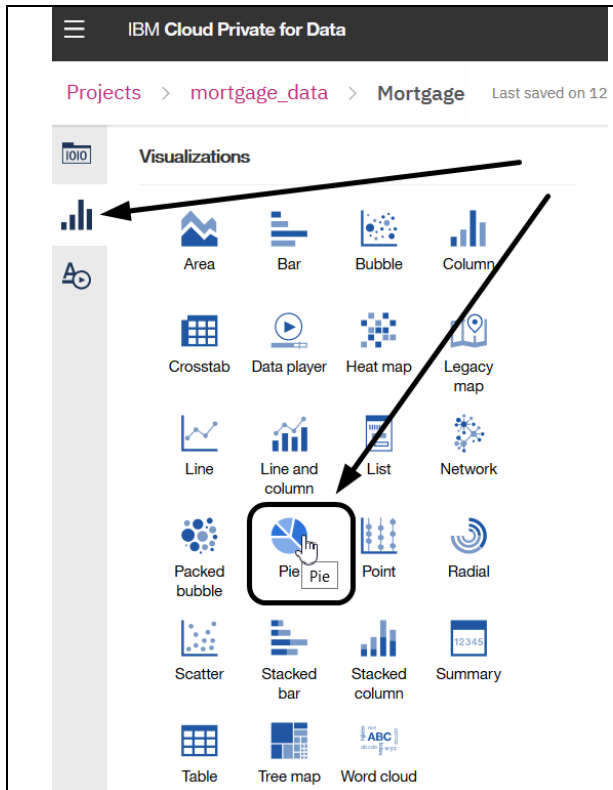


Once the data source has been added it can be viewed in the left pane.

Below Data Sources are Visualizations. There are many different types of graphs that you can use to visualize the data from the selected data source connection. Drill down into your source connection and select the data segments that you want to visualize.

Below Visualizations there are Widgets, such as text, media, web pages, images, and shapes to be added to the dashboard.

11.4. Add a graph



Select the pie graph.

11.5. Modify the graphs

The screenshot shows the Cloud Pak for Data interface. On the left, a navigation pane lists various data fields. In the center, a pie chart is displayed, showing the total value of properties for each location. The legend indicates the following values: 10,797,500 (Location 100), 1,940,500 (Location 120), 2,832,500 (Location 130), 3,035,500 (Location 101), and 4,342,500 (Location 110). On the right, a line and column chart is displayed, showing the total income of applicants (blue bars) and the value of mortgages granted (green line) for each location. The legend indicates the following values: INCOME (Sum) and LOAN_AMOUNT (Sum). Annotations show how to drag and drop data fields (LOCATION, SALE_PRICE, INCOME, LOAN_AMOUNT) and how to collapse the graph.

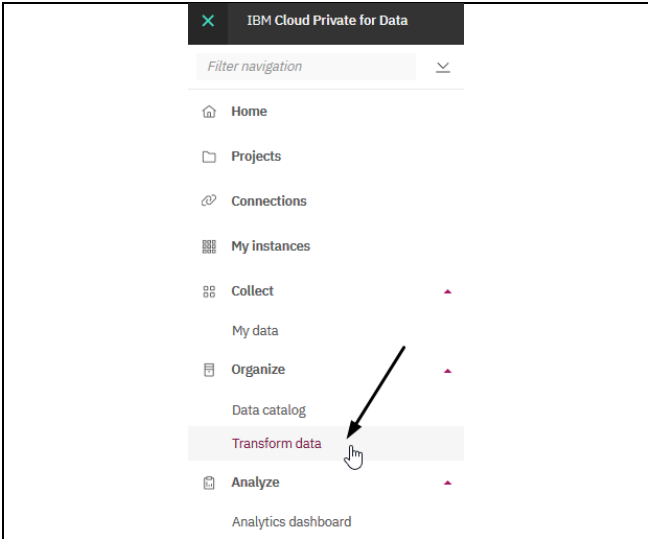
Drag and drop Location from the data source to Segments in the graph, and Sale Price for size. This shows total value of properties for each location. Click Collapse in the top right corner of the graph.

Add a 2nd graph from the Visualization menu: Line and column. Fill in the data as on the screenshot. This graph shows total income of applicants and the value of mortgages granted for each location.

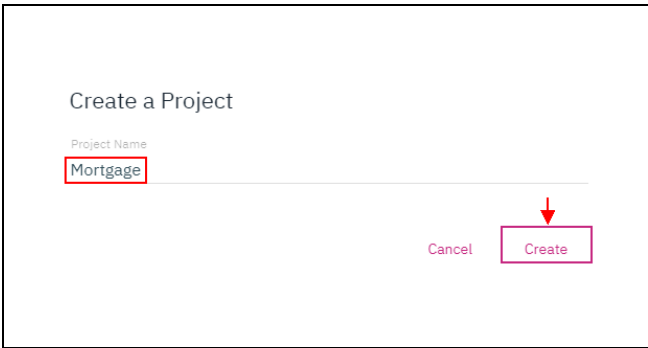

12. Transform Data

With Cloud Pak for Data, you can create, edit, load, and run transformation jobs. Cloud Pak for Data has features like built-in search, automatic metadata propagation, and simultaneous highlighting of all compilation errors, which make developers more productive.

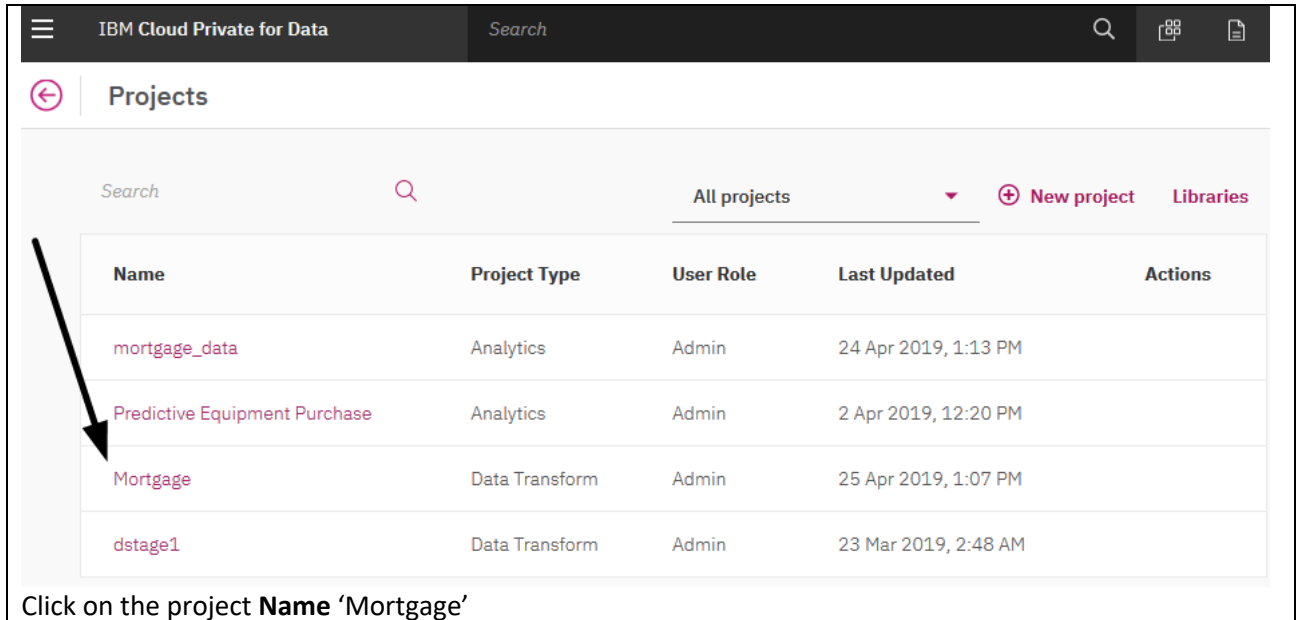
12.1. Navigate to transform data

| | |
|---|---|
|  | <p>Let's transform the data now. Go back to Organize option on the left pane and choose Transform data.</p> |
|---|---|

12.2. Create a Project

| | |
|---|--|
|  | <p>Next create a project by clicking on  Create icon on top left corner.</p> <p>On Create a Project window use the Project Name as 'Mortgage'.</p> <p>Click on Create</p> <p>It may take few minutes to complete.</p> |
|---|--|

Once the project is created it will be listed under the **Projects**.



IBM Cloud Private for Data

Search

Projects

Search

All projects

+ New project

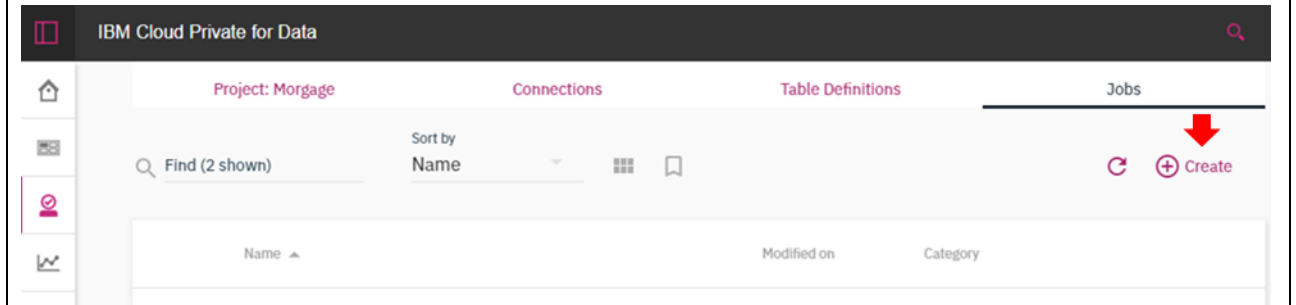
Libraries

| Name | Project Type | User Role | Last Updated | Actions |
|-------------------------------|----------------|-----------|----------------------|---------|
| mortgage_data | Analytics | Admin | 24 Apr 2019, 1:13 PM | |
| Predictive Equipment Purchase | Analytics | Admin | 2 Apr 2019, 12:20 PM | |
| Mortgage | Data Transform | Admin | 25 Apr 2019, 1:07 PM | |
| dstage1 | Data Transform | Admin | 23 Mar 2019, 2:48 AM | |

Click on the project **Name** 'Mortgage'

12.3. Create a job

Let's create a job by clicking on  **Create** icon on top left corner.



IBM Cloud Private for Data

Project: Mortgage

Connections

Table Definitions

Jobs

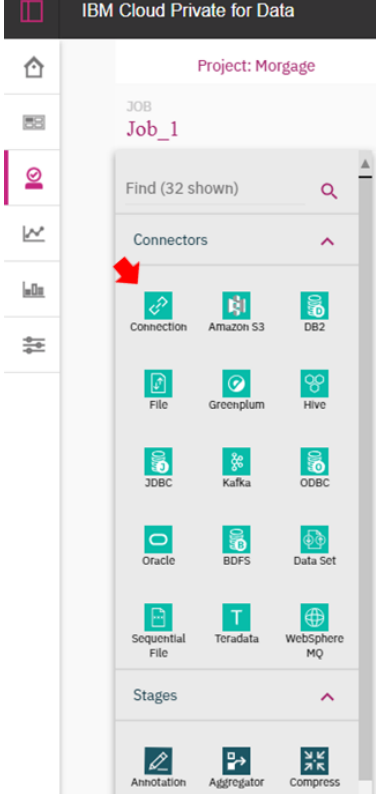

Find (2 shown)

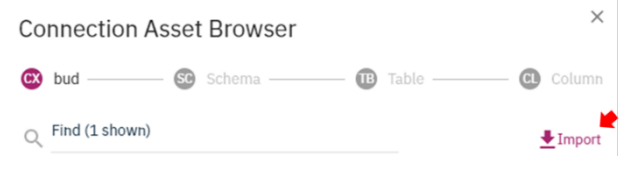
Sort by Name

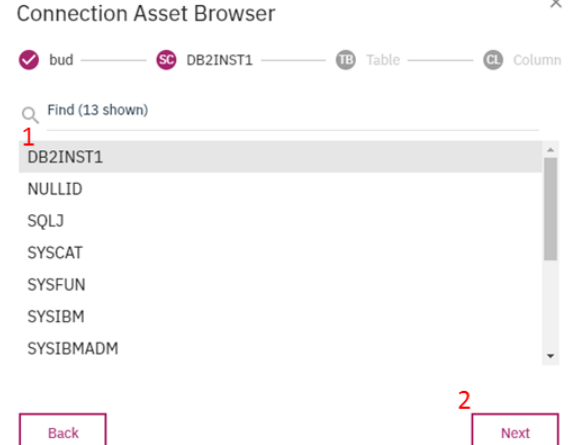
+ Create

| Name | Modified on | Category |
|------|-------------|----------|
| | | |

12.4. Add tables from asset browser

| | |
|--|---|
|  | <p>The create job operation will open a palette on the left.</p>  <p>Click on the Connection icon, drag it on the right pane and click once again. This will open the Connection Asset Browser window.</p> |
|--|---|

| | |
|---|--|
|  | <p>On the Connection Asset Browser window, Click on the Import to use the connection that you created earlier on step 4.2.</p> <p>If connection name already exists, just select it and click Next.</p> |
|---|--|

| | |
|---|---|
|  | <ol style="list-style-type: none">1. Chose the schema named 'DB2INST1'2. Click Next |
|---|---|

Connection Asset Browser

✓ bud

DB2INST1

TB MORTGAGE_...

CL Column

Find (4 shown)

1

MORTGAGE_CUSTOMER

MORTGAGE_DEFAULT

MORTGAGE_JOIN

MORTGAGE_PROPERTY

Back

2Next

1. Choose table named 'MORTGAGE_CUSTOMER'

2. Click **Next**

Connection Asset Browser

✓ bud

DB2INST1

✓ MORTGAGE_...

CL Column

✓

Name

Type

Length

✓

ID

INTEGER

0

✓

INCOME

INTEGER

0

✓

APPLIED_ONLI...

CHAR

1

✓

RESIDENCE

CHAR

1

✓

YRS_CURRENT...

SMALLINT

0

Back

Add to Job

Review the column name and datatype from table 'MORTGAGE_CUSTOMER' and click **Add to Job**.

MORTGAGE_CUS...

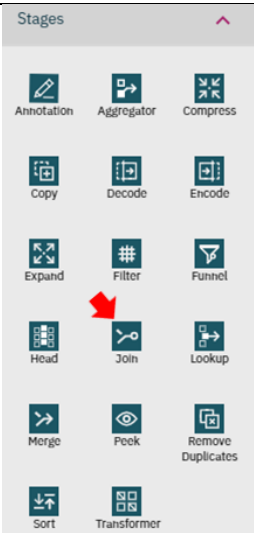

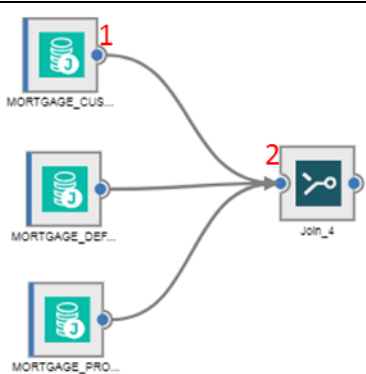
MORTGAGE_DEF...

MORTGAGE_PRO...

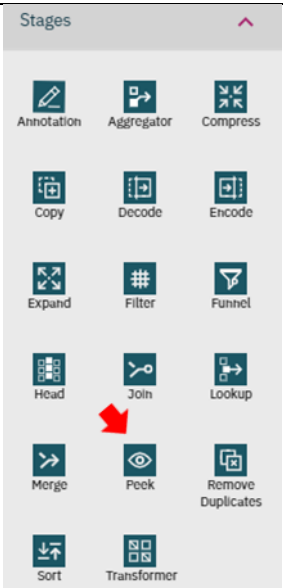

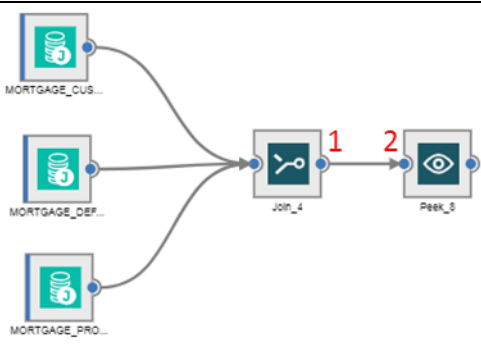
Repeat the above step in 5.4. for add tables 'MORTGAGE_DEFAULT' and 'MORTGAGE_PROPERTY' to the job.

Once all three tables added to the job, you should have three tiles on right pane.

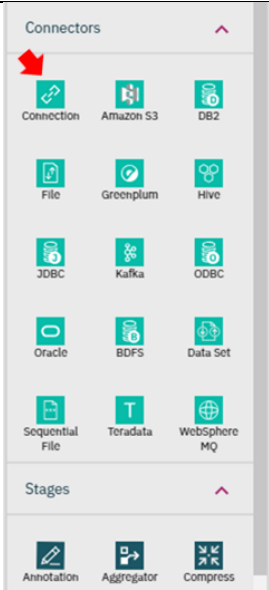
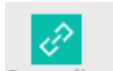
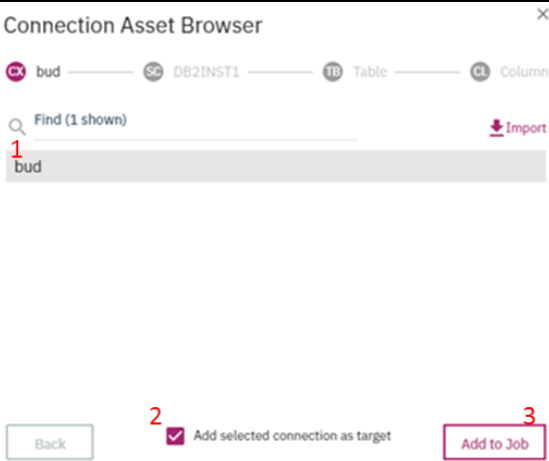
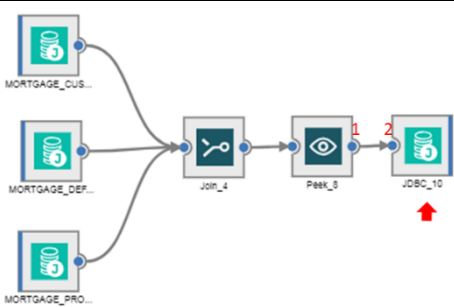
12.5. Join tables

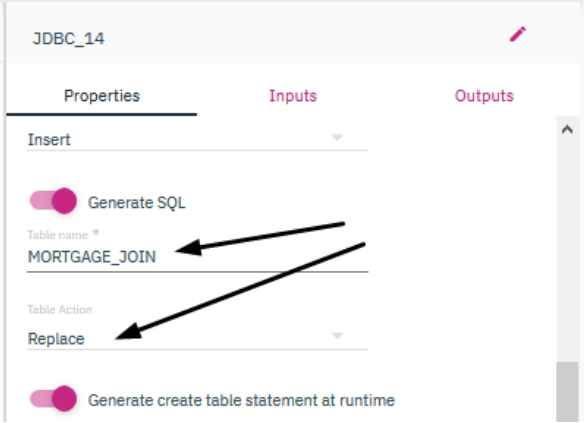
| | |
|--|--|
|  | <p>Now join the data.</p>  <p>Click on the Join icon from the palette on left, drag it on the right pane and click once again.</p> |
|  | <p>Connect one table tile at a time to the join tile.</p> <p>To connect tiles, click on two blue dots one at a time.</p> |

12.6. Preview output from join

| | |
|---|--|
| <p>Stages</p>  <p>Annotation Aggregator Compress</p> <p>Copy Decode Encode</p> <p>Expand Filter Funnel</p> <p>Head Join Lookup</p> <p>Merge Peek Remove Duplicates</p> <p>Sort Transformer</p> | <p>Add peek to look into join output.</p> <p>Click on the  icon from the palette on left, drag it on the right pane and click once again.</p> |
|  | <p>Connect join to the peek tile.</p> <p>To connect tiles, click on two blue dots one at a time.</p> |

12.7. Store output from join

| | |
|---|--|
|  | <p>Next save the persistent data from join to the target at Db2 database.</p>  <p>Click on the Connection icon, drag it on the right pane and click once again. This will open the Connection Asset Browser window.</p> |
|  | <ol style="list-style-type: none">1. On the Connection Asset Browser window, click on connection that you created earlier in step 4.2.2. Use check box Add selected connection as target3. Click Add to Job |
|  | <p>Join the target table tile with the peek</p> <p>To connect tiles, click on two blue dots one at a time.</p> <p>Once join completed, double-click on the new target table tile to make some adjustment.</p> |




On the Job properties pane

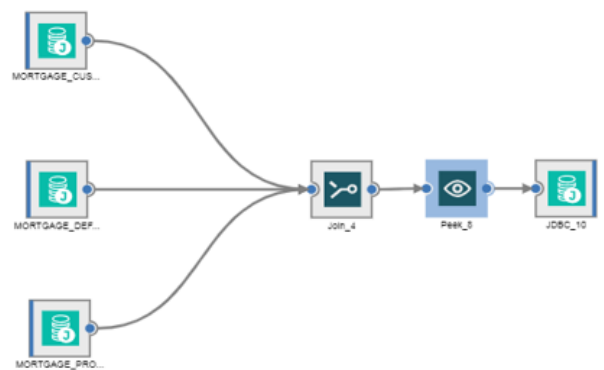
1. Use 'MORTGAGE_JOIN' as target **Table name**
2. From the **Table Action** dropdown menu chose 'Replace'
3. Click on **OK**

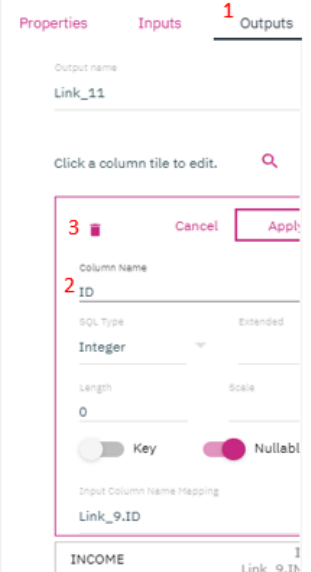
12.8. Transform output data

Let's go back to the Peek tile and double-click on it.

1. Choose the **Outputs** tab on the top right
2. Remove the column name **ID** by clicking on that column.
3. Click on the  icon.
4. Click OK

For machine learning to predict mortgage default, it will use all columns, except the **ID**.

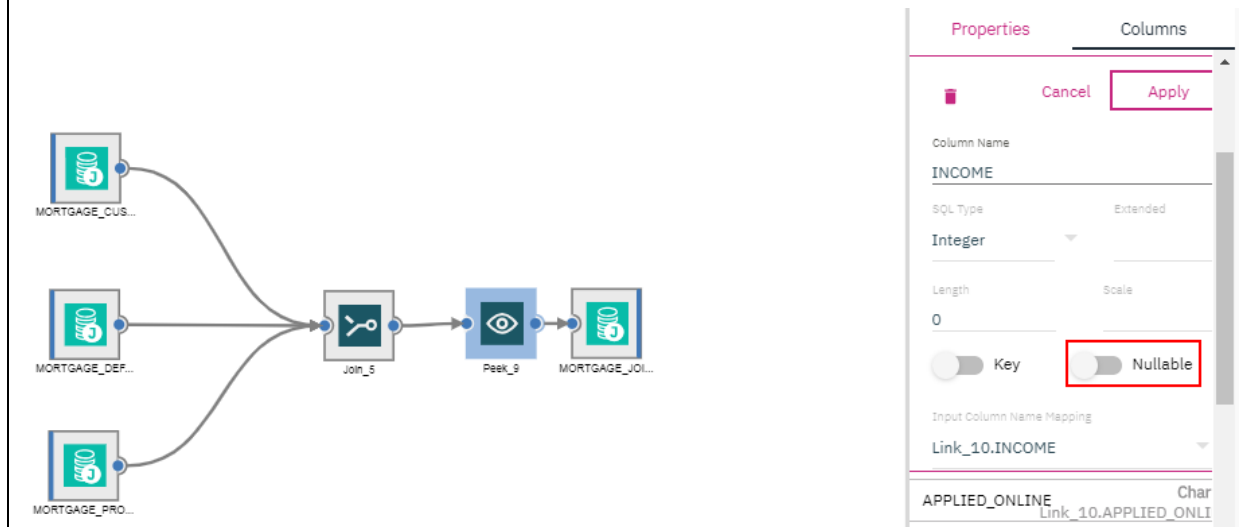




12.9. Apply governance rule

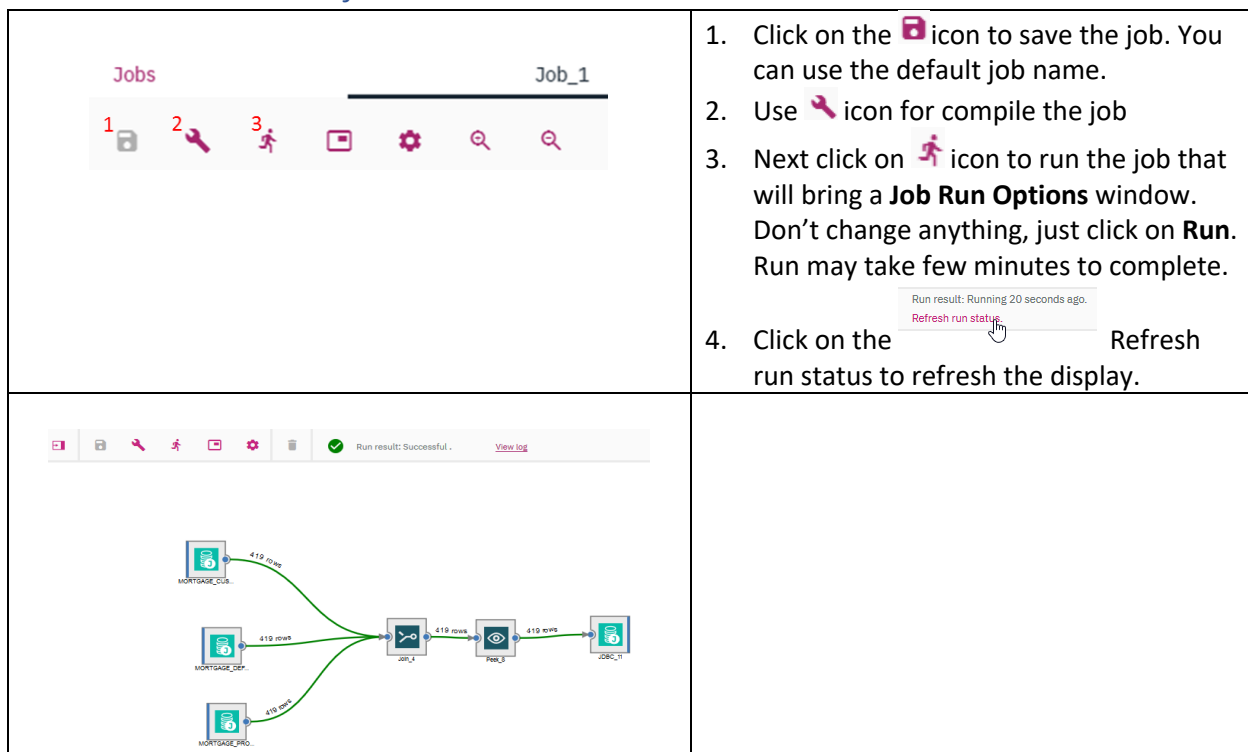
Go back to the **Peek** tile and double-click on it.

1. Choose the **Outputs** tab on the top right
2. Click on the INCOME column
3. Turn off the **Nullable** option, according to the “Income cannot be null” rule
4. Click **OK**







The screenshot shows a job flow with three input tiles (MORTGAGE_CUS..., MORTGAGE_DEF..., MORTGAGE_PRO...) feeding into a Join tile, which then feeds into a Peek tile, and finally into an output tile (MORTGAGE_JOI...). The Properties panel for the Peek tile is open, showing the INCOME column. The Nullable option is highlighted with a red box and is currently turned off.

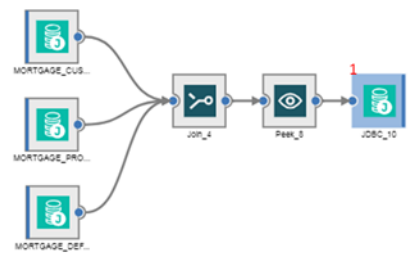
12.10. Execute job



The screenshot shows the Jobs panel with a list of jobs. The Job_1 job is selected. The job flow diagram is shown below the list, with data flow from the input tiles to the Join tile, then to the Peek tile, and finally to the output tile. The job flow is labeled with '419 rows' at each stage.

1. Click on the  icon to save the job. You can use the default job name.
2. Use  icon for compile the job
3. Next click on  icon to run the job that will bring a **Job Run Options** window. Don't change anything, just click on **Run**. Run may take few minutes to complete.
4. Click on the  Refresh run status to refresh the display.

12.11. Preview output data



Properties

CONNECTION

URL *
jdbc:db2://9.30.160.15:50000/demodb

Username
db2inst1

Password

Attributes

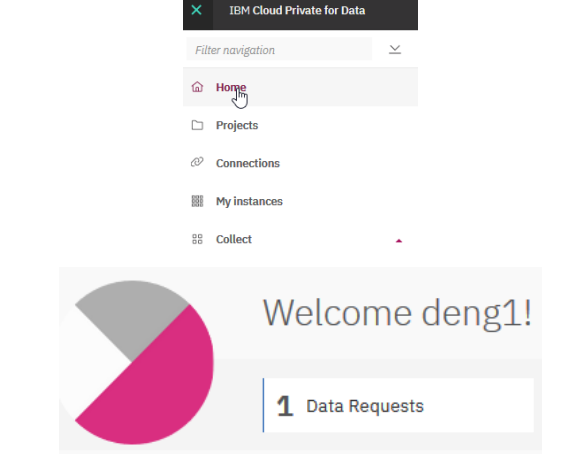
Cancel OK


Let's take a quick look into the final data.

1. Click on the new target table tile
2. Click on **View Data**

The View Data will pop up a window with all the data. Once you done with review the data, close the window.

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. Use the IP address of master-1 node in case of JDBC URL.

New data request

OverviewColumnsSource

Source

Data source name

mortgage_joinDB2

Username

db2inst1

Password

JDBC URL

169.45.83.218

+ Add new dataset

| | Remote data set name | Description | Schema | Table |
|---|----------------------|-------------|----------|---------------|
| 1 | mortgage_join | | db2inst1 | mortgage_join |

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

| NAME | ID | STATUS | REQUESTED BY | ACCEPTED BY | LAST UPDATED | ACTIONS |
|----------------------|----|-----------|--------------|-------------|----------------------|--|
| mortgagedata1 | 7 | Delivered | dst1 | deng1 | 6 Aug 2018, 12:55 PM | |
| Mortgage_Data_Access | 9 | Accepted | dst1 | deng1 | 15 Aug 2018, 1:17 AM | <div><div></div><div>Deliver</div><div>Decline</div><div>Close</div></div> |

IBM Cloud Private for Data

HomeData RequestsData CatalogData Governance

Welcome deng1

2 Data Requests

Signed in as: deng1

Getting Started

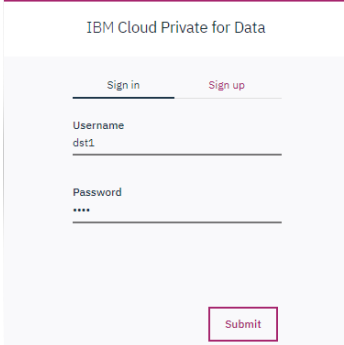
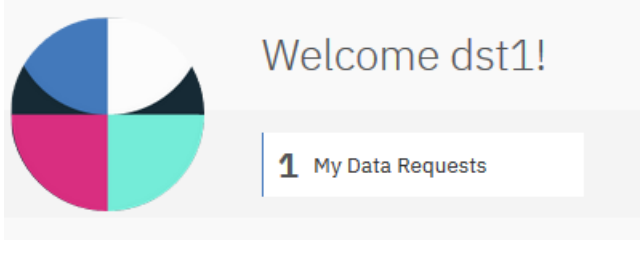

Settings

Sign Out

Sign out from user **deng1**

13. 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> |
|  | <p>At this point data engineer deliver the data set for the data you requested. You can go to the home page by clicking on  icon from left pane</p> |

13.1. Navigate to analytics project

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

13.2. Create a model

| | |
|---|--|
|  | Next, choose the Launch Terminal with Python from top right corner. |
|---|--|

Copy (scp using root) a predefined Jupyter notebook from `~/ICP4XTutorial/assets/mortgage-001/MortgageNotebook.jupyter-py36.ipynb` on master-1 node; to `./jupyter/` directory under current project. Jupyter notebook was downloaded earlier from the Git repository.

(This step is needed just for this tutorial to create a model easily. In real life a data scientist will not have access of root on master-1 node.)

| |
|--|
| <p>Projects > mortgage_data > Terminal</p> <pre>1002@jupyter-py36-server-1544906152778-1002-846b99f76c-nbnh6:~/DSX_Projects/mortgage_data\$ scp root@ :/root/ICP4XTutorial/assets/mortgage-001/MortgageNotebook.jupyter-py36.ipynb ./jupyter/ Warning: Permanently added ' ' (ECDSA) to the list of known hosts. Are you sure you want to continue connecting (yes/no)? yes Warning: Permanently added ' ' (ECDSA) to the list of known hosts. root@ : password: MortgageNotebook.jupyter-py36.ipynb 100% 83KB 8.9MB/s 00:00 1002@jupyter-py36-server-1544906152778-1002-846b99f76c-nbnh6:~/DSX_Projects/mortgage_data\$</pre> |
|--|

| | |
|--|--|
| <p>Projects > mortgage_data > Terminal</p> | Go back to project mortgage_data by clicking on the project name from top left. |
|--|--|

| | |
|---|---|
| <p>mortgage_data</p> <p>Assets 1 Data Sources 0 Jobs 0 Environments 4 Collaborators 1</p> <p>Recent</p> <ul style="list-style-type: none"> Data sets 0 Notebooks 1 Scripts 0 Models 0 Model groups 0 Analytics dashboards 0 Data Refinery flows 0 Modeler flows 0 <p>MortgageNotebook</p> <p>Jupyter notebook • 27 Mar 2019, 11:30 AM</p> | Open the predefined notebook called MortgageNotebook . |
|---|---|

13.3. 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.

First go the Step 1 and update the **dsn_hostname** value with the IP address of master node-1.

mortgage_data > Notebooks > MortgageNotebook

File Edit View Insert Cell Kernel Widgets Help

Run

Predicting Mortgage Default in Loan Marketplace

In this notebook you will learn how to build a predictive model with Spark machine learning API (SparkML) and deploy it for scoring in Machine Learning

This notebook walks you through these steps:

- Build a model with SparkML API
- Save the model in the ML repository
- Create a Deployment in ML (via UI)
- Test the model (via UI)

In [7]: *# Check Python version. This notebook is implemented for Python 3.5.x. Not all cells may work in other versions of Python*

```
import platform
print(platform.python_version())
```

3.6.6

In [8]:

```
import ibm_db
import pandas
import ibm_db_dbi
```

Step 1: Load data --- Update the **dsn_hostname** value with your Web Console IP

In [9]: *#Enter the values for you database connection*

```
dsn_driver = "IBM DB2 ODBC DRIVER"
dsn_database = "MORTGAGE"
dsn_hostname = <Hostname/IP>
dsn_port = "50000"
dsn_protocol = "TCPIP"
dsn_uid = "db2inst1"
dsn_pwd = "password"
```

e.g. "MORTGAGE"
e.g. "Use the same IP as Web Console"
e.g. "50000"
i.e. "TCPIP"
e.g. "dash104434"
e.g. "7dBZ3jWt9xN6\$o0JiX!m"

Run through it so that you generate a model. The easiest way to do this is to open the notebook, scroll down to Step 6, click on it, then in the menu select Cell -> Run all above.

The screenshot shows the Jupyter Notebook interface for a project named 'mortgage_data'. The notebook is titled 'MortgageNotebook'. The 'Cell' menu is open, and the 'Run All Above' option is highlighted. The notebook content includes a 'Step 6: Test' section with instructions: '1. Save the notebook', '2. Under Models', and '3. Click the Test link'. The 'Input' section shows fields for 'INCOME *' (44202), 'APPLIED_ONLINE *' (Y), 'RESIDENCE *' (O), 'YRS_CURRENT_ADD *' (8), 'YRS_CURRENT_EMP *' (0), and 'NO_OF_CARDS *'. The 'Result' section displays a pie chart with two segments, one labeled 'Y' and the other 'N'.

13.4. Test the model

Save the notebook and switch to the Models tab of the project (hint: right click the project name link, **mortgage_data**, at the top, and open with another tab in your browser).

The screenshot shows the Jupyter Notebook interface for the 'mortgage_data' project. The notebook is titled 'MortgageNotebook'. The 'Step 6: Test Saved Model with Test UI' section is visible, containing instructions: '1. Save the notebook and switch to the Models tab of the project (hint: right click the project name link, mortgage_data, at the top, and open with another tab in your browser).', '2. Under Models, find and click into your saved model.', and '3. Click the Test link to Real-time score the model. You can use the data for testing that already loaded:'. The 'Input' section shows fields for 'INCOME *' (44202), 'APPLIED_ONLINE *' (Y), 'RESIDENCE *' (O), 'YRS_CURRENT_ADD *' (8), 'YRS_CURRENT_EMP *' (0), and 'NO_OF_CARDS *'. The 'Result' section displays a pie chart with two segments, one labeled 'Y' and the other 'N'.

mortgage_data

Assets 3

Data Sources 0

Jobs 0

Environments 4

Collaborators

Recent

Data sets 1

Notebooks 1

Scripts 0

Models 1

Model groups 0

Analytics dashboards 0

Data Refinery flows 0

Modeler flows 0

MortgageNotebook

Jupyter notebook • 27 Mar 2019, 11:3...

Mortgage_Prediction_Model

Spark model • 27 Mar 2019, 11:34 AM

Chose the **Mortgage_Prediction_Model**

Mortgage_Prediction_Model v1

LAST MODIFIED

27 Mar 2019, 11:34 AM

TYPE

Spark

Overview

Real-time score

Batch score

Evaluate

Accuracy

62%

Accuracy history

100%

80%

60%

40%

20%

0%

Click on the **Real-time score** to test the model.

Once your model is open, check the the mortgage default predeiction based on the sample data in the **Input** section.

mortgage_data > Models > Mortgage_Prediction_Model

Mortgage_Prediction_Model v1

LAST MODIFIED

15 Dec 2018, 4:15 PM

TYPE

Spark

ALGORITHM

PipelineModel (Classification)

ENGINE

Python 3.6

Overview

Real-time score

Batch score

Evaluate

Input

Installed Packages

Result

INCOME *

44202

APPLIED_ONLINE *

Y

RESIDENCE *

0

YRS_CURRENT_ADO *

0

YRS_CURRENT_EMP *

0

Result

0%

41

Cloud Pak for Data – Tutorial

CARD_DEBIT *

748

CURRENT_LOANS *

0

LOAN_AMOUNT *

10455

SALE_PRICE *

170000

LOCATION *

100

Clear

Submit

If you want you can change some values in the Input section.
Then clien on **Submit**.

According on input values, model will predict the possibilities of mortgage default and produce a pie chart.

