



Governed MLOps Workshop
**Setup Users,
Catalogs, Data**

Document version: June 2023

DISCLAIMER

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.

The development, release, and timing of any future features or functionality described for our products remains at our sole discretion I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results like those stated here.

Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. **This document is distributed "as is" without any warranty, either express or implied. In no event, shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity.** IBM products and services are warranted per the terms and conditions of the agreements under which they are provided.

IBM products are manufactured from new parts or new and used parts. In some cases, a product may not be new and may have been previously installed. Regardless, our warranty terms apply."

Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.

Performance data contained herein was generally obtained in controlled, isolated environments. Customer examples are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary. References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.

Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.

It is the customer's responsibility to insure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer follows any law.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products about this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. **IBM expressly disclaims all warranties, expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a purpose.**

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.

IBM, the IBM logo, and ibm.com are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at: www.ibm.com/legal/copytrade.shtml.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

OpenShift is a trademark of Red Hat, Inc.

UNIX is a registered trademark of The Open Group in the United States and other countries.

© 2023 International Business Machines Corporation. No part of this document may be reproduced or transmitted in any form without written permission from IBM.

Table of Contents

Introduction	4
Create Platform Connection	4
Create the required users in Cloud Pak for Data	8
Create users using CPD's REST API.....	12
Create Catalog for Churn Data Assets	16
Default Catalog Access	19
Platform Connections Access.....	21
Data Virtualization Access	23
Data Upload.....	24
Refresh OpenPages Db2 Certificates	30
Summary	32
Appendix	33
Data Upload (Manual).....	33

Introduction

This document contains the details for initially configuring your Cloud Pak for Data environment for purposes of this workshop. These instructions assume you already have access to a Cloud Pak for Data environment as admin user.

You will execute the following tasks:

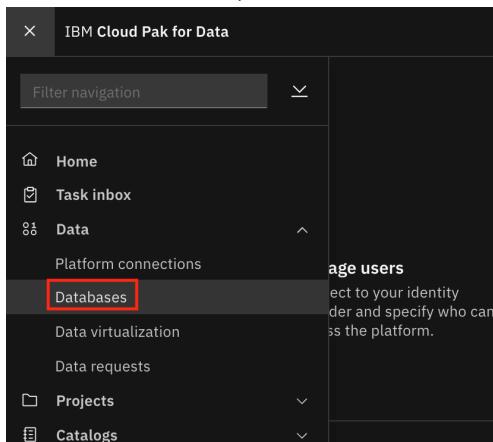
1. Create Platform connection
2. Create the required users in Cloud Pak for Data
3. Create Catalog for Churn Data Assets
4. Default Catalog Access
5. Platform connections access
6. Data Virtualization Access
7. Data Upload

Create Platform Connection

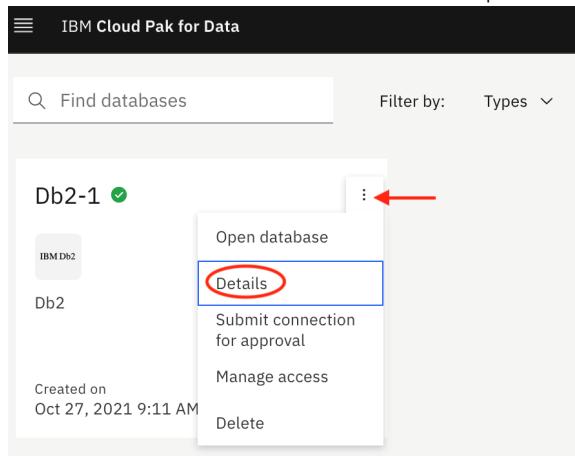
To start with, you will step through the creation of a platform connection to Db2 database running on the same Cloud Pak for Data cluster. This Db2 will include relevant data assets that will be consumed throughout this workshop.

To setup that platform connection to Db2, you need to find the deployment ID for your Db2 instance, by executing the following steps:

- 1- Navigate to your Cloud Pak for Data url and log in as admin user.
- 2- Navigate to databases by clicking on the Navigation menu (top left hamburger icon) and selecting Data → Databases (annotated with red rectangle).



- 3- Click the open and close list of options menu (annotated with red arrow) next to your Db2-1 database and select Details from the drop down (annotated with red oval)



- 4- On the database details page, find the Deployment id. In the example below, the Deployment id is db2oltp-1681603303498757 (annotated with red oval). Your Db2 instance may show different values.

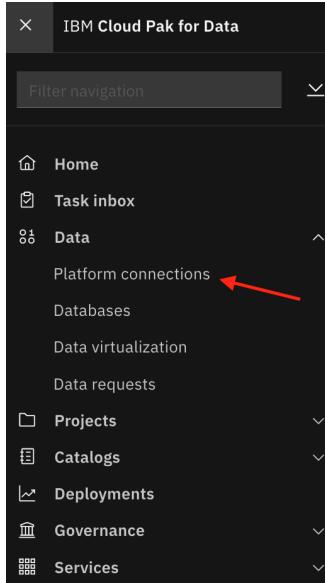
About this database		Storage	
Database name	BLUDB	Storage class (System storage)	managed-nfs-storage
Database type	db2oltp	Size (System storage)	100 GiB
Database software version	11.5.8.0-cn1-x86_64	Storage class (User storage)	managed-nfs-storage
Processor	x86-64	Size (User storage)	100 GiB
Deployment id	db2oltp-1681603303498757	Storage class (Backup storage)	managed-nfs-storage
Created on	Apr 15, 2023 8:01 PM	Size (Backup storage)	100 GiB
Status	Available	Size (Archive logs storage)	managed-nfs-storage
Nodes		Size (Archive logs storage)	100 GiB
HOSTNAME	CPU	MEMORY	
worker7	2.1 cores	5.5 GiB	Storage class (Transaction logs storage)

- 5- Capture the Db2 deployment ID as you will need it later. For the example referenced here, the values would be:

```
"Deployment ID": db2oltp-1681603303498757
```

- 6- Next, you will create platform connections to access the data assets referenced earlier which are needed for the churn prediction project. Platform connections are available to be consumed by all services of the platform provided the services support the data source type.

- 7- Navigate to Platform connections by clicking on the Navigation menu (top left hamburger icon) and selecting Data → Platform connections (annotated with red arrow).



- 8- On the Platform connections page, click New connection + button (annotated with red arrow).

The screenshot shows the 'Platform connections' page. At the top right, there is a blue button labeled 'New connection' with a '+' sign, which is annotated with a red arrow. The page also includes a search bar, a filter dropdown, and a table header with columns: Name, Type, Created by, Modified by, and Last updated.

- 9- Select IBM Db2 connection type (annotated with red rectangle) and click Select (annotated with red arrow).

The screenshot shows the 'New connection' dialog. On the left, under 'Provider', 'IBM' is selected. In the center, 'Find connection types' is used to search for 'IBM Db2', which is highlighted with a red rectangle. On the right, 'Selected connection type' shows 'IBM Db2' selected. Below it, 'Details' provide information about IBM Db2. At the bottom right, there is a large blue 'Select' button, which is annotated with a red arrow.

- 10- Provide a Name <Db2 Customer Personal Information> and an optional Description for the connection and provide the required connection details to access the Db2 instance which were obtained earlier. The username and password credentials should be the credentials for the user who has access to that Db2 instance; in this case, it is the admin.

```
"Database": BLUDB,
"Hostname or IP address": c-<YOUR-DEPLOYMENT-ID>-db2u-engn-svc
"port": 50000
"username": admin
"password": your_admin_password
```

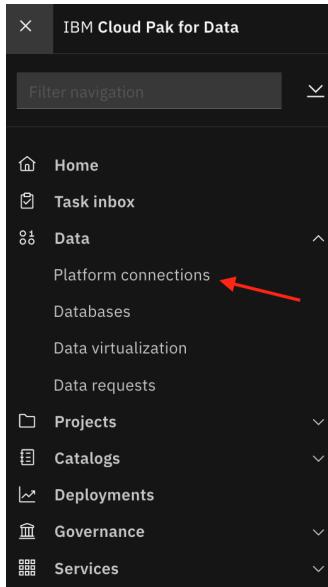
The Hostname or IP address is the Deployment ID you captured earlier, with the “c-“ prefix and “-db2u-engn-svc” suffix.

Make sure the box next to Port is SSL-enabled on the bottom left of the page is unchecked. Then click Test connection (annotated with red oval). You should see a message in green on top of the page confirming that “The test was successful” (annotated with red rectangle).

Once you see the successful test message, click Create (bottom right of the page).

The screenshot shows the 'Create connection: IBM Db2' form. The 'Connection details' tab is active. A green box at the top displays the message 'The test was successful.' Below it, the 'Test connection' button is circled in red. The 'Username' field ('admin') and 'Password' field (redacted) are highlighted with yellow boxes. Other fields like 'Hostname or IP address' and 'Port' are also visible.

- 11- Navigate back to the Platform connections page by clicking on the Navigation menu (top left hamburger icon) and selecting Data → Platform connections (annotated with red arrow).



12- You should have two platform connections, the Db2 Customer Personal Information connection you just created which connects to the on-prem Db2 database and the db2cloud customer transaction data connection which was pre-created for you and that connects to the managed Db2 database on IBM Cloud.

The screenshot shows the 'Platform connections' page in IBM Cloud Pak for Data. The 'Connections' tab is selected. The page displays a list of connected data sources:

Name	Type	Created by	Modified by	Last updated
Db2 Customer Personal Information	IBM Db2	Dataengineer	dataengineer	Jan 28, 2023
db2cloud customer transaction	IBM Db2 on Cloud	Admin	admin	Jan 27, 2023
Data Virtualization	IBM Watson Query	System	System	Jan 26, 2023

Create the required users in Cloud Pak for Data

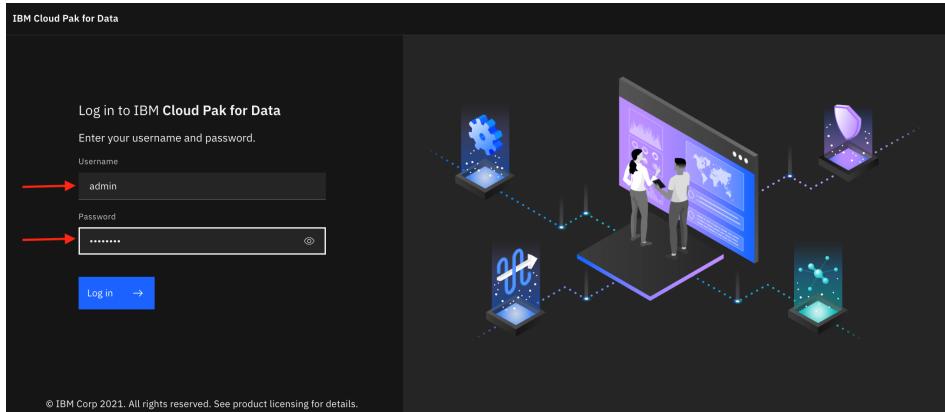
As an enterprise Data and AI platform, Cloud Pak for Data supports separation of roles and responsibilities. In this workshop, you will execute different tasks while assuming different user personas with different roles and permissions. Specifically, you will execute tasks as one of the following users:

- datasteward (Data Steward role and permissions)
- dataengineer (Data Engineer role and permissions)
- dqanalyst (Data Quality Analyst role and permissions)

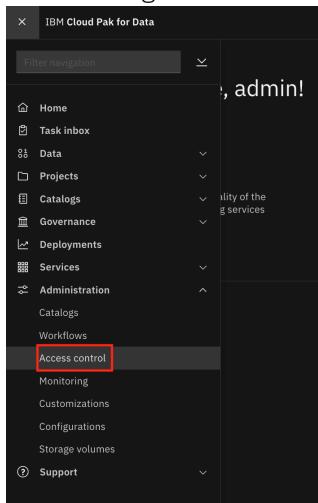
- datascientist (Data Scientist role and permissions)
- dslead (Data Scientist and Data Engineer role and permissions)
- deploy (Data Scientist role and permissions)

The Cloud Pak for Data environment initially only has the admin user. You will log in as the admin user and create the different users and associate the corresponding roles.

- 1- Navigate your favorite browser to the Cloud Pak for Data url
- 2- Log into Cloud Pak for Data as admin user.
username: admin
password: CP4DDataFabric



- 3- Navigate to Access control by clicking on the navigation menu (top left hamburger menu) and then clicking Administration → Access control (annotated with red rectangle).



- 4- Click on Add user + (annotated with red rectangle).

Governed MLOps Workshop – Setup Users, Catalogs, Data

The screenshot shows the 'Access control' section of the IBM Cloud Pak for Data interface. The 'Users' tab is selected. At the top right of the user table, there is a blue 'Add user' button with a dropdown arrow, which is highlighted with a red box.

- 5- Create the datasteward user. Provide a full name (optional), username, email (optional), and password, then provide the password again to confirm it. Click Next.
 Please note the password for each user you create as you will need that to log into Cloud Pak for Data as that user and execute the associated tasks.

The screenshot shows the 'Add user' profile information step. The 'Profile information' section is active. The 'Full name (optional)' field contains 'datasteward'. The 'Username' field contains 'datasteward'. The 'Email (optional)' field contains 'datasteward@cpd.com'. The 'Password' and 'Confirm password' fields both contain '*****'. At the bottom right, the 'Next' button is highlighted with a blue box.

- 6- Click Next to accept the default of Assign roles directly.

The screenshot shows the 'Add users' platform access step. The 'Platform access' section is active. It shows two options: 'Assign roles directly' (selected) and 'Add to user group'. Both options have small icons next to them. At the bottom right, the 'Next' button is highlighted with a blue box.

- 7- On the next screen, choose the Data Steward role and click Next.

Governed MLOps Workshop – Setup Users, Catalogs, Data

Add user
Authorize user access to the platform.

Profile information
 Platform access
 Roles
 Summary

Roles
Assign roles to this user or [create a new role](#).

Find roles

	Description	Modified on
<input type="checkbox"/> Administrator	Administrator role	Sep 2, 2021 8:24 AM
<input type="checkbox"/> Business Analyst		
<input type="checkbox"/> Data Engineer		
<input type="checkbox"/> Data Quality Analyst		
<input type="checkbox"/> Data Scientist		
<input checked="" type="checkbox"/> Data Steward	Data steward role	Sep 2, 2021 8:24 AM
<input type="checkbox"/> Developer		
<input type="checkbox"/> User		

14 permissions, 53 actions [Expand all](#)

Access governance artifacts
Administer platform
Create deployment spaces
Create projects

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

- 8- Finally, review all the values you have provided and click Add.

Add user
Authorize user access to the platform.

Profile information
 Platform access
 Roles
 Summary

Summary
Review the following summary. When you're ready to authorize the user, click Add.

Profile information

Name	datasteward
Username	datasteward
Email	datasteward@cpd.com
Password	*****

Roles

Data Steward	Data steward role
--------------	-------------------

Access

[Cancel](#) [Back](#) **Add**

- 9- The datasteward user will be created and listed in the list of users.

Access control

[IBM Cloud Pak for Data](#) All [Search](#) [LDAP configuration](#)

[Users](#) [User groups](#) [Roles](#)

Filter by: All roles

Name	User ID	Username	Created on	Roles
datasteward	1000331003	datasteward	Sep 2, 2021 8:54 AM	Data Steward
admin	1000330999	admin	Jul 1, 2021 12:59 PM	Administrator + 6 more

Add user +

Click Add user + again and repeat steps 5 to 8 to create the following users:

- a. dataengineer having the role of Data Engineer
- b. dqanalyst having the role of Data Quality Analyst
- c. datascientist having the role of Data Scientist
- d. dslead having the role of Data Scientist
- e. deploy having the role of Data Scientist

- 10- You have now created all the Cloud Pak for Data users required for the MLOps Workshop.

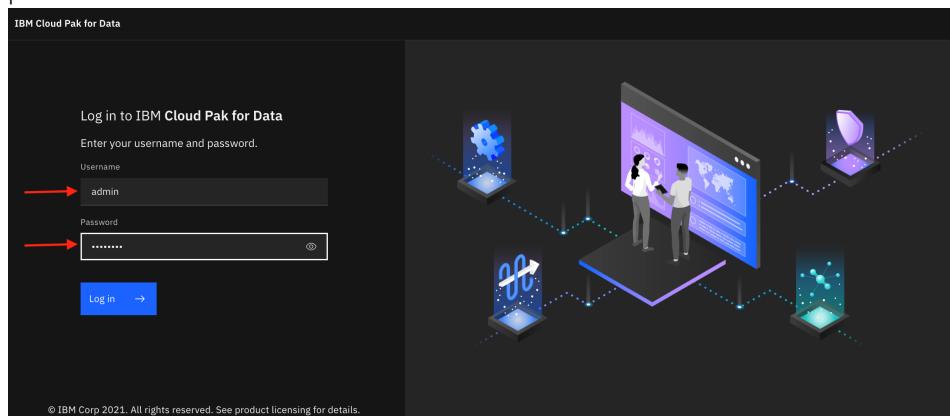
Name	Username	Email	Previous Session	User ID	Roles
deploy	deploy	—	—	1000331008	Data Scientist
dslead	dslead	—	—	1000331007	Data Scientist
datascientist	datascientist	—	—	1000331006	Data Scientist
dqanalyst	dqanalyst	—	—	1000331005	Data Quality Analyst
dataengineer	dataengineer	—	—	1000331004	Data Engineer
datasteward	datasteward	—	—	1000331003	Data Steward
admin	admin	--	April 16, 2023 8:42AM	1000330999	Administrator + 6 more

It is worthwhile spending some time reviewing the pre-defined Roles and associated permissions for each role as that influences what each user can and can NOT do on the Cloud Pak for Data platform.

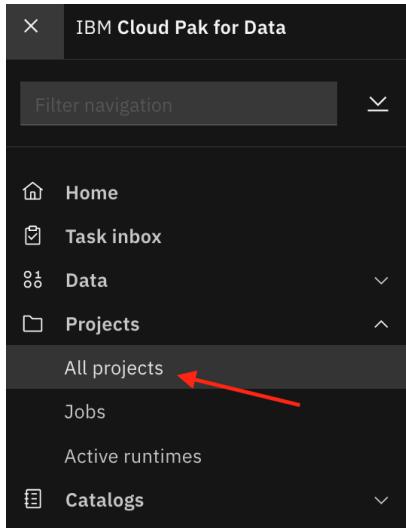
Create users using CPD's REST API

You can also add users via the Cloud Pak for Data REST API. You will now add the three users below by executing a Jupyter notebook in a project in Cloud Pak for Data.

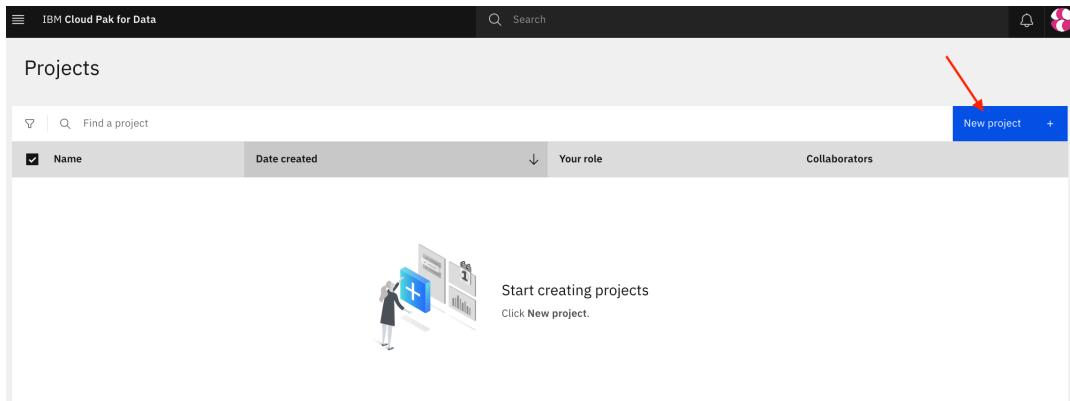
- 1- If logged out, log back into Cloud Pak for Data as admin user.
username: admin
password: CP4DDataFabric



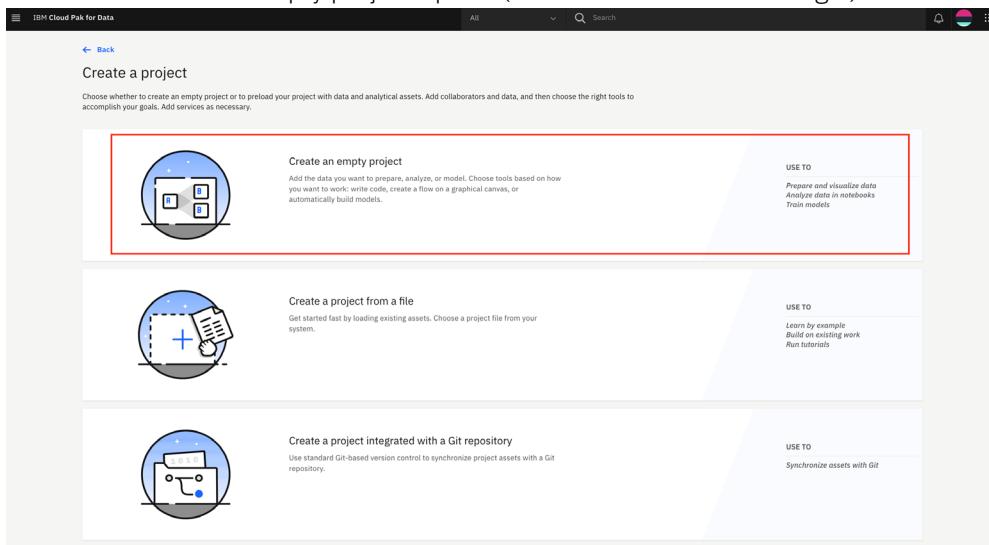
- 2- Select All projects by clicking on the Navigation menu (top left hamburger icon) and selecting Projects → All projects (annotated with red arrow).



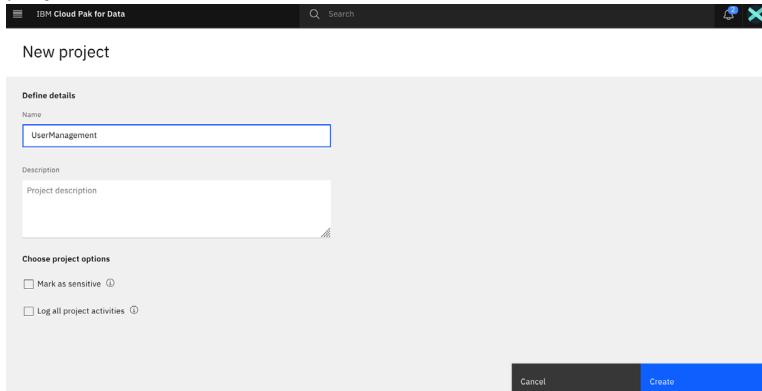
- 3- Click on New project (annotated with red arrow) to create a new project.



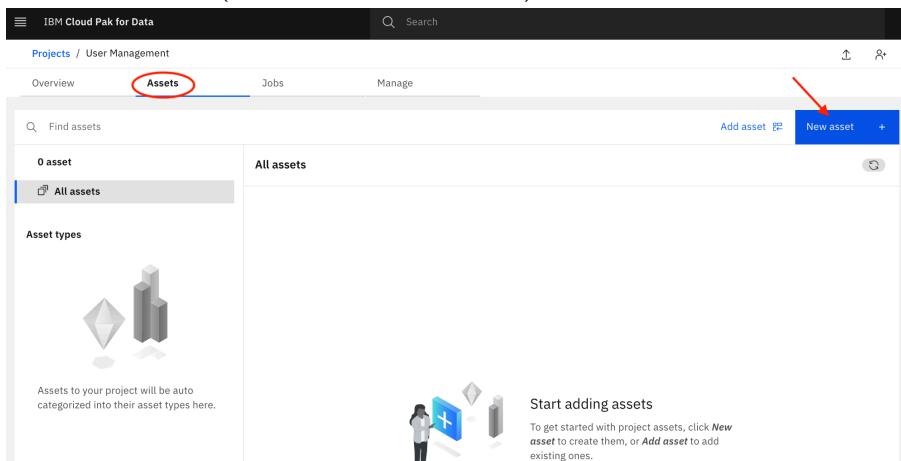
- 4- Select the Create an empty project option (annotated with red rectangle).



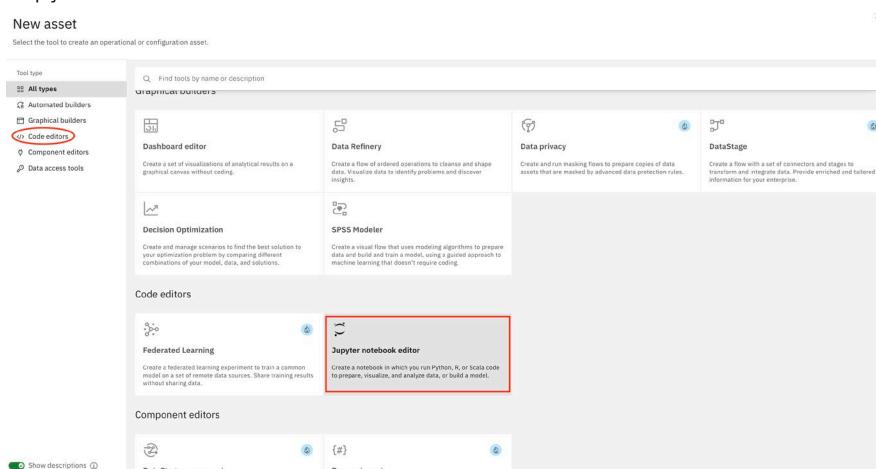
- 5- On the New project page, provide a Name (User Management) and Description (optional) for the project. Click Create.



- 6- On the User Management project page, click the Assets tab (annotated with red oval) and click the New asset button (annotated with red arrow).



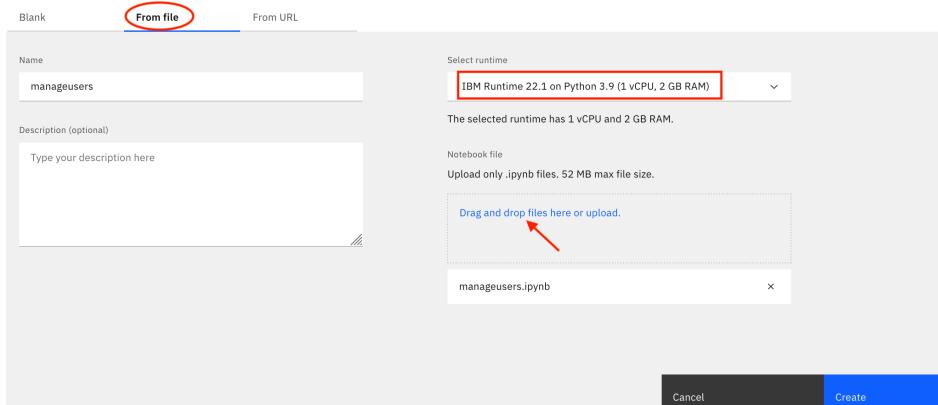
- 7- Scroll down and select the Jupyter notebook editor (annotated with red rectangle). Note that you can filter asset types by selecting the Code editors (annotated with red oval) to quickly find Jupyter notebook editor.



- 8- On the New notebook page, click From file tab (annotated with red oval) and click the Drag and drop files here or upload (annotated with red arrow). Select the manageusers.ipynb notebook to

upload. You should have downloaded this notebook from box to your computer (if not, download the manageusers.ipynb now). Verify the selected runtime is IBM Runtime 22.2 on Python 3.10 (1 vCPU, 2GB RAM) (annotated with red rectangle). Click Create.

New notebook



- 9- Once the notebook loads, review the cells and scroll to the bottom of the notebook to find the cell where you provide cpdurl, username, password, and user_number details (annotated with red arrows). Change the values to match your credentials and user id; for example:

```
# Cloud Pak for Data url
# Since you are running this notebook internal to the cluster, need to use the internal Cloud Pak for Data url
cpdurl = "https://cpd-cpd.apps.ocpinstall.gym.lan"
# Admin user credentials
username = "admin"
password = "CP4DDDataFabric"
# User number
user_number = "7"
```

```
# Cloud Pak for Data url
# Since you are running this notebook internal to the cluster, need to use the internal Cloud Pak for Data url
cpdurl = "<your_cloud_pak_for_data_url>" ←
# Admin user credentials
username = "<user_id_with_admin_privileges>" ←
password = "<password_for_username>" ←
# User number
user_number = "<user_number>" ←
```

Please note that if the notebook already existed in your project, you will need to click edit (the pencil icon) to make the updates in the notebook.

- 10- After you change that cell, click Cell (annotated with red oval) → Run All (annotated with red rectangle) to run all the cells in the notebook.

```

In [1]: # Provide user details for admin user, and the number for the user you'd like to create
        # Method will be POST
        # URL: /cpd/api/v1/users
        # Body: {
        #   "user": {
        #     "name": "admin",
        #     "email": "admin@cpd.com",
        #     "password": "CP4DdataFabric"
        #   }
        # }

        # Create users
        # Method will be POST
        # URL: /cpd/api/v1/users
        # Body: [
        #   {
        #     "name": "admin",
        #     "email": "admin@cpd.com",
        #     "password": "CP4DdataFabric"
        #   },
        #   {
        #     "name": "dataengineer",
        #     "email": "dataengineer@cpd.com",
        #     "password": "CP4DdataFabric"
        #   },
        #   {
        #     "name": "datasteward",
        #     "email": "datasteward@cpd.com",
        #     "password": "CP4DdataFabric"
        #   },
        #   {
        #     "name": "dataengineer2",
        #     "email": "dataengineer2@cpd.com",
        #     "password": "CP4DdataFabric"
        #   },
        #   {
        #     "name": "datasteward2",
        #     "email": "datasteward2@cpd.com",
        #     "password": "CP4DdataFabric"
        #   }
        # ]

        # Create users
        # Method will be POST
        # URL: /cpd/api/v1/users
        # Body: [
        #   {
        #     "name": "admin",
        #     "email": "admin@cpd.com",
        #     "password": "CP4DdataFabric"
        #   },
        #   {
        #     "name": "dataengineer",
        #     "email": "dataengineer@cpd.com",
        #     "password": "CP4DdataFabric"
        #   },
        #   {
        #     "name": "datasteward",
        #     "email": "datasteward@cpd.com",
        #     "password": "CP4DdataFabric"
        #   },
        #   {
        #     "name": "dataengineer2",
        #     "email": "dataengineer2@cpd.com",
        #     "password": "CP4DdataFabric"
        #   },
        #   {
        #     "name": "datasteward2",
        #     "email": "datasteward2@cpd.com",
        #     "password": "CP4DdataFabric"
        #   }
        # ]
    
```

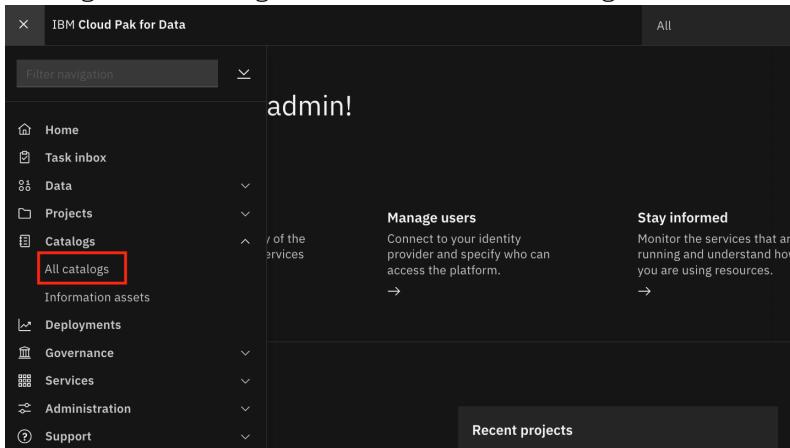
- 11- Review the print out from usrList which include the list of created users. You can also navigate to navigation menu → Administration → Access Control to verify all these users were created with the specified roles.

Create Catalog for Churn Data Assets

In this section, you step through the process of creating a new catalog to collect governed data assets that are relevant for purposes of customer churn prediction. Deciding how to organize your data into catalogs is your decision and depends on how you'd like to provide access to your end data consumers.

In this lab, we will create a sample catalog where data assets will be cataloged and made accessible to data consumers.

- 1- If you're logged out, navigate your favorite browser to Cloud Pak for Data url and login as admin user.
- 2- Navigate to All catalogs by clicking the Navigation menu (top left hamburger icon) and selecting Catalogs → All catalogs (annotated with red rectangle).



- 3- On the Your catalogs page, click New Catalog (annotated with red arrow).

Admin
Platform assets catalog
Platform catalog

Creator: admin Date created: Jan 27, 2023 12:20 PM

The catalog where all platform users can share assets and connections.

- 4- Provide a Name (Churn Data Catalog in the below example) and a Description (optional), click the checkbox to Enforce data protection rules (annotated with red arrow), select Overwrite original assets (annotated with red oval) and click Create.
 Hover over the information icon next to these various options to understand better how this selection affects the behavior of the catalog.

Name
Churn Data Catalog

Description
An intelligent catalog to organize governed assets for customer churn prediction use case.

Can't be undone.
You can't disable rule enforcement for a catalog after you enabled it.

Permanently enforce rules ⓘ
 Enforce data protection rules

Duplicate asset handling ⓘ
 Update original assets ⓘ
 Overwrite original assets ⓘ
 Allow duplicates ⓘ
 Preserve original assets and reject duplicates ⓘ

Cancel Create

- 5- On the Churn Data Catalog page, add the datasteward user as an Admin of that catalog. Navigate to the Access control page (annotated with red oval) and click Add Collaborators and then Add user (annotated with red arrow).

Assets Access control Settings

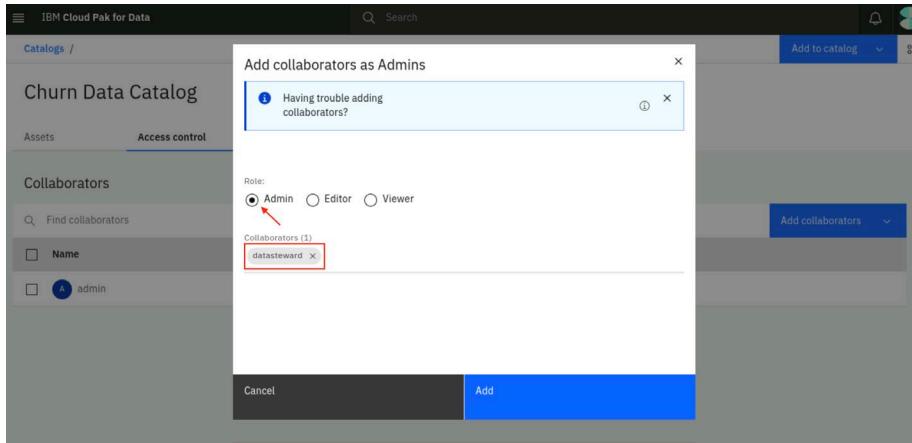
Collaborators

Name	Email	Role	Date added
admin	...	Admin	Aug 21, 2022

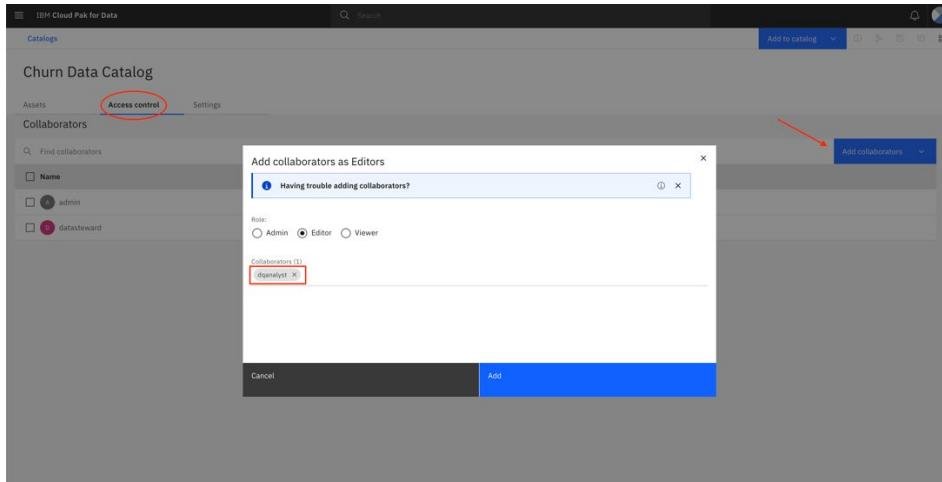
Add collaborators ^

Add user group
Add user

Select the Role as Admin (annotated with red arrow) and select datasteward user by typing 'datasteward' and selecting the 'datasteward' user (annotated with red rectangle). Click Add. This effectively gives the datasteward user permissions to be the administrator of the Churn Data Catalog.



- 6- Add the dqanalyst user (Data Quality Analyst role) as Editor of the Churn Data Catalog. Click the Access control tab (annotated with red oval) and click Add collaborators (annotated with red arrow), followed by Add user. Select dqanalyst user by typing dqanalyst (annotated with red rectangle), click the Editor radio button and click Add to give dqanalyst user the permissions to edit the Churn Data Catalog.
- Repeat to add datascientist user as editor of the Churn Data Catalog so that use can publish assets, such as AI models, to the Churn Data Catalog.



- 7- Give All users group Viewer access to the Churn Data Catalog. On the Access control tab (annotated with red oval), click Add collaborators, followed by Add user group (annotated with red arrow)

The screenshot shows the 'Churn Data Catalog' page. The 'Access control' tab is active. In the 'Collaborators' section, there is a table with columns: Name, Email, Role, and Date added. Four users are listed: admin (Admin, Apr 21, 2023), datasteward (Unavailable, Admin, Apr 21, 2023), dqanalyst (Unavailable, Viewer, Apr 21, 2023), and datascientist (Unavailable, Viewer, Apr 21, 2023). To the right of the table is a 'Add collaborators' dropdown menu with options: 'Add user group' (highlighted with a red box) and 'Add user'. A red arrow points from the 'Add user group' option to the 'All users' group in the 'Add access groups' dialog.

You can start typing All in the field and then select All users group. Select Viewer access level for this group. Click Add.

The screenshot shows the 'Add access groups' dialog. It includes a message 'Having trouble adding groups?', an 'Add access level' section with 'Viewer' selected, and a 'Groups (1)' section containing 'All users'. Below the dialog, the main catalog page shows the 'Access groups level: Viewer (8)' message. A red arrow points from the 'Add' button in the dialog to the 'Add' button in the 'Add collaborators' dropdown menu.

Default Catalog Access

In this section, you give access to the All users group to view the Default catalog.

- 1- If you're logged out navigate your favorite browser to Cloud Pak for Data url and login as admin user.
- 2- Navigate to All catalogs by clicking the Navigation menu (top left hamburger icon) and selecting Catalogs → All catalogs (annotated with red rectangle).

The screenshot shows the IBM Cloud Pak for Data interface. On the left, there's a navigation sidebar with links like Home, Task inbox, Data, Projects, Catalogs (which is expanded), Deployments, Governance, Services, Administration, and Support. Under Catalogs, 'All catalogs' is highlighted with a red rectangle. The main area has a dark background with white text. It features a 'Manage users' section and a 'Stay informed' section. At the bottom, there's a 'Recent projects' bar.

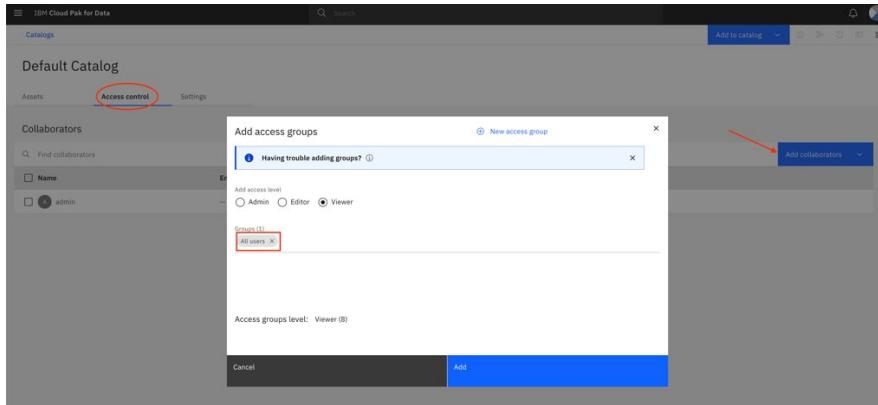
- 3- On the Your catalogs page, click the Default Catalog (annotated with red rectangle).

The screenshot shows the 'Your catalogs' page. It lists three catalogs: 'Platform assets catalog', 'Churn Data Catalog', and 'Default Catalog'. The 'Default Catalog' entry is highlighted with a red rectangle. Each catalog entry shows its creator ('admin'), creation date ('Jul 01, 2021 3:21 PM' or 'Oct 26, 2021 9:23 PM'), and a brief description. A 'Create Catalog' button is visible at the top right.

- 4- On the Default Catalog page, click the Access control tab (annotated with red oval) and click the Add collaborators, followed by Add user group (annotated with red arrow).

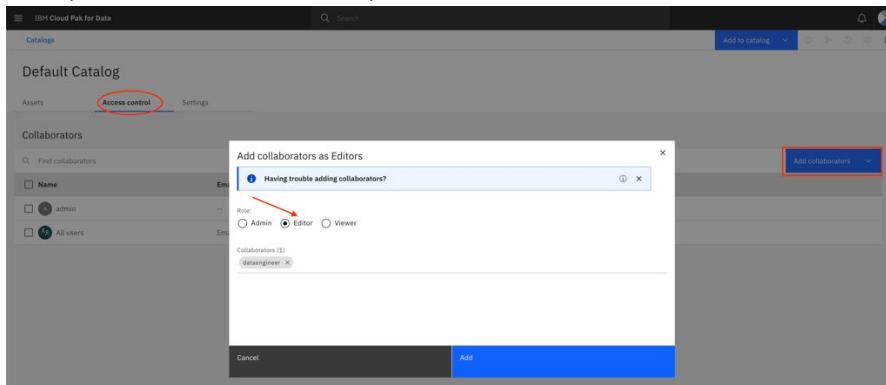
The screenshot shows the 'Default Catalog' page. The 'Assets' tab is selected, but a red oval highlights the 'Access control' tab. Below the tabs, there's a 'Collaborators' section with a table showing one user ('admin'). To the right, a modal window is open with a red border. Inside the modal, there are three buttons: 'Add collaborators', 'Add user group' (which is highlighted with a red arrow), and 'Add user'.

You can start typing All in the field and then select All users group (annotated with a red rectangle). Select Viewer access level for this group. Click Add.



This provides access to all users to view the Default catalog and access the cataloged data assets.

- 5- Make the dataengineer an Editor of the Default Catalog. This is required so the dataengineer can run profiling on the assets published in the Default Catalog. On Default Catalog page, click Access control tab (annotated with red oval) and then click Add collaborators (annotated with red rectangle), followed by Add user. Select dataengineer user by typing dataengineer and then selecting the identified dataengineer user. Select the Editor role to assign to the dataengineer user (annotated with red arrow) and click Add.

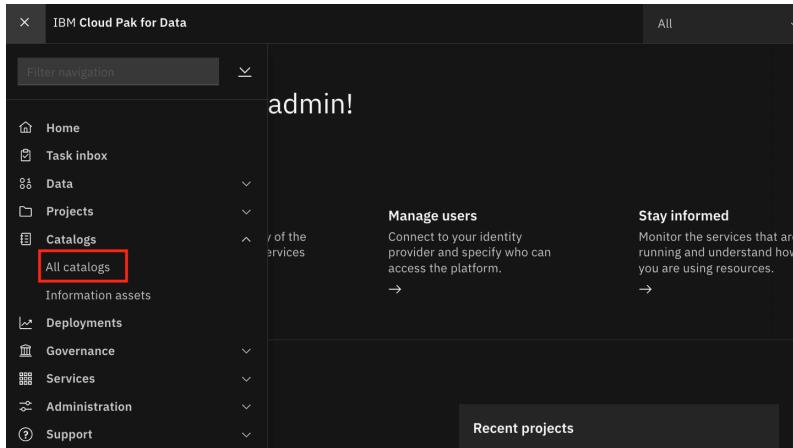


Platform Connections Access

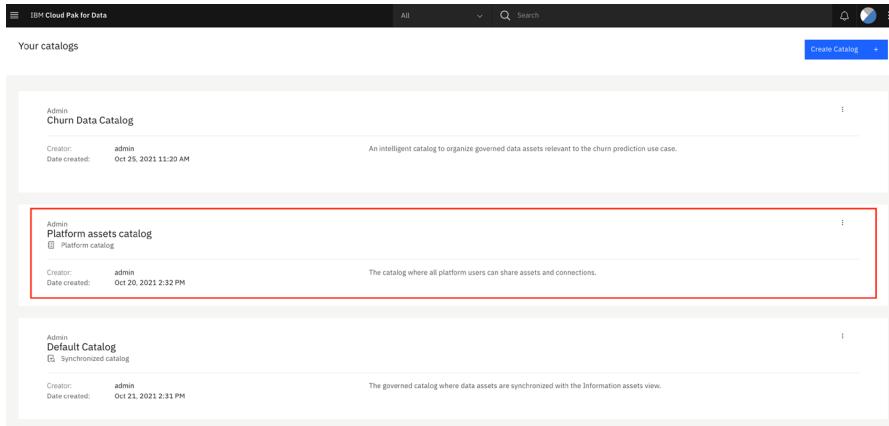
In this section, you will add the dataengineer user as an editor of Platform Assets catalog which in turn enables that user to create Platform Connections.

(optional) Before executing the steps in this section, feel free to try and log in as dataengineer, navigate to Data → Platform Connections and you will see that the dataengineer user does NOT have access to create new platform connections.

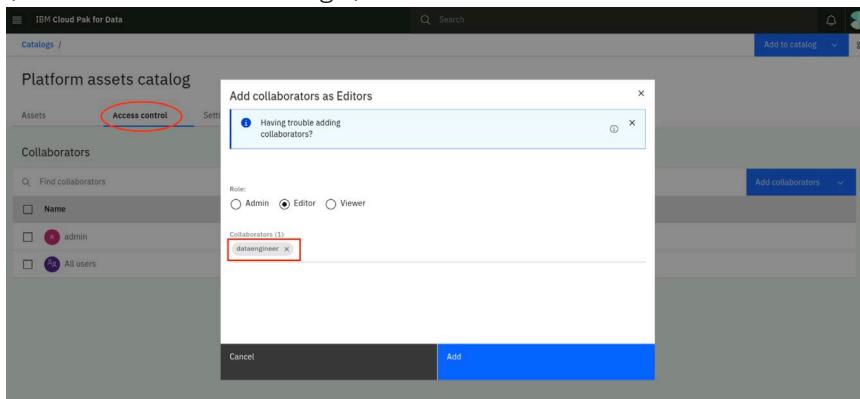
- 1- If you're logged out navigate your favorite browser to Cloud Pak for Data url and login as admin user.
- 2- Navigate to All catalogs by clicking the Navigation menu (top left hamburger icon) and selecting Catalogs → All catalogs (annotated with red rectangle).



- 3- On the Your catalogs page, click the Platform assets catalog (annotated with red rectangle).



- 4- On the Platform assets catalog page, navigate to the Access control tab (annotated with red oval) and click Add Collaborators (annotated with red arrow), followed by Add user. Select the Role as Editor and select dataengineer user by typing 'dataengineer' and selecting the 'dataengineer' user (annotated with red rectangle). Click Add.



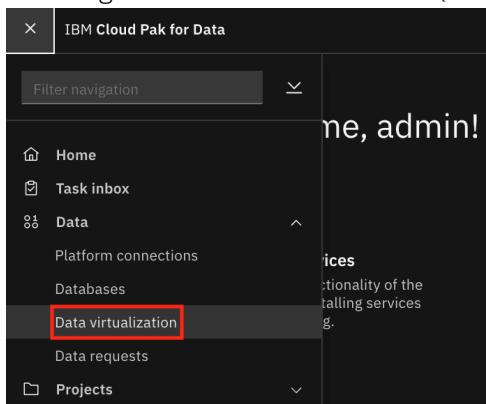
Adding the dataengineer as a user on the Platform assets catalog enables that user to create Platform Connections. For more details, review the Connecting to data sources at the platform level documentation.

Repeat the steps above to add the **datascientist** user as a **Viewer** on the **Platform assets catalog**.

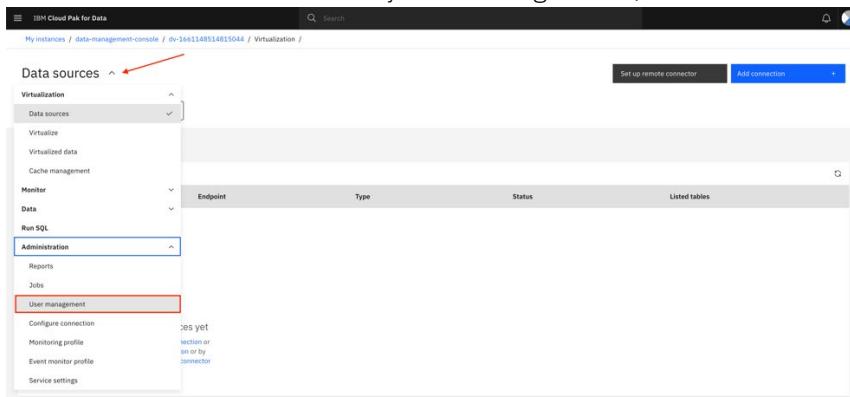
Data Virtualization Access

One of the tasks in this lab is to virtualize data assets from different sources. To enable the dataengineer to perform that task, the admin needs to provide the dataengineer with access to Data Virtualization service.

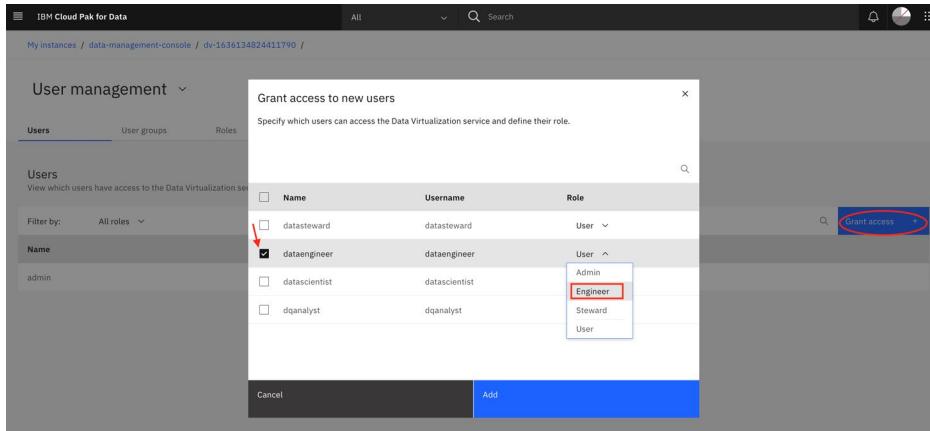
- 1- If you're logged out, navigate your favorite browser to Cloud Pak for Data url and login as admin user.
- 2- Navigate to Data Virtualization by clicking the Navigation menu (top left hamburger icon) and selecting Data → Data Virtualization (annotated with red rectangle).



- 3- On the Data Virtualization page, select the drop-down menu (annotated with red arrow) and then select Administration followed by User management (annotated with red rectangle).



- 4- On the User management page, click Grant access + (annotated with red oval). On the pop-up Grant access to new users window, check the box next to dataengineer (annotated with red arrow) and select the role as Engineer (annotated with red rectangle). This provides the dataengineer user access to Data Virtualization capabilities with the permissions associated with the Engineer role. Click Add.



- 5- Repeat the process to give access to datasteward, datascientist, and dqanalyst users as User role.

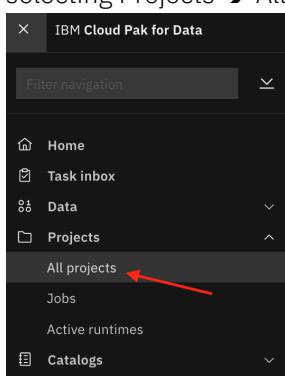
<input checked="" type="checkbox"/> datasteward	datasteward	User ▾
<input checked="" type="checkbox"/> datascientist	datascientist	User ▾
<input checked="" type="checkbox"/> dqanalyst	dqanalyst	User ▾

[Cancel](#) [Add](#)

Data Upload

In this section, you will upload data to Db2 database which is deployed on the same Cloud Pak for Data cluster. Our intent with this module is to illustrate some of the activities you need to setup the cluster so it has some usable data that you can leverage in the rest of the workshop. You can upload the csv files manually as described in the Data Upload (Manual) approach documented for reference in the appendix. However, in the spirit of automation, we developed a project and notebook that does the upload automatically and you will execute those steps next.

- 1- If you're logged out, navigate your favorite browser to Cloud Pak for Data url and login as **admin** user.
- 2- Navigate to your Projects view by clicking on the Navigation menu (top left hamburger icon) and selecting Projects ➔ All Projects (annotated with red arrow).

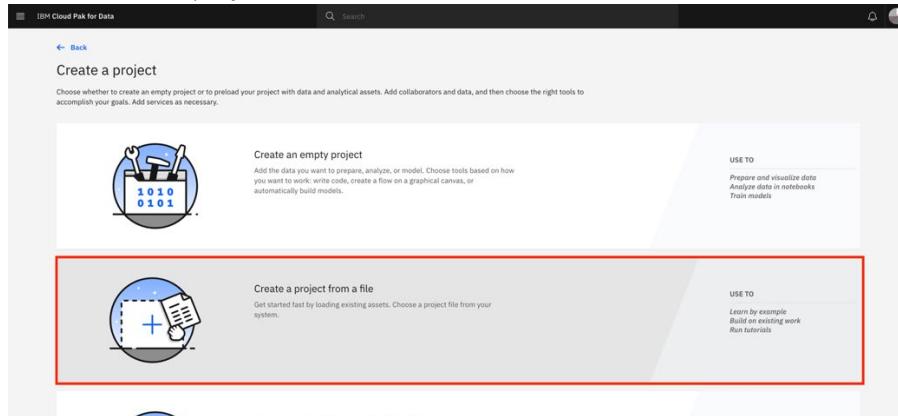


Governed MLOps Workshop – Setup Users, Catalogs, Data

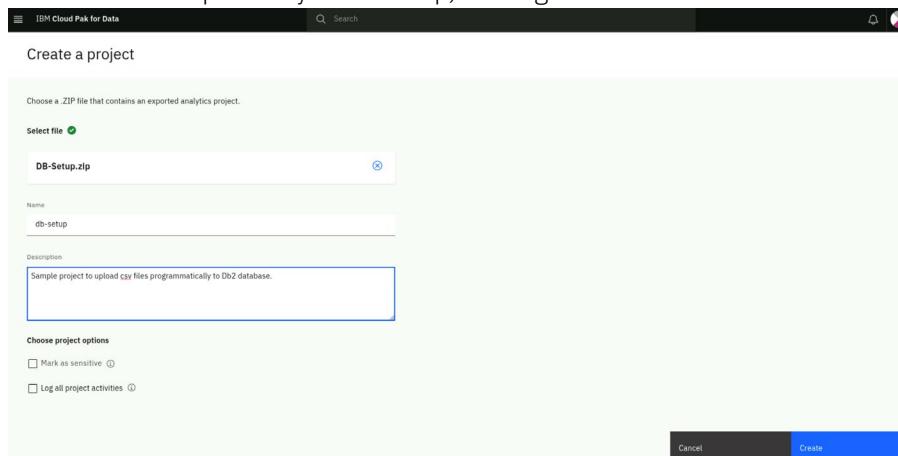
- 3- Create a new project by clicking the New Project button.



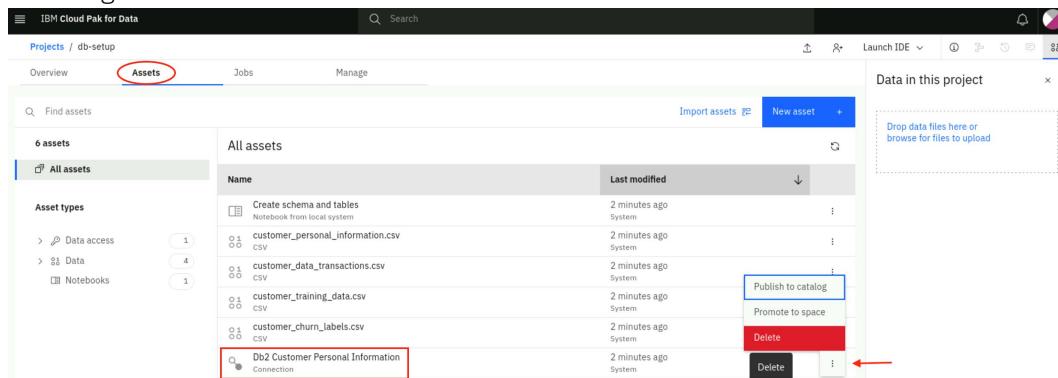
- 4- Select Create a project from a file.



- 5- Browse for the “[Db-Setup.zip](#)” zip file on your desktop (if you didn’t download earlier, please download that zip file to your desktop) or drag the file to the browser window.



- 6- Navigate to the newly imported project (db-setup) and click **Assets** tab (annotated with red oval). If you see an existing **Db2 Customer Personal Information** connection (annotated with red rectangle), delete that connection by selecting the actions menu (annotated with red arrow) and selecting **Delete**.



- 7- Next, click **New asset** button (annotated with red arrow) to add a connection to the project.

The screenshot shows the 'Assets' tab selected in the navigation bar. Below it is a table titled 'All assets' with columns for 'Name', 'Last modified', and actions. To the left, there's a sidebar for 'Asset types' with sections for 'Data' and 'Notebooks'. On the right, a sidebar titled 'Data in this project' says 'Drop data files here or browse for files to upload'.

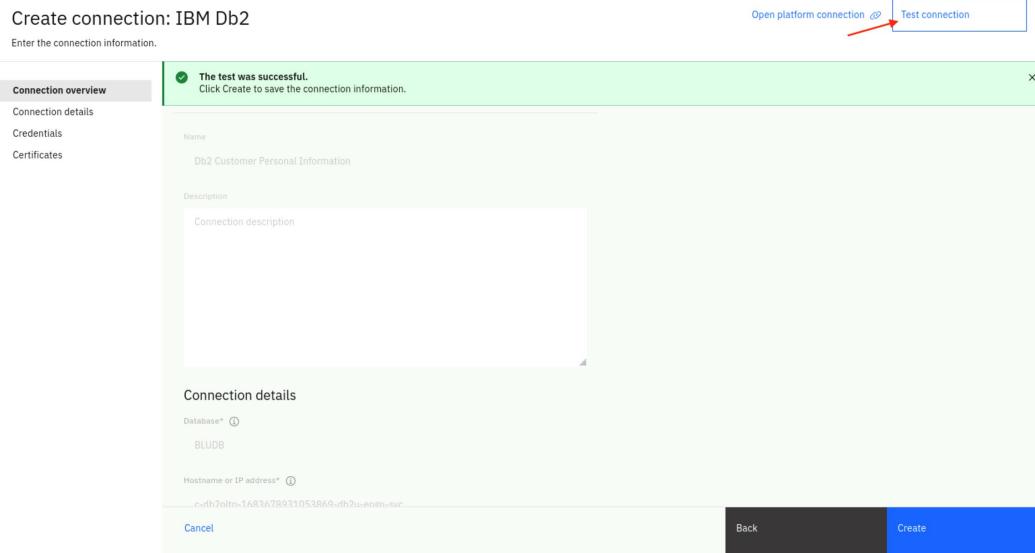
- 8- Select **Connection** tile (annotated with red rectangle).

The 'New asset' dialog has a sidebar for 'Tool type' with 'All types' selected. Under 'Data access tools', there are three tiles: 'Connection' (highlighted with a red rectangle), 'Metadata import', and 'Model'. Below this is a section for 'Automated builders' and 'Graphical builders'. At the bottom, there's a 'Show descriptions' link.

- 9- On the New connection page, click the **From platform** tab (annotated with red oval) and select the **Db2 Customer Personal Information** connection (annotated with red rectangle) and click **Select**.

The 'New connection' dialog has a 'Provider' dropdown set to 'IBM'. Under 'All connection types', the 'Db2 Customer Personal Information' connection is highlighted with a red rectangle. To the right, a sidebar shows 'Selected platform connection' with 'Db2 Customer Personal Information' listed, 'Connection type' as 'IBM Db2', and a 'Description' field. At the bottom are 'Cancel' and 'Select' buttons.

10- Next, click **Test connection** button (annotated with red arrow) and once the test returns successful, click **Create**.



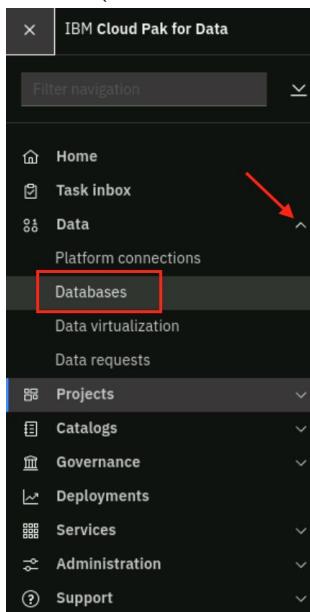
11- Back on the project Assets tab, find the **Create schema and tables** notebook (annotated with red rectangle) and open that notebook in edit mode by selecting the actions menu (annotated with red arrow) and selecting **Edit** (annotated with red oval).

12- Execute the notebook cell-by-cell and confirm all cells execute correctly. Select the cell and click **Run** (annotated with red arrow) and repeat until all cells are run.

```
In [1]: connectionName='DB2 Customer Personal Information'

In [2]: import itc.utils.flightservice as itcfs
from ibm.watson.studio.lib import access.project_or_space
import pyarrow.flights as flight
import pyarrow as pa
```

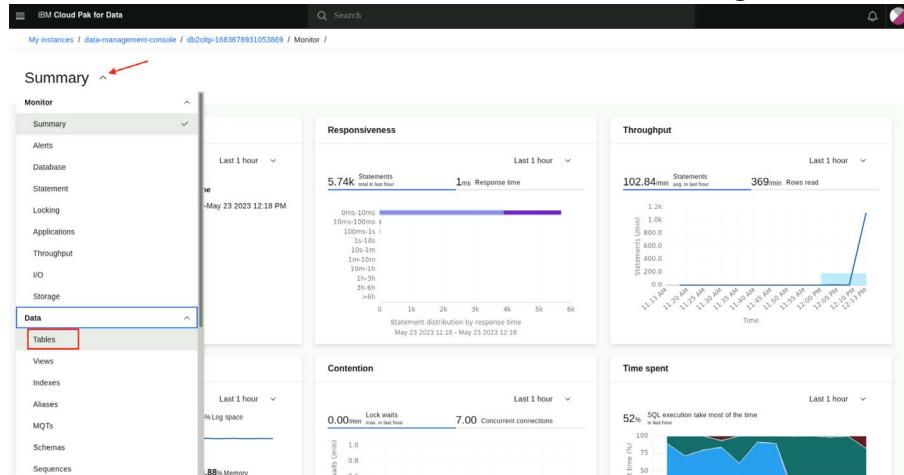
- 13- At this point, the Db2 database should be populated by a set of tables under CUSTOMER schema. To confirm, navigate to databases by selecting the Navigation menu (top left), expanding the **Data** section (annotated with red arrow) and selecting **Databases** (annotated with red rectangle).



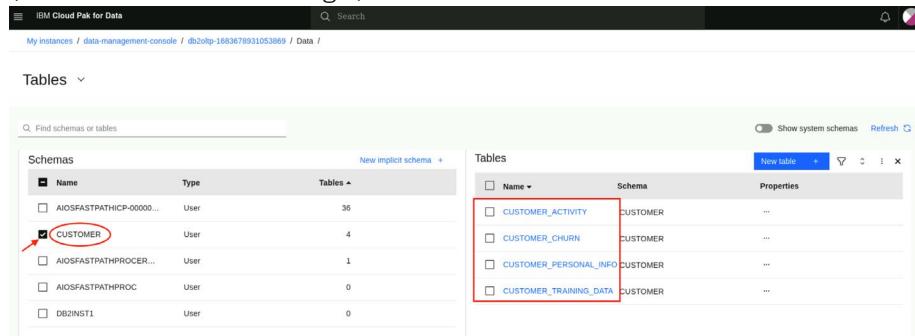
- 14- On the Databases page, find the **Db2-1** tile, click the actions menu (annotated with red arrow) and select **Open database** (annotated with red rectangle).

The image shows the 'Databases' page in IBM Cloud Pak for Data. A tile for 'Db2-1' is selected. An actions menu (three dots) is open above the tile, and the 'Open database' option is highlighted with a red rectangle. Other options in the menu include 'Details', 'Submit connection for approval', 'Manage access', and 'Delete'. A red arrow points to the three-dot menu icon.

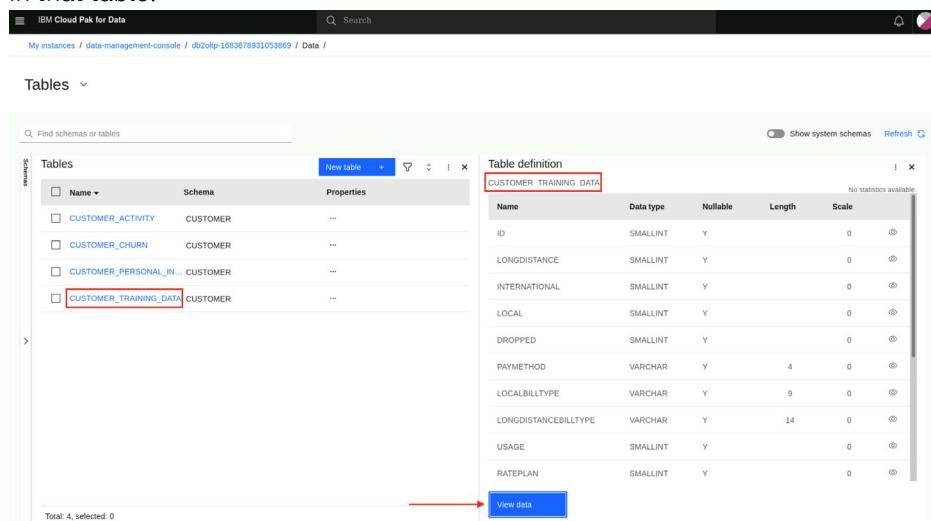
- 15- On the database UI, select the drop down next to Summary (annotated with red arrow), expand the Data section and select **Tables** (annotated with red rectangle).



- 16- Observe the **CUSTOMER** schema (annotated with red rectangle) which was created using the notebook you executed. Select the checkbox next to CUSTOMER schema (annotated with red arrow) and observe the 4 database tables which were created using the notebook as well (annotated with red rectangle).



- 17- Select any of the tables (for example, **CUSTOMER_TRAINING_DATA**), and review the table definition. Click the **View data** button (annotated with red arrow) to review the actual data stored in that table.



Refresh OpenPages Db2 Certificates

In this section, we describe how to refresh the Db2 certificates for the OpenPages db2u pod if needed. Note that the steps below need to be executed from a terminal in the bastion VM and you need credentials to access OpenShift.

- From your TechZone reservations page, capture the username (ocpadmin) and password (annotated with red arrow). Next click the url link for the Bastion Remote Desktop (annotated with red rectangle).

The screenshot shows a reservation page with the following details:

- Purpose:** Proof-of-Technology
- Opportunity ID(s):** 3P054AYV1LXVPMDUOUDSSCP4HQ1DK88D1
- Opportunity description:** Sep 13 AI Gov
- Share your feedback with us:**
 - How would you rate the reservation product? (Scale 0-10)
 - How would you rate the reservation infrastructure? (Scale 0-10)
 - How would you rate the reservation experience? (Scale 0-10)
- Environment:**
 - Reservation ID: 650220056b272c0017749a4e
 - Type: IBM Cloud
 - Request method: vmware-template
 - Cloud Account: ITZSQUAD
 - Region: us-east
 - Region: vmware-template
 - Environment: 650220056b272c0017749a4e
 - Timeout action: Cloud Pak for Data URL (when using VPN) <https://cpd-ond-gaps.ocpinstall.gym.lan>
 - CP4D credentials: User=admin, Password=CP4DDataFabric
 - OCF URL (when using VPN) <https://console-openshift-console.apps.ocpinstall.gym.lan>
 - OCF credentials: Super User : kubeadmin, Password : 4o6Xu-cSUNk-dijav-o5oGp, Admin User : ocpadmin, Password : X1ObejCUMWkVTBz35gqj
 - Bastion Remote Desktop URL: <https://remote-cloud.techzone.ibm.com/guacamole/#/?username=itz-550220056b272c0017749a4e&password=o2WCFB8mfd-Etgv>

- Open up a Terminal in the Bastion VM.
- On the Terminal, log into OpenShift using the username (ocpadmin) and the password you captured in step 1.

```
oc login -u ocpadmin -p X1ObejCUMWkVTBz35gqj
```

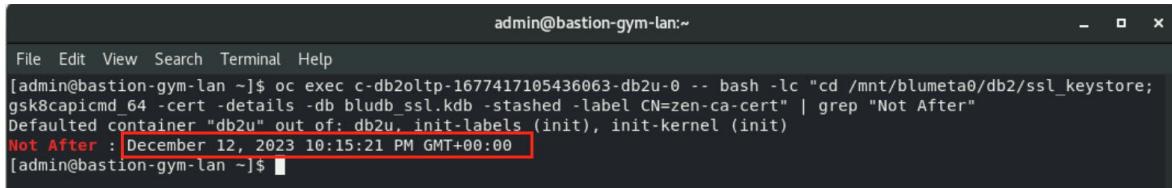
```
Activities Terminal Sep 25 10:56 AM
admin@bastion-gym-lan:~ File Edit View Search Terminal Help
[admin@bastion-gym-lan ~]$ oc login -u ocpadmin -p X1ObejCUMWkVTBz35gqj
Login successful.

You have access to 70 projects, the list has been suppressed. You can list all projects with 'oc projects'

Using project "cpd".
[admin@bastion-gym-lan ~]$
```

- Check whether the db2 certificates are expired by executing the following command:

```
oc exec c-db2oltp-1677417105436063-db2u-0 -- bash -lc "cd /mnt/blumeta0/db2/ssl_keystore; gsk8capicmd_64 -cert -details -db bludb_ssl.kdb -stashed -label CN=zen-ca-cert" | grep "Not After"
```



```
admin@bastion-gym-lan:~$ oc exec c-db2oltp-1677417105436063-db2u-0 -- bash -lc "cd /mnt/blumeta0/db2/ssl_keystore; gsk8capicmd_64 -cert -details -db bludb_ssl.kdb -stashed -label CN=zen-ca-cert" | grep "Not After"
Defaulted container "db2u" out of: db2u, init-labels (init), init-kernel (init)
Not After : December 12, 2023 10:15:21 PM GMT+00:00
[admin@bastion-gym-lan ~]$
```

If the Not After date (annotated with red rectangle) is before the current date, then the certificates are expired and need to be renewed by following the steps below. Otherwise, the certificates are valid and you can skip the remaining steps in this section.

- 5- Download the following two scripts from Box to your Bastion VM. In your Bastion VM, start your Firefox browser and navigate to the links below and download these scripts.

fix_openpages_db2_certs.sh: <https://ibm.box.com/s/5icfjhsld8np222z7eriw7m0th600s>

refresh_openpages_db2_certs.sh: <https://ibm.box.com/s/o2sdvm8wzhj7mo6rpwu13y5xq0mamjsx>

- 6- On your Terminal in the bastion VM, execute the following commands to execute the script to refresh db2 certificates:

```
mkdir /home/admin/scripts
cd /home/admin/scripts
mv /home/admin/Downloads/fix_openpages_db2_certs.sh /home/admin/scripts/
mv /home/admin/Downloads/refresh_openpages_db2_certs.sh /home/admin/scripts/
chmod 755 fix_openpages_db2_certs.sh
chmod 755 refresh_openpages_db2_certs.sh
./fix_openpages_db2_certs.sh
```

- 7- The script should run for a couple of minutes. When it concludes, re-run the command in step 4 to check the status of the certificates. Verify the date is after the current date.

```
oc exec c-db2oltp-1677417105436063-db2u-0 -- bash -lc "cd /mnt/blumeta0/db2/ssl_keystore; gsk8capicmd_64 -cert -details -db bludb_ssl.kdb -stashed -label CN=zen-ca-cert" | grep "Not After"
```

Summary

In this exercise, you executed multiple steps to setup and configure your Cloud Pak for Data cluster including creating users, uploading data to Db2 and providing access and permissions to the various users so they can perform their tasks. In an enterprise Data and AI platform, separation of roles is critical and its important that users with given roles have the right permissions to be able to perform their jobs efficiently.

Appendix

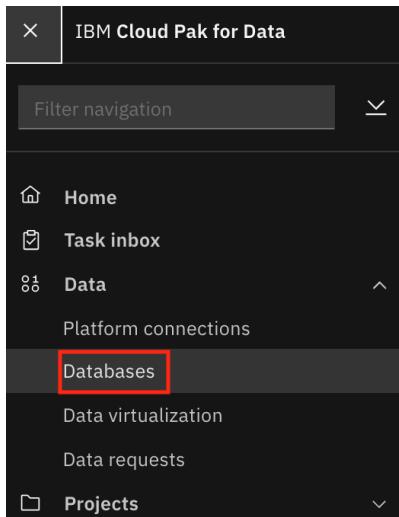
Data Upload (Manual)

In this section, you will upload a sample csv file to the Db2 Warehouse instance on your Cloud Pak for Data cluster. In practice, the databases would already exist and be populated with data but for purposes of this lab, since this is a fresh install, you don't have any data so we'll emulate the functionality by uploading csv files to Db2 Warehouse. For this task, download the csv files to your local machine.

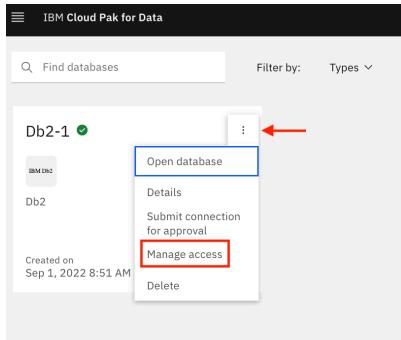
The instructions below assume that a Db2 Warehouse instance that is provisioned on the same Cloud Pak for Data cluster. If you don't see Db2 Warehouse or Db2, and you have Db2 installed but not provisioned, follow the provisionDb2_on_CPDv4.mp4 video to provision a Db2 instance. The rest of the instructions assume you have a Db2 Warehouse instance but they also apply if you have a Db2 instance provisioned in your cluster.

Start by providing the dataengineer user with access to the Db2 instance.

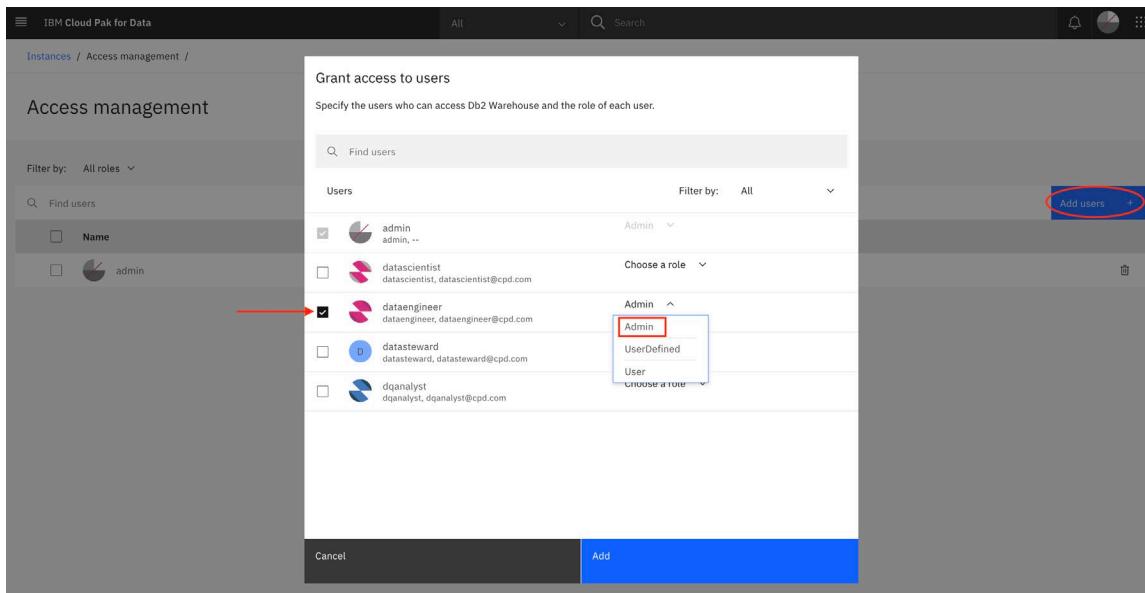
- 1- If you're logged out, navigate your favorite browser to Cloud Pak for Data url and login as admin user.
- 2- Navigate to Databases by clicking the Navigation menu (top left hamburger icon) and selecting Data → Databases (annotated with red rectangle).



- 3- On the databases page, find your Db2 instance, click the open and close list of options menu (annotated with red arrow) and click Manage access (annotated with red rectangle).

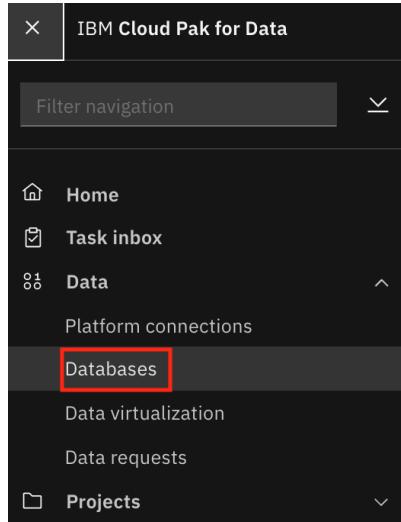


- 6-** On the Access management page, click Add users + (annotated with red oval). On the pop-up Grant access to users window, check the box next to dataengineer (annotated with red arrow) and select the role as Admin (annotated with red rectangle). This provides the dataengineer admin access to the Db2 instance. Click Add.

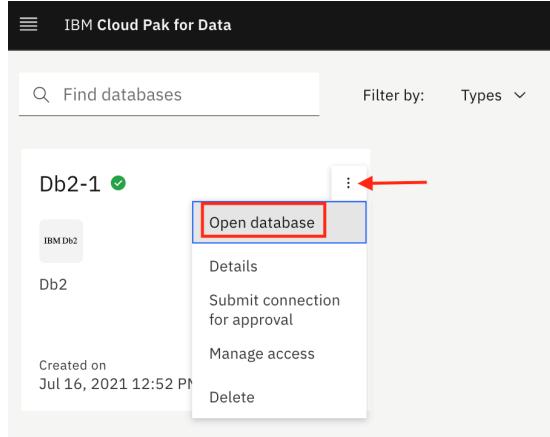


Once the dataengineer user has access rights to the Db2 Warehouse instance, you can login as the dataengineer user to load the data.

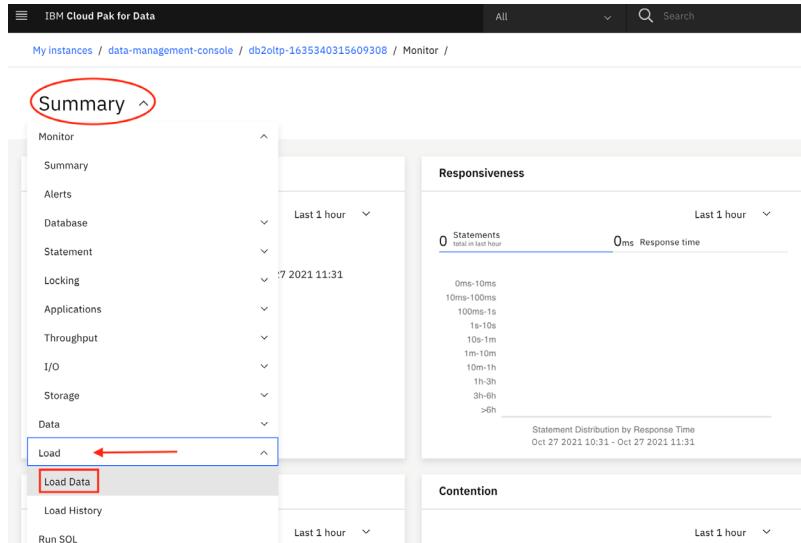
- 1- Navigate your favorite browser to Cloud Pak for Data url and login as dataengineer user.
- 2- Navigate to Databases by clicking the Navigation menu (top left hamburger icon) and selecting Data → Databases (annotated with red rectangle).



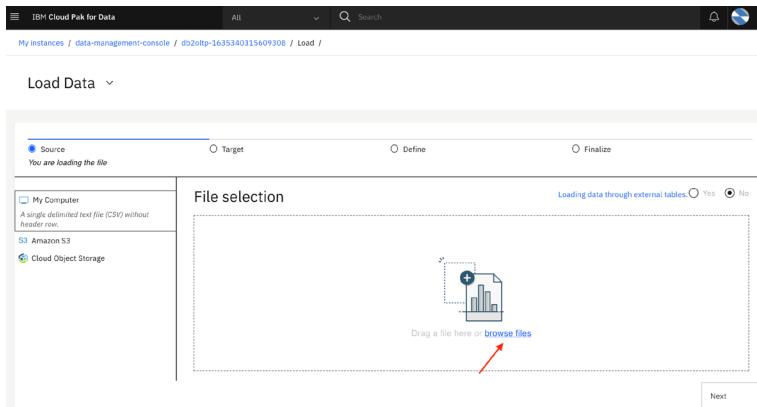
- 3- On the databases page, find your Db2 instance, click the options menu (annotated with red arrow) and click Open database (annotated with red rectangle).



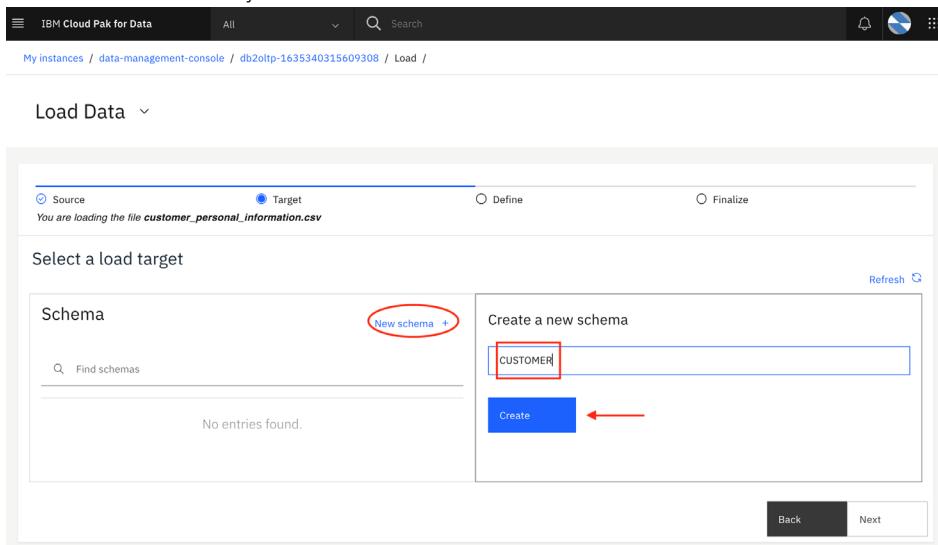
- 4- Click the Summary drop down (annotated with red oval), expand Load drop down (annotated with red arrow) and select Load Data (annotated with red rectangle).



- 5- On the Load Data menu, click the browse files link (annotated with red arrow) and select the customer_personal_information csv file (you would have the csv file downloaded to your laptop). Click Next.



- 6- On the schema page, click New schema + (annotated with red oval), specify the schema name as CUSTOMER (annotated with red rectangle) and click Create (annotated with red arrow). For future uploads of data sets to the CUSTOMER schema, you can select the CUSTOMER schema if it is already created.



- 7- On the Table page, click New table + (annotated with red oval), provide a table name as CUSTOMER_PERSONAL_INFO (annotated with red rectangle) and click Create (annotated with red arrow).

Governed MLOps Workshop – Setup Users, Catalogs, Data

The screenshot shows the 'Load Data' interface in IBM Cloud Pak for Data. The 'Target' tab is selected. On the left, under 'Schema', 'CUSTOMER' is selected. On the right, under 'Table', 'CUSTOMER_PERSONAL_INFO' is highlighted with a red rectangle. A modal window titled 'Create a new table' contains the table name 'CUSTOMER_PERSONAL_INFO' and a 'Create' button, which is also highlighted with a red rectangle and has a red arrow pointing to it.

- 8- Select the CUSTOMER_PERSONAL_INFO as the table to upload data to (annotated with red rectangle) and click Next (annotated with red arrow).

The screenshot shows the 'Load Data' interface in IBM Cloud Pak for Data. The 'Table definition' step is shown. The table 'CUSTOMER_PERSONAL_INFO' is selected. A red rectangle highlights the table name 'CUSTOMER_PERSONAL_INFO' in the 'Table definition' section, and a red arrow points to the 'Next' button at the bottom right.

- 9- Review the data sample that appears on the next page. You need to change the DOB column data type from DATE to VARCHAR. To do so, click the pencil icon (annotated with red arrow) next to DATE type field under the DOB column.
Note that you may need to scroll right and/or down to find the DOB column.

Governed MLOps Workshop – Setup Users, Catalogs, Data

The screenshot shows the 'Load Data' interface in IBM Cloud Pak for Data. It's loading a CSV file named 'customer_personal_information.csv' into the 'CUSTOMER.CUSTOMER_PERSONAL_INFO' table. The target schema is defined with columns: ID (SMALLINT), GENDER (VARCHAR(1)), STATUS (VARCHAR(1)), CHILDREN (SMALLINT), ESTINCOME (DECIMAL(8,2)), CAROWNER (VARCHAR(1)), AGE (DECIMAL(14,5)), CREDITCARD (BIGINT), and DOB (DATE). The 'DOB' column is highlighted with a red rectangle. The 'Define' tab is selected.

Change the data type from DATE to VARCHAR (annotated with red rectangle) and click OK.

This screenshot shows the 'Edit column data type' dialog for the 'DOB' column. The current data type is 'DATE'. A dropdown menu lists several options: INTEGER, SMALLINT, BIGINT, DECIMAL, and VARCHAR. The 'VARCHAR' option is highlighted with a red rectangle. At the bottom, there are 'Close' and 'OK' buttons.

After changing the data type from DATE to VARCHAR for DOB column, click Next.

10- Click Begin Load (annotated with red arrow) to begin loading the csv file into the CUSTOMER_PERSONAL_INFO table on your Db2 instance which is provisioned on the same Cloud Pak for Data cluster.

This screenshot shows the 'Review settings' dialog. It contains two sections: 'Summary' and 'Option'. The 'Summary' section includes fields for Code page (1208), Separator (,), Time format (HH:MM:SS), Date format (YYYY-MM-DD), Timestamp format (YYYY-MM-DD HH:MM:SS), and String delimiter (Default). The 'Option' section has a field for Maximum number of warnings set to 1000. At the bottom right, there are 'Back' and 'Begin Load' buttons, with 'Begin Load' highlighted by a red rectangle.

11- The load operation will run for a minute or two and you can monitor progress through the steps (annotated with red rectangle below).

Governed MLOps Workshop – Setup Users, Catalogs, Data

The screenshot shows the 'Load Data' interface in IBM Cloud Pak for Data. A file named 'customer personal information.csv' is being uploaded to a target table 'CUSTOMER.CUSTOMER_PERSONAL_INFO'. The status bar indicates 'Uploading' and '100% completed'. A tooltip provides instructions: 'Did you know? On the Run SQL page, you can create tables and procedures, run queries, and export query results.' Below the status bar, a vertical stack of four steps is shown: 'Auto-create table', 'Upload', 'Load data', and 'Complete', with 'Upload' highlighted by a red box. The right side of the screen shows error and warning counts.

12- When data is loaded, you should see a summary message indicating the data load job succeeded.

The screenshot shows the 'Load Data' interface after the job has succeeded. The summary panel displays '1,416 Rows read' and '1,416 Rows loaded' with '0 Rows rejected'. It also shows the start and end times: 'Start time 10/27/2021 12:24:48 PM' and 'End time 10/27/2021 12:27:10 PM'. A prominent message states 'The data load job succeeded. You can now work with your data.' A note below it says 'No errors' but 'But, there are 21 messages.' The right side of the screen shows error and warning counts.

13- Repeat steps 4-12 to load 3 other csv files into your Db2 instance and create new tables.

Please note that in step 6, you just select CUSTOMER schema as it is already created and you can skip step 9 as that issue is specific to the first dataset only.

customer_data_transactions csv file ➔ CUSTOMER_ACTIVITY table

customer_churn_labels csv file ➔ CUSTOMER_CHURN table

customer_training_data csv file ➔ CUSTOMER_TRAINING_DATA table