

IBM CLOUD PAK FOR DATA 4.5

DATA VIRTUALIZATION HANDS ON LAB

Sandeep Ved
CUSTOMER SUCCESS MANAGER ARCHITECT, IBM SINGAPORE

Table of Contents

Data Virtualization Lab	3
Access Your Cluster.....	3
Manage Users	3
Service Provisioning	6
DV and DB2.....	7
PostgreSQL	9
Db2 data loading	9
Check DB2 Connection Details.....	14
Setting up Data Virtualization.....	17
Task 1: Create a project.....	17
Task 2: Add a connection to your Db2 data source	18
Task 3: Add a connection to a PostgreSQL data source	21
Task 4: Add tables to your virtualized data	23
Task 5: Join and Publish virtualized data to project	26
Next steps.....	31
Analyze the data in a notebook	31
Visualize the data with a dashboard	33
Appendix.....	36
DB2 Instance Provisioning.....	36

Data Virtualization Lab

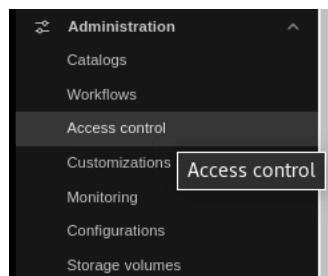
Access Your Cluster

- 1) Find out the OpenShift cluster Details as per the Given Cluster URL and credential provided.
- 2) Open the Chrome/Firefox Browser in Incognito or Private mode to Open the Cloud pak for Data Web Console.eg.
cpd-cpd-instance.apps.itzocp-664003qtcr-mn9o.selfservice.aws.techzone.ibm.com
- 3) Login to Web Console using **admin** ID and credentials provided.

Manage Users

- 1) Once login, you will see the CP4D Console Home Page.

- 2) An Admin User can further create more users eg. their team members to share the platform functionalities with them with appropriate access control. To create users, Navigate to **Administration -> Access Control**.



- 3) You can see the current list of Users. Click on **Add User +** to add more users.

IBM Cloud Pak for Data

Search

Access control

LDAP configuration [🔗](#)

Users User groups Roles

Filter by: All roles [▼](#)

Find users [Add users +](#)

Name	User ID	Username	Created on	Roles
 admin	1000330999	admin	6 Sept 2022 23:34	Administrator + 2 more

4) Define the new user profile information. Click **Next**.

Add users

Authorize user access to the platform.

Profile information

Platform access

Summary

Profile information
Specify the identification and authentication information for the users you want to add.

Full name (optional)
Sandeep Ved

Username
sved

Email (optional)
sandeep.ved1@ibm.com

Password
..... [👁](#)

Confirm password
..... [👁](#)

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

5) Select **Assign Roles Directly** and click on **Next**.

Add users

Authorize user access to the platform.

Profile information

Platform access

Roles

Summary

Platform access
Privileges on the platform are controlled by permissions. Users are administered permissions through role assignment. Users can be assigned roles directly or inherit them from groups they are added to.

 Assign roles directly

 Add to user group

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

6) Select the Data Engineer Role. To see the full details of all roles and permissioned, refer [here](#).

Add users

Authorize user access to the platform.

Profile information

Platform access

Roles

Summary

Administrator

Description
Administrator role

Modified on
Aug 30, 2022 7:57 AM

13 permissions, 73 actions [Expand all](#)

Access governance artifacts

Administer platform

Create deployment spaces

Create projects

Create service instances

Manage asset discovery

Cancel Back Next

7) Click Next.

Add users

Authorize user access to the platform.

Profile information

Platform access

Roles

Summary

Administrator

Business Analyst

Data Engineer

Description
Data engineer role

Modified on
12 Sept 2022 22:13

1 permission, 1 action [Expand all](#)

Create projects

Cancel Back Next

8) Click Add on the Summary Window.

Add users

Authorize user access to the platform.

Profile information

Platform access

Roles

Summary

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

Profile information

Username	Name	Email
sved	Sandeep Ved	sandeep.ved1@ibm.com

Roles

Role name	Description
Data Engineer	Data engineer role

Access

Permission	Actions
Create projects	Create new projects

Cancel Back Add

9) The New User will be added. Similarly you may add more users as needed for your team members.

Access control

[LDAP configuration](#)

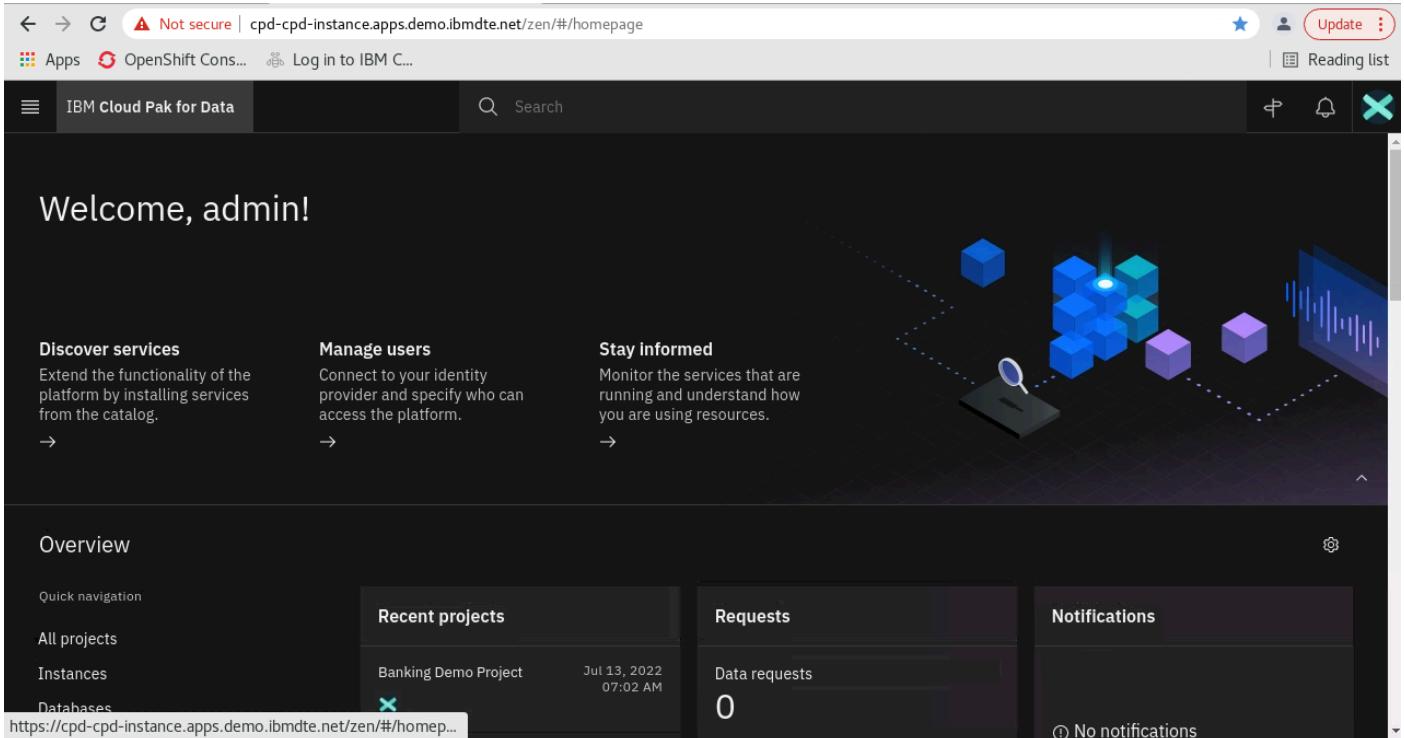
Users

Name	User ID	Username	Created on	Roles
SV Sandeep Ved	1000331001	svd	13 Sept 2022 18:13	Data Engineer
admin	1000330999	admin	6 Sept 2022 23:34	Administrator + 2 more

Service Provisioning

This lab requires the Data Virtualisation service as well as a Db2 service to prepare the data source.

- 1) Navigate back to CP4D Web Console Home Page.



Not secure | cpd-cpd-instance.apps.demo.ibmdte.net/zen/#/homepage

Apps OpenShift Cons... Log in to IBM C...

Welcome, admin!

Discover services
Extend the functionality of the platform by installing services from the catalog.

Manage users
Connect to your identity provider and specify who can access the platform.

Stay informed
Monitor the services that are running and understand how you are using resources.

Overview

Recent projects: Banking Demo Project (Jul 13, 2022 07:02 AM)

Requests: 0

Notifications: 0

Quick navigation: All projects, Instances, Databases

- 2) Review all the Menu Items. Click on **Services -> Instances** to check the status of all instances.

DV and DB2

- 3) Review the currently provided instances and its status. The status for all the services should be **green**. Data Virtualization Instance (dv) and DB2 (db2oltp) should already be provisioned, along with Data Management Console (dmc) to manage the DB2/Data Virtualization Data.
- 4) Click **New Instance +** Button to see what all services are already enabled.

Name	Type	Created by	vCPU requests	Memory requests (GiB)	Users	Status	Created on
data-virtualization	dv	admin	11.50	38.50 Gi	1	green	13 Sept 2022
data-management-console IBM Db2 Data Management Console	dmc	admin	5.20	21.40 Gi	1	green	8 Sept 2022
Db2-1 Service instance for db2oltp-16625688212...	db2oltp	admin	0.60	4.25 Gi	1	green	8 Sept 2022
openscale-defaultinstance IBM Watson OpenScale	aos	admin	-	-	1	green	7 Sept 2022

- 5) Review the status for all services. You will see that many services like Watson Studio or Cognos Dashboard are already enabled.

IBM Cloud Pak for Data

Instances / Select a service /

Select a service

Select the service for which you want to create a new instance.

AI 3 items

 Watson Machine Learning IBM Build and train machine learning models with tools for all skill levels. Deploy and manage models at scale. Enabled ✓	 Watson OpenScale IBM Infuse your AI with trust and transparency. Understand how your AI models make decisions to detect and mitigate bias. Enabled ✓	 Watson Studio IBM Unleash the power of your data. Build custom models and infuse your business with AI and machine learning. Enabled ✓
--	---	---

Analytics 1 item

 Data Refinery IBM Simplify the process of preparing large amounts of raw data for analysis. Enabled ✓
--

IBM Cloud Pak for Data

Instances / Select a service /

Select a service

Select the service for which you want to create a new instance.

Dashboards 1 item

 Cognos Dashboards IBM Identify patterns in your data with sophisticated visualizations. No coding needed. Enabled ✓
--

Data governance 1 item

 Watson Knowledge Catalog IBM Organize and govern data. Automatically discover, classify, profile, and protect your data so data scientists can find trusted data fast. Enabled ✓

6) For the *Db2 instance*, and select **Manage Access**.

IBM Cloud Pak for Data

Instances

Filter by: Type ▾ Status ▾

Find instances

Name	Type	Created by	vCPU requests	Memory requests (GiB)	Users	Status	Created on
data-virtualization	dv	admin	11.50	38.50 Gi	1	Green	13 Sept 2022
data-management-console	dmc	admin	5.20	21.40 Gi	1	Green	8 Sept 2022
Db2-1	db2oltp	admin	0.60	4.25 Gi	1	Green	8 Sept 2022
openscale-defaultinstance	aios	admin	-	-	1	Green	7 Sept 2022

Open

Manage access

Delete

7) Review the users with Access to DB2.

Access management: Db2-1

Filter by: All roles ▾

Find users

Name	Username	Service role
<input type="checkbox"/> admin	admin	Admin ▾

Add users +

- 8) You may add new users by clicking on **Add User+** button for them to have access to manage DB2 Services.

Grant access to users and user groups

Specify the users who can access Db2 and the role of each user.

Find users and user groups

Users and user groups	Filter by:
<input checked="" type="checkbox"/> admin admin ..	Admin ▾
<input checked="" type="checkbox"/> SV Sandeep Ved sved, sandeep.ved1@ibm.com	Admin ▾

Cancel Add

Access management: Db2-1

Filter by: All roles ▾

Find users

Name	Username	Service role
<input type="checkbox"/> admin	admin	Admin ▾
<input type="checkbox"/> SV	Sandeep Ved	sved Admin ▾

Add users +

PostgreSQL

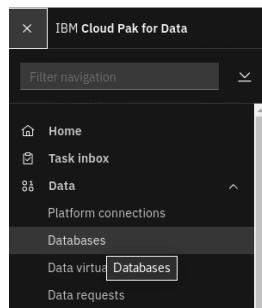
- 9) A external PostgreSQL DB database has been already prepared for the lab with the following connection details.
- Database Name: vlyawtap
 - Host Name: tiny.db.elephantsql.com
 - Port: 5432
 - User Name: vlyawtap
 - Password: fXt4TyCB_W0d0LCaCaPF7MbLKWlpti60

Db2 data loading

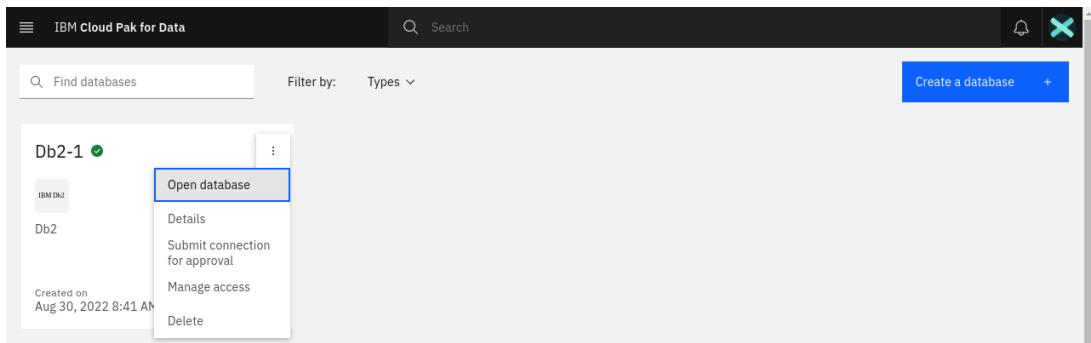
The data sets you will use in this tutorial are available in the Practicum Github. You need to load those data sets into Db2.

1. Access these two data sets in the GitHub, and download them to your local machine:
 - **Customers data set -> customers.csv** (To upload data in DB2 Database)
 - **Sales data set -> sales.csv** (For Reference only, as its already loaded in PostgreSQL DB)

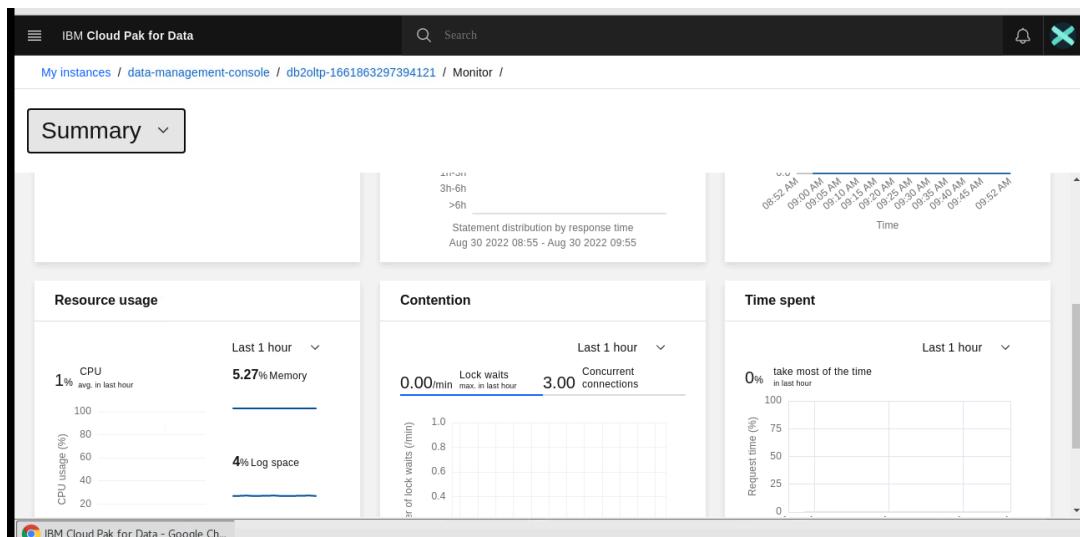
2. Navigate to Data -> Databases.



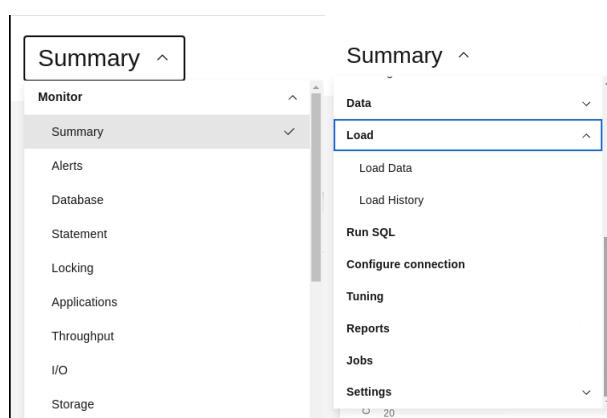
3. Click on Open Database



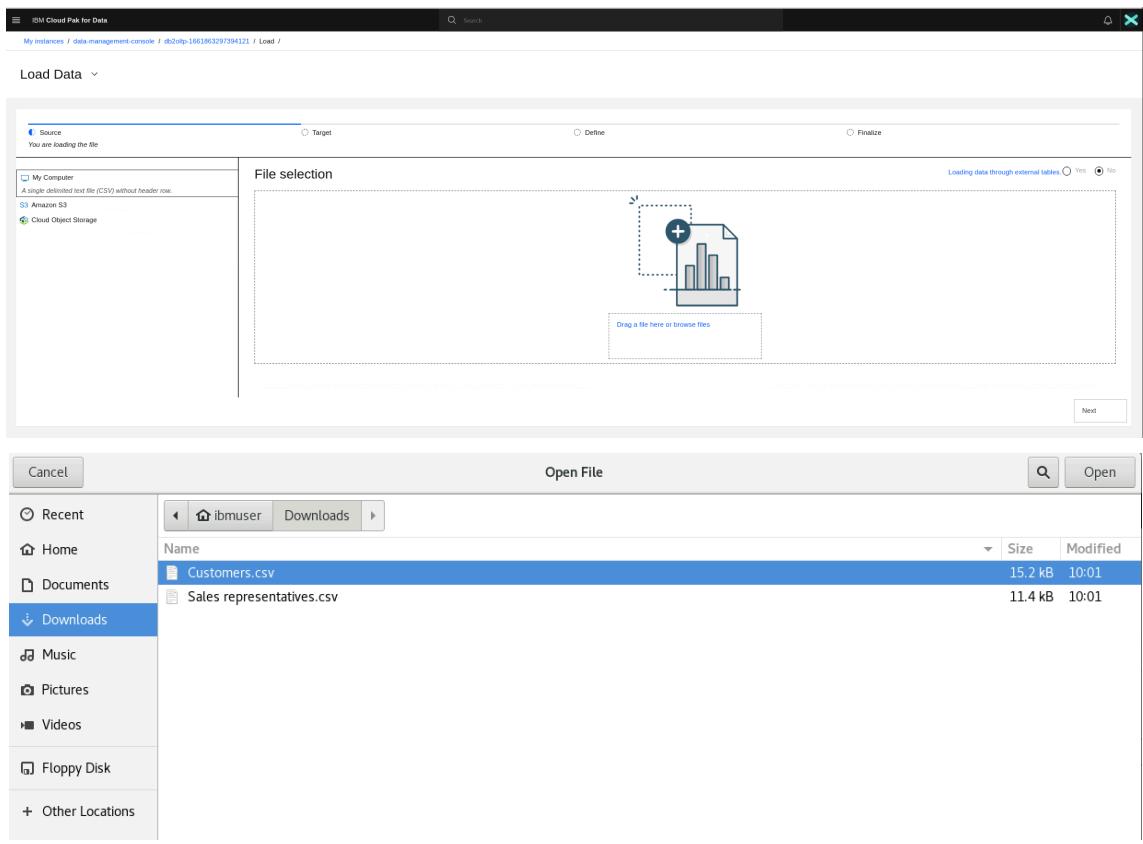
4. You will see the DB2 Database Summary. Click on the Summary Drop Down.



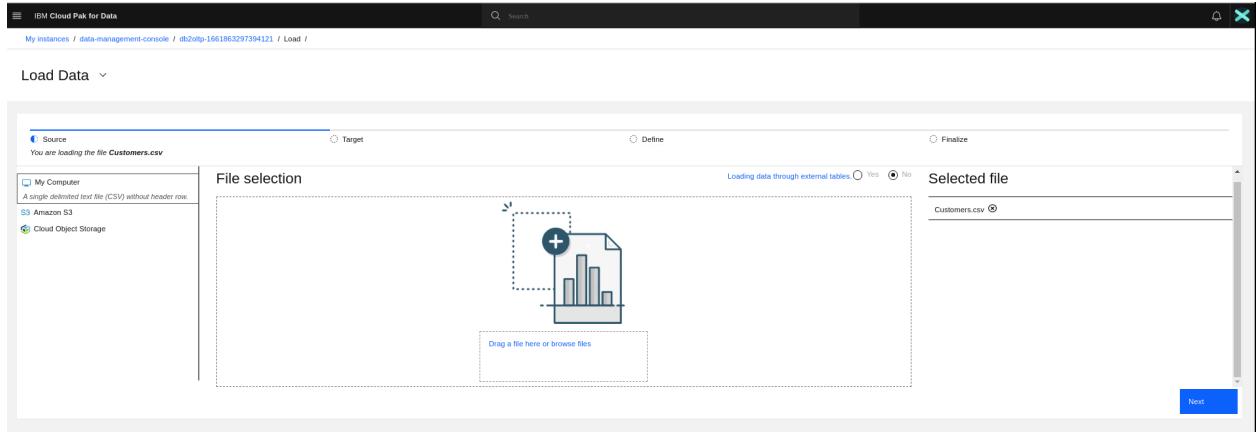
5. Under the Summary Drop down, click on Load -> Load Data.



6. The DB2 Load Data Page will appear. Select or Drag the **customers.csv** file from your local machine into the *File Selection* panel in the *Db2* console.



7. Click **Next**.



8. Click **New Schema+**.

IBM Cloud Pak for Data

My instances / data-management-console / db2oltp-1662568821240436 / Load /

Load Data ▾

Source Target Define Finalize

You are loading the file **Customers.csv**

Select a load target

Schema

Find schemas

TUTORIAL
aiopenscale00

Refresh Back Next

This screenshot shows the 'Load Data' interface in IBM Cloud Pak for Data. The 'Target' tab is selected. A message at the top says 'You are loading the file Customers.csv'. Below it, a 'Select a load target' section shows a 'Schema' dropdown with 'TUTORIAL' and 'aiopenscale00' listed. A 'New schema +' button is also present. At the bottom are 'Back' and 'Next' buttons.

9. Enter a suitable name, eg. IBMADMIN. Click **Create**.

IBM Cloud Pak for Data

My instances / data-management-console / db2oltp-1662568821240436 / Load /

Load Data ▾

Source Target Define Finalize

You are loading the file **Customers.csv**

Select a load target

Schema

Find schemas

TUTORIAL
aiopenscale00

Create a new schema

IBADMIN

Create

Refresh Back Next

This screenshot shows a 'Create a new schema' dialog box overlaid on the main interface. It contains a single input field with 'IBADMIN' typed into it and a 'Create' button below it. The main interface background is visible through the dialog.

10. Click **New table +**.

IBM Cloud Pak for Data

My instances / data-management-console / db2oltp-1662568821240436 / Load /

Load Data ▾

Source Target Define Finalize

You are loading the file **Customers.csv**

Select a load target

Schema

Find schemas

IBADMIN
TUTORIAL
aiopenscale00

Table

Find tables in IBADMIN

No entries found

New table +

Refresh Back Next

This screenshot shows the 'Table' section of the load interface. The 'IBADMIN' schema is selected in the schema dropdown. The table list area shows 'No entries found'. A 'New table +' button is located at the top right of this section. The main interface background is visible.

11.Type **customers** for the table name, and click **Create**.

The screenshot shows the 'Load Data' interface in IBM Cloud Pak for Data. On the left, under 'Schema', 'IBADMIN' is selected. In the center, under 'Table', there is a search bar with 'Find tables in IBMADMIN' and a result list with 'No entries found.' On the right, a 'Create a new table' panel has 'Customers' entered in the input field and a blue 'Create' button. At the bottom right of the main area are 'Back' and 'Next' buttons.

12.Click **Next** to continue.

The screenshot shows the 'Load Data' interface. The 'Source' tab is selected, showing 'IBADMIN' as the schema and 'Customers.csv' as the file. The 'Target' tab is selected, showing 'IBADMIN.CUSTOMERS' as the table. The 'Define' and 'Finalize' tabs are also visible. At the bottom right are 'Back' and 'Next' buttons, with 'Next' being highlighted.

13.Review the data to be loaded, and click **Next**.

The screenshot shows the 'Load Data' interface with the 'Target' tab selected. It displays the data from 'Customers.csv' into 'IBADMIN.CUSTOMERS'. The data is presented in a table format with columns: CUST_ID, CUSTNAME, CITY, STATE, COUNTRY_CODE, POSTAL_CODE, EMAIL_ADDRESS, PHONE_NUMBER, and YTD_SALES. The 'Next' button at the bottom right is highlighted.

CUST_ID	CUSTNAME	CITY	STATE	COUNTRY_CODE	POSTAL_CODE	EMAIL_ADDRESS	PHONE_NUMBER	YTD_SALES
1	Michael Golden	Abbadia Alpina	TO	IT	10060	Michael.E.golden@spambob.com	724-454-8453	90.3
2	Renee Mullins	Columbus	OH	US	45101	Rene.K.mullins@dodgeit.com	229-990-2162	0
3	Allen Schmidt	Abetao	PG	IT	6040	Allen.M.schmidt@spambob.com	289-202-8653	304
4	Robert May	Houston	TX	US	79601	Robert.C.may@spambob.com	630-492-6535	304
5	Joe Cruz	Harrisonburg	VA	US	24210	Joe.D.cruz@spambob.com	448-788-1089	180.3
6	Rebecca White	Acate	RG	IT	97011	Rebecca.white@spambob.com	828-268-4303	52
7	Gary Neal	Chicago	IL	US	61701	Gary.N.neal@dodgeit.com	267-929-9454	673
8	Steve Huynh	Detroit	MI	US	48614	Steve.A.huynh@dodgeit.com	644-101-7549	904.86
9	Anthony Johnson	Achstetten	DE		86480	Anthony.H.johnson@pookmail.com	907-111-5490	354
10	Alberto Fabian	Acquabona	BL	IT	32043	Alberto.T.fabian@pookmail.com	113-785-7672	352

14. Review the summary, and click **Begin Load**.

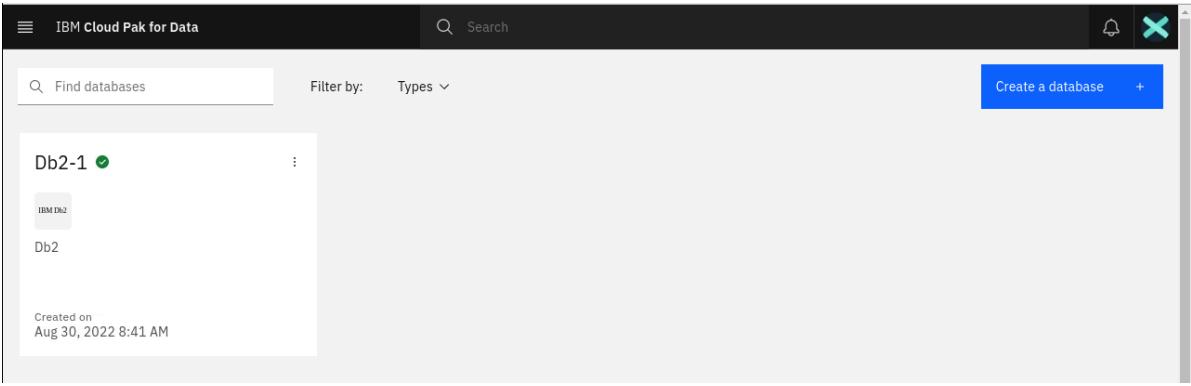
15. Verify that the load is successful. Click **View Table**.

16. Table details will appear.

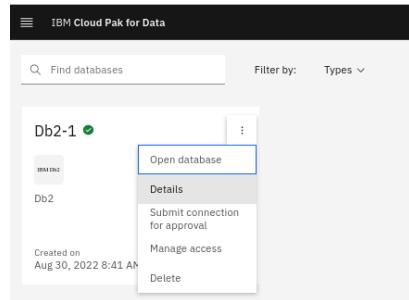
17. Repeat steps 5-10 to load the **sales.csv** data set into the same schema with the table name **sales**. *<not needed as same table is uploaded and available from PostgreSQL DB>*

Check DB2 Connection Details

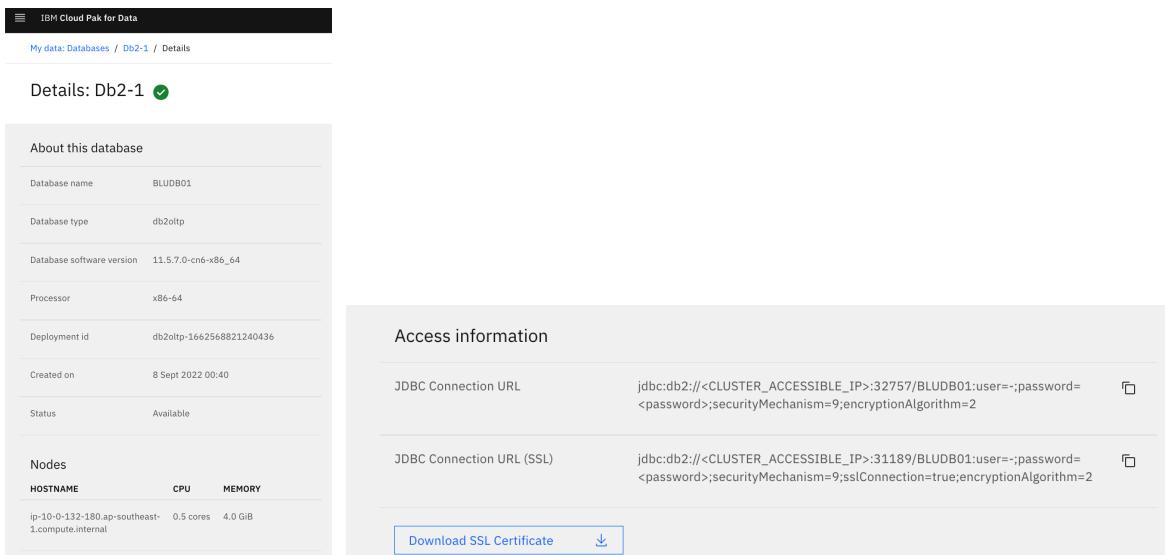
18. Navigate back to **Data -> Databases** from the main navigation menu.



19.Click on Details .



20.Note down the connection details for this DB instance like hostname, Database Name, Port (non SSL & SSL). You can use the admin user id credential to connect to the database later.



Nodes		
HOSTNAME	CPU	MEMORY
ip-10-0-132-180.ap-southeast-1.compute.internal	0.5 cores	4.0 GiB

21.Again open the DB2 Database and Open the Summary Menu and select Run SQL.

The screenshot shows the 'Authorization' section of the Data Management Console. On the left, a sidebar lists various database objects like Tables, Views, Indexes, etc. The 'Authorization' section is expanded, showing a list of users and groups with their roles. At the top, there are buttons for 'Add role', 'Remove role', 'Grant multiples', and 'Revoke multiples'. A search bar at the top right allows filtering by name.

22. You can use the SQL interface to update the DB2 database as usual.

The screenshot shows the 'Run SQL' interface. On the left, a sidebar shows 'Data objects' with a list of system tables like IBMADMIN, IBMCONSOLE, etc. The main area is titled 'Untitled - 1' and contains a SQL editor with the following code:

```
grant dbadm on database to user sved;
grant secadm on database to user sved;
```

23. For example you can run below commands to grant full access to the database to newly created users.

The screenshot shows the 'Run SQL' interface after the commands have been run. The SQL editor now displays the results of the executed commands. Below the editor, a 'History' panel shows the script and its execution details:

Script	Date	Status	Runtime
db2oltp-1662568821240436: Untitled - 1	Sep 16, 2022 3:18:05 PM	✓ 2	0.012 s
grant dbadm on database to user sved		✓	0.006 s
grant secadm on database to user sved		✓	0.006 s

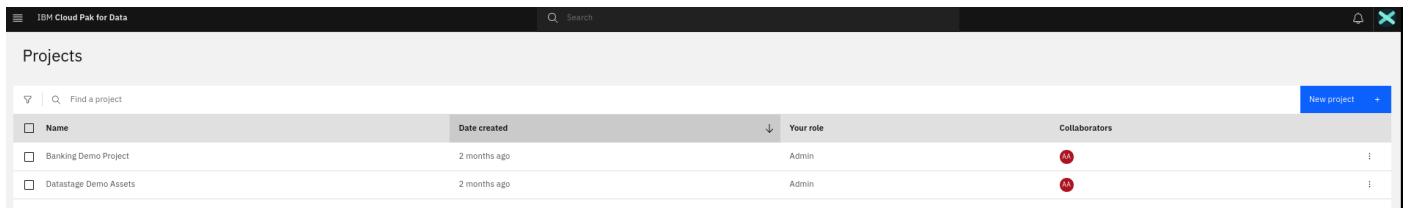
Setting up Data Virtualization

You can use **Data Virtualization** to create a virtual table to segment or combine data from one or more tables. **Data Virtualization** connects multiple data sources into a single self-balancing collection of data sources or databases.

Task 1: Create a project

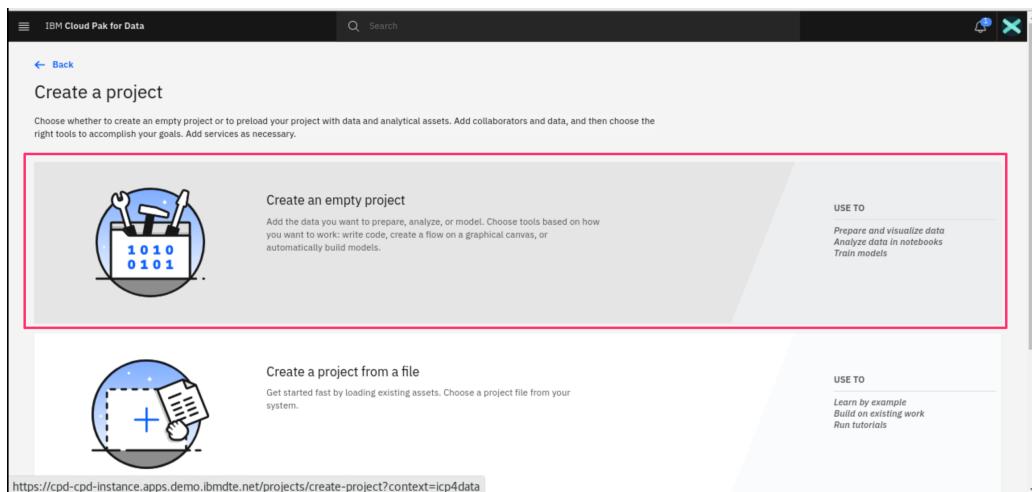
You need a project to store the virtualized data.

1. Login to CPD Platform. Click on the Hamburger Icon on the top left.
2. Select **View all Projects**. Alternatively All Projects can be accessed from Home Page under Overview Section.
3. If you have an existing project and want to reuse it, open it. If you don't have an existing project or want to use fresh new project, click **New project** on your **Projects** page.



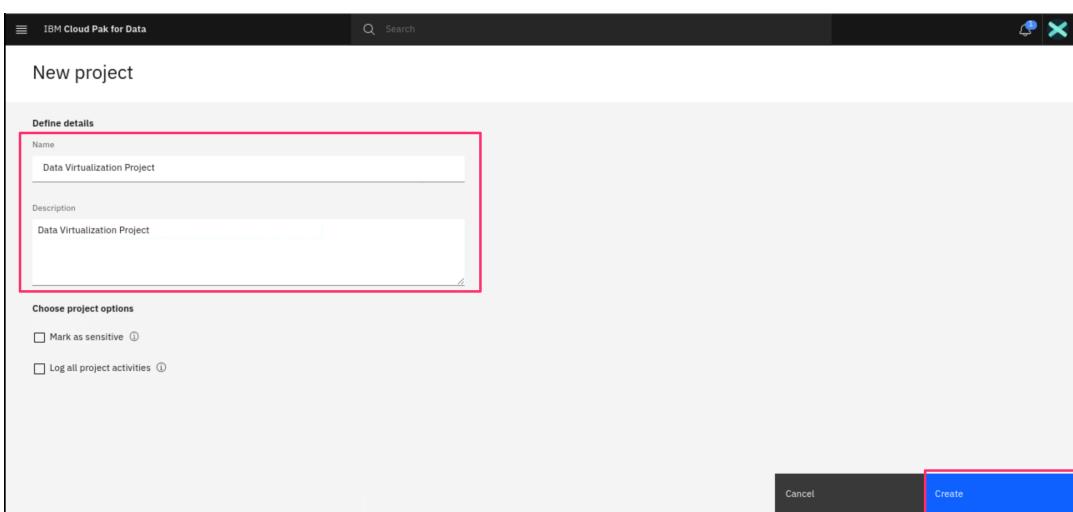
The screenshot shows the 'Projects' section of the IBM Cloud Pak for Data interface. It lists two projects: 'Banking Demo Project' and 'Datastage Demo Assets'. Both projects were created 2 months ago and are assigned to the 'Admin' role. There are two collaborator icons next to each project entry.

4. Select **Create an empty project**.



The screenshot shows the 'Create a project' screen. It has two main sections: 'Create a project' (selected) and 'Create a project from a file'. The 'Create a project' section includes a sub-section for 'Create an empty project' with a description and a 'USE TO' section listing 'Prepare and visualize data', 'Analyze data in notebooks', and 'Train models'. A red box highlights the 'Create a project' section.

5. On the **Create a project** screen, add a name and optional description for the project. Click **Create**.



The screenshot shows the 'New project' creation form. It has sections for 'Define details' (Name: 'Data Virtualization Project', Description: 'Data Virtualization Project') and 'Choose project options' (checkboxes for 'Mark as sensitive' and 'Log all project activities'). A red box highlights the 'Define details' section. The 'Create' button at the bottom right is also highlighted with a red box.

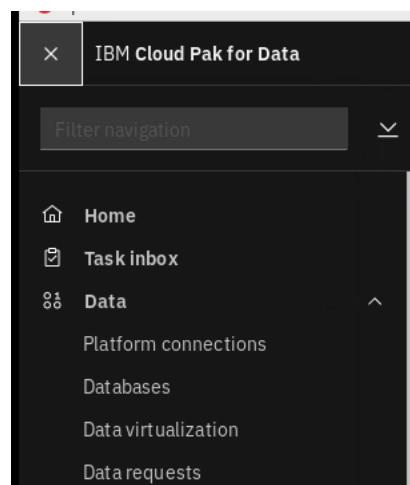
6. The Project will be created successfully.

The screenshot shows the 'IBM Cloud Pak for Data' interface. At the top, there's a navigation bar with 'IBM Cloud Pak for Data' and a search bar. Below it, a 'Projects / Data Virtualization Project' section has tabs for 'Overview', 'Assets', 'Jobs', and 'Manage'. The 'Overview' tab is selected. It displays sections for 'Assets' (with a note about creating assets), 'Project size' (0 Bytes), and 'Project history' (No notifications). There's also a 'Readme' section where users can type project notes.

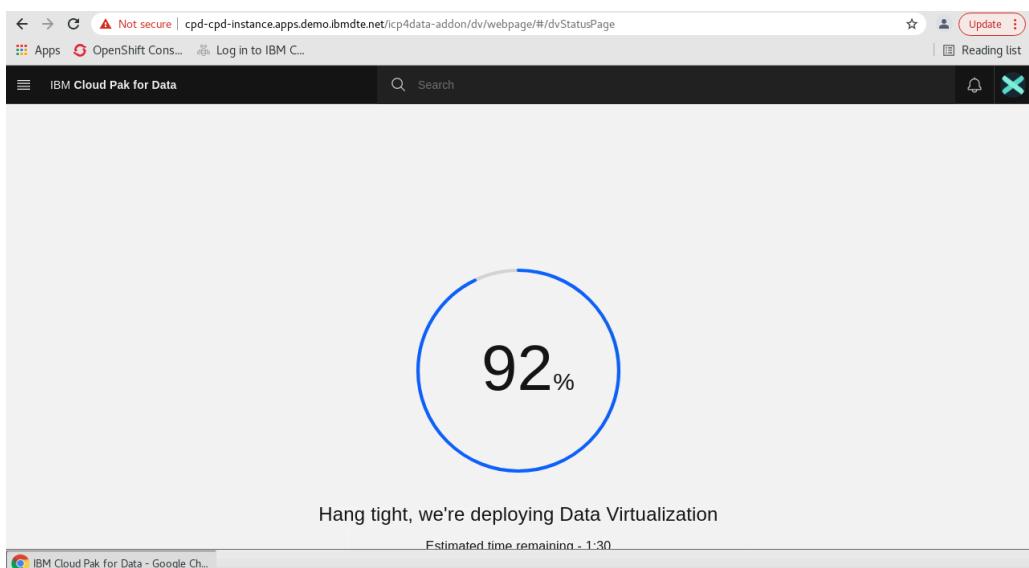
Task 2: Add a connection to your Db2 data source

Before you can virtualize the data, you need to create a connection to the data source.

1. From the main navigation menu, select **Data > Data virtualization**.



2. If opening for the first time, the Data Virtualization component may start to get initialized. Just wait until its fully deployed and Next screen appears.



3. After initial initialization, the Data Virtualization Data Sources Page appears.

Connections: 0 | Remote connectors: 0

Find by Name, Endpoint, Type

Name	Endpoint	Type	Status	Listed tables
?				

IBM Cloud Pak for Data - Google Ch...

4. Click **Add connection > New connection**.

Connections: 0 | Remote connectors: 0

Find by Name, Endpoint, Type

Name	Endpoint	Type	Status	Listed tables
?				

You don't have any data sources yet
Add a data source by creating a new connection or
select from an existing platform connection or by
searching for a data source via a remote connector

5. Select **Db2**. Click **Select**.

New connection

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

Provider: IBM

db2

All connection types

- Db2 Event Store
- Db2 for z/OS
- Db2 on Cloud
- Db2 Warehouse
- Db2 for i
- Db2 Hosted
- Db2 Big SQL

Selected connection type

Db2

Details

IBM Db2 database

Compatible services

Catalogs

DataStage

Data Virtualization

Data discovery and Data quality

Metadata import (Lineage)

Metadata import (Discovery)

Watson Studio

Select

6. Complete the connection details based on the credentials you copied for your db2 instance .

Create connection: Db2

Enter the connection information.

Connection overview

Connection details

Credentials

Certificates

Connection overview

Name: CP4D DB2 Connection

Description: CP4D DB2 Connection for Customers Data

The details for this type of connection cannot be validated until you use the connection. Ensure that the credentials and details are correct. If you encounter an issue, you can edit the connection later.

Connection details

Additional properties: ⓘ

Cancel Back Create

7. Enter Database Name, Host Name and Non SSL Port as we noted earlier. Select the Credential Setting as **Personal**.

Create connection: Db2

Enter the connection information.

Connection overview

Connection details

Credentials

Certificates

Database*: ⓘ Customer

Hostname or IP address*: ⓘ worker1

Port*: ⓘ 30812

Credentials

Credential setting: ⓘ

Personal Shared

Cloud Pak for Data authentication ⓘ

Cancel Back Create

8. Enter **Personal** Credential of **admin** User. Click **Create**.

Create connection: Db2

Enter the connection information.

Connection overview

Connection details

Credentials

Certificates

Personal Shared

Cloud Pak for Data authentication ⓘ

Username*: ⓘ admin

Password*: ⓘ

Username and Password Security Mechanism ⓘ

Default

Username and Password Encryption Algorithm ⓘ

Default

Certificates

Port is SSL-enabled ⓘ

Cancel Back Create

9. Click **Skip**

IBM Cloud Pak for Data

My instances / data-management-console / dv-1657090051946032 /

Add to a remote connector (optional)

You can add connections to remote connectors to enhance parallelism during processing and improve query performance.

Find Search

Remote connector Hostname Description Port Username



You haven't set up any remote connectors yet.
You must set up a remote connector to assign data sources to the connector. You must have the required permissions to set up remote connectors.

Set up remote connector +

Skip Add to connector Refresh

10.The DB2 Data Source will be added as a Data Virtualization Source.

The screenshot shows the 'Data sources' page in the IBM Cloud Pak for Data interface. At the top, there are buttons for 'Set up remote connector' and 'Add connection'. Below this, a feedback survey is displayed. The main area shows a table of data sources:

Name	Endpoint	Type	Status	Listed tables
CP4D DB2 Connection	worker1: 30812	Db2 Family	Active	318 / 318
Customer	30812	admin	-	CP4D DB2 Connection for Customers Data

Task 3: Add a connection to a PostgreSQL data source

For the purpose of Validating the Data Virtualization from Multiple Types of Data Sources, We have prepared a **PostgreSQL DB** as well with the *Sales.csv* data uploaded in table named *Sales_Rep*. Use below steps and given credentials to use the data from this PostgreSQL Data Source.

1. From the main navigation menu , select **Data > Platform Connections**. The list of configured *Data sources* displays.
2. Click **New connection**.

The screenshot shows the 'Data sources' page with the 'Add connection' button highlighted in blue. The rest of the interface is identical to the previous screenshot, showing the list of configured data sources.

3. Select **PostgreSQL** and click **Select**.

New connection

Supported connection types

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

Provider

- IBM
- Third-party
- User-defined

All connection types

- Amazon RDS for PostgreSQL
- Databases for PostgreSQL
- PostgreSQL

Cancel Select

4. Enter the Connection Name and Description.

Create connection: PostgreSQL

Enter the connection information.

Connection overview

Connection details

Credentials

Certificates

Connection overview

Show more

Name

Description

Cancel Back Create

5. Enter the Connection Details and Credentials as below. Click **Next**.

- Database Name: vlyawtap
- Host Name: tiny.db.elephantsql.com
- Port: 5432
- Credential: Personal
- User Name: vlyawtap
- Password: fXt4TyCB_W0d0LCaCaPF7MbLKWLpti60

Create connection: PostgreSQL

Enter the connection information.

Connection overview

Connection details

Credentials

Certificates

Database*

Hostname or IP address*

Port*

Credentials

Credential setting Personal Shared

Username*

Password*

Certificates

Port is SSL-enabled

Cancel Back Create

6. Click **Skip**.

Add to a remote connector (optional)

[Skip](#) [Add to connector](#)

You can add connections to remote connectors to enhance parallelism during processing and improve query performance.

Q Find Refresh ⌂

Remote connector	Hostname	Description	Port	Username
				

You haven't set up any remote connectors yet.

7. The new data source will be added.

IBM Cloud Pak for Data Search Set up remote connector Add connection

Data sources

Table view Constellation view

Connections: 2 | Remote connectors: 0

Name	Endpoint	Type	Status	Listed tables
PostgreSQL db Elephantsql	tiny.db.elephantsql.com: 5432	PostgreSQL	Active	2 / 2
CP4D DB2 Connection	worker1: 30812	Db2 Family	Active	318 / 318

Task 4: Add tables to your virtualized data

With the connection defined, you can virtualize data from that data source.

- From the *Data sources* menu, select **Virtualization > Virtualize**.

IBM Cloud Pak for Data

My instances / data-management-console / dv-1657090051946032 / Virtualization /

Data sources ^

Virtualization

- Data sources
- Virtualize
- Virtualized data
- Cache management

Monitor

- Endpoint
- tiny.db.elephantsql.com: 5432

Data

- Run SQL
- Administration
- worker1: 30812

- Click on **Add Virtual Objects**.

IBM Cloud Pak for Data Search

Virtualized data

Find virtual objects

Filter by: All types

Table	Schema name	Created on	Statistics last collected on
			

You don't have any virtualized objects yet.

Add connections to your data sources, then you can create virtual tables and views. [Learn more](#)

Add virtual objects

- You will see the available connections from where data can be virtualized.

IBM Cloud Pak for Data

My instances / data-management-console / dv-1657090051946032 / Virtualization /

Virtualize

Tables Files

Explore List

Connections /

Connections

PostgreSQL on ElephantSQL...>

CP4D DB2 Connection (318) >

You haven't made a selection yet.

Select a connection to browse schemas and find tables that you want to virtualize.

4. Select CP4D DB2 Connection -> IBMADMIN -> CUSTOMERS Table and click **Add to cart**.

IBM Cloud Pak for Data

My instances / data-management-console / dv-1657090051946032 / Virtualization /

Virtualize

Tables Files

Explore List

Connections / CP4D DB2 Connection / IBMADMIN /

< □ CP4D DB2 Connection (1)

I selected Add to cart

□ IBMADMIN (1)

□ CUSTOMERS

You haven't made a selection yet.

Select the tables that you want to virtualize and add them to the cart. Click a table to preview the table contents or click the checkbox to add it to the cart without previewing it.

5. Click on **Connections** link again.

IBM Cloud Pak for Data

My instances / data-management-console / dv-1657090051946032 / Virtualization /

Virtualize

Tables Files

Explore List

Connections / CP4D DB2 Connection / IBMADMIN /

< □ CP4D DB2 Connection (1)

Find table

□ IBMADMIN (1)

□ CUSTOMERS

You haven't made a selection yet.

Select the tables that you want to virtualize and add them to the cart. Click a table to preview the table contents or click the checkbox to add it to the cart without previewing it.

6. Select **PostgreSQL on ElephantSQL** -> **sample** -> **sales_rep** tables from the list, and click **Add to cart**.

IBM Cloud Pak for Data

My instances / data-management-console / dv-1657090051946032 / Virtualization /

Virtualize

Tables Files

Explore List

Connections / PostgreSQL on ElephantSQL / sample /

< □ sample (1)

I selected Add to cart

□ SALES_REP

You haven't made a selection yet.

Select the tables that you want to virtualize and add them to the cart. Click a table to preview the table contents or click the checkbox to add it to the cart without previewing it.

7. Click **View cart**.

The screenshot shows the 'Virtualize' step in the data management console. A sidebar on the left lists 'Tables' and 'Files'. Below it, under 'Connections / PostgreSQL ob ElephantSQL / sample /', there's a list of tables: 'sample (1)' and 'SALES_REP'. A message says 'You haven't made a selection yet.' with instructions to select tables and add them to the cart.

8. Select Assign to **Virtualized data** to add these two tables to your list of virtualized data.
Click **Virtualize**.

The screenshot shows the 'Review cart and virtualize tables' step. It displays the selected objects in the cart: 'CUSTOMERS' and 'SALES_REP'. The 'Assign to' dropdown is set to 'Virtualized data'. The 'Publish to' dropdown is set to 'Catalog'. The 'Virtualize' button is visible at the top right.

9. Click **Confirm**.

The screenshot shows a confirmation dialog box titled 'Confirm virtualization'. It contains a message: 'If you click Confirm and start virtualizing, you can't cancel the request or leave this screen until the virtualization is complete. If you click Cancel, your object will not be virtualized.' There is a checkbox 'Do not show this message again' and two buttons: 'Cancel' and 'Confirm'.

10. Click **Go to virtualized data**.

The screenshot shows the 'Virtualize objects' confirmation screen. It lists the virtualized objects: 'CUSTOMERS' and 'SALES_REP', both with 'Success' status in both 'Virtualization status' and 'Publish status'. The estimated virtualization time is 2 seconds. At the bottom, there are buttons for 'Virtualize more data' and 'Go to virtualized data'.

11. You can see the **virtualized data**.

IBM Cloud Pak for Data

Search

My instances / data-management-console / dv-1657090051946032 / Virtualization /

Virtualized data ▾

Find virtual objects

1 item selected		Manage access		Join	Assign	Publish to catalog	Cancel
Table	Schema name	Created on	Statistics last collected on				
<input type="checkbox"/> SALES_REP	ADMIN	Sep 2, 2022 11:21 AM	Not collected				
<input type="checkbox"/> CUSTOMERS	ADMIN	Sep 2, 2022 11:21 AM	Not collected				

Task 5: Join and Publish virtualized data to project

Next, join two tables to create a virtualized asset and publish that to a catalog and project.

1. On the *Virtualized data* screen, select the **customers** and **sales_rep** tables from the list, and click **Join**.

IBM Cloud Pak for Data

Search

My instances / data-management-console / dv-1657090051946032 / Virtualization /

Virtualized data ▾

Find virtual objects

2 items selected		Manage access		Join	Assign	Publish to catalog	Cancel
Table	Schema name	Created on	Statistics last collected on				
<input checked="" type="checkbox"/> SALES_REP	ADMIN	Sep 2, 2022 11:21 AM	Not collected				
<input checked="" type="checkbox"/> CUSTOMERS	ADMIN	Sep 2, 2022 11:21 AM	Not collected				

2. Connect the **SALESREP_ID** columns in the two tables. Click **Preview**.

IBM Cloud Pak for Data

Search

My instances / data-management-console / dv-1657090051946032 /

Join virtual objects

Click and drag from one table row to the other to create a join key.

Table 1: SALES_REP

Column name	Data type
<input checked="" type="checkbox"/> SALESREP_ID	CLOB
<input checked="" type="checkbox"/> FIRST_NAME	CLOB
<input checked="" type="checkbox"/> LAST_NAME	CLOB
<input checked="" type="checkbox"/> NATIONALITY	CLOB
<input checked="" type="checkbox"/> NATIONAL_ID	CLOB
<input checked="" type="checkbox"/> PHONE_NUMBER	CLOB
<input checked="" type="checkbox"/> AGE	CLOB
<input checked="" type="checkbox"/> SEX	CLOB
<input checked="" type="checkbox"/> TERRITORY	CLOB

Table 2: CUSTOMERS

Column name	Data type
<input checked="" type="checkbox"/> STATE	VARCHAR
<input checked="" type="checkbox"/> COUNTRY_CODE	VARCHAR
<input checked="" type="checkbox"/> POSTAL_CODE	VARCHAR
<input checked="" type="checkbox"/> EMAIL_ADDRESS	VARCHAR
<input checked="" type="checkbox"/> PHONE_NUMBER	VARCHAR
<input checked="" type="checkbox"/> YTD_SALES	DECIMAL
<input checked="" type="checkbox"/> SALESREP_ID	CLOB
<input checked="" type="checkbox"/> NATIONALITY	VARCHAR
<input checked="" type="checkbox"/> NATIONAL_ID	VARCHAR

Open in SQL editor [🔗](#)

Join keys Filters

After you select at least two columns of different data types, click **Preview** to ensure that the columns were properly joined. Preview can take a while if you are joining large tables. Click **Next** to continue joining these tables.

SALES REP	CUSTOMERS
SALESREP_ID	SALESREP_ID

3. Close the Preview Popup and Click **Next**.

Table 1: SALES_REP

New join preview

SALESREP_ID FIRST_NAME LAST_NAME NATIONALITY NATIONAL_ID PHONE_NUMBER AGE SEX TERRITORY

NC160 Betsy Adams UK 124-168-918 229-990-2162 35 F NorthCentral

NC166 Penney Hayes FR 516-264-270 630-492-6535 50 F NorthCentral

NC169 Fredericka King ES 766-887-613 448-788-1089 88 F NorthCentral

NC169 Fredericka King ES 766-887-613 448-788-1089 88 F NorthCentral

NC232 Lakeesha Jenkins IT 910-870-499 828-268-4303 21 F NorthCentral

NC232 Lakeesha Jenkins IT 910-870-499 828-268-4303 21 F NorthCentral

AGE CLOB SALESREP_ID VARCHAR
SEX CLOB NATIONALITY VARCHAR
TERRITORY CLOB NATIONAL_ID VARCHAR

4. Review the joined table, and click **Next**.

SALESREP_ID	FIRST_NAME	LAST_NAME	NATIONALITY	NATIONAL_ID	PHONE_NUMBER	AGE
NC160	Betsy	Adams	UK	124-168-918	229-990-2162	35
NC166	Penney	Hayes	FR	516-264-270	630-492-6535	50
NC169	Fredericka	King	ES	766-887-613	448-788-1089	88
NC169	Fredericka	King	ES	766-887-613	448-788-1089	88
NC232	Lakeesha	Jenkins	IT	910-870-499	828-268-4303	21
NC232	Lakeesha	Jenkins	IT	910-870-499	828-268-4303	21

5. For the view name, type joined_customers_sales_table. Select a project from the list that was created earlier, eg. Data Virtualization Project. Click **Create view**.

Assign to (all tables will be assigned to the same project)

Data request Project Virtualized data

Publish to

Catalog

Data Virtualization Project

View name Schema name

joined_customers_sales_table ADMIN

6. When the process completes, Click **Go to Virtualized Data**.

Join view created

Assigned to: Data Virtualization Project [View](#)
Published to: Default Catalog's catalog [View](#)

View	Schema	View creation status	Assignment status	Publish status
joined_customers_sales_table	ADMIN	Success	Success	Success

[Virtualize more data](#) [Go to virtualized data](#)

7. Navigate to Projects -> Data Virtualization Project to view the project to preview the virtualized data.

IBM Cloud Pak for Data

Search

My Instances / data-management-console / dv-1657090051946032 / Virtualization /

Virtualized data

Find virtual objects

Filter by: All types

Table	Schema name	Created on	Statistics last collected on
joined_customers_sales_table	ADMIN	Sep 2, 2022 11:28 AM	Not applicable
SALES_REP	ADMIN	Sep 2, 2022 11:21 AM	Not collected
CUSTOMERS	ADMIN	Sep 2, 2022 11:21 AM	Not collected

Add virtual objects +

8. Click on the project created earlier.

IBM Cloud Pak for Data

Search

Projects

Find a project

Name	Date created	Your role	Collaborators
Data Virtualization Project	3 days ago	Admin	AA
Banking Demo Project	2 months ago	Admin	AA
Datastage Demo Assets	2 months ago	Admin	AA

New project +

9. Navigate to **Assets** Tab. You will see the virtual view and the connect object to the Data Virtualization Service.

10. Normally, You will need an Credential or API key to view the data in the project. This credential can be saved in the connection object so that its not prompted again and again. Click on the Connection object eg. DSXXXXXX.

IBM Cloud Pak for Data

Projects / Data Virtualization Project

Overview Assets Jobs Manage

Find assets

All assets

Name	Last modified
ADMIN.joined_customers_sales_table	4 minutes ago admin (You)
DS16621325392290910	4 minutes ago Connection

Asset types

- 1 Data access
- 1 Connection
- 3 Data
- 1 Data asset

New asset +

Data in this project

Drop data files here or browse for files to upload

11. Click on **Credentials**.

IBM Cloud Pak for Data

Search

Edit connection: Data Virtualization

Review the connection information.

Test connection

Connection overview

Connection details

Credentials

Certificates

Name
DS16621325392290910

Description
Connection description

Connection details

Database*

Cancel Save

12. Select one of the Authentication Method as the personal connection.

IBM Cloud Pak for Data

Search

Edit connection: Data Virtualization

Review the connection information.

Test connection

Connection overview

Credentials Personal Shared

Authentication method*

Certificates

Port is SSL-enabled

SSL certificate

-----BEGIN CERTIFICATE-----
MIIDBzCCEgAwIBAgIRAIyQHEkIpGPu/kYybPFWFUwDQYJKoZIhvNAQELBQA
HTEBMBkGA1UEAxMSemVuLWNhLNcnRpZmljYXRlMB4XDThMDYzMDIxMDQyN1oX
DTI1MDYyOTIxMDkyN1owHTEbBkGA1UEAxMSemVuLWNhLNcnRpZmljYXRlMIIB
TIANBekohkiG9w0RAOFAAOCAQ8AMITBcEKAQFAOTY5Nkiv3s9H66Jii/vCifTx

Cancel Save

13. Choose User Name and Password.

IBM Cloud Pak for Data

Search

Edit connection: Data Virtualization

Review the connection information.

Test connection

Connection overview

Credentials Personal Shared

Authentication method*

API key

Username and password

X.509 certificate

-----BEGIN CERTIFICATE-----
MIIDBzCCEgAwIBAgIRAIyQHEkIpGPu/kYybPFWFUwDQYJKoZIhvNAQELBQA
HTEBMBkGA1UEAxMSemVuLWNhLNcnRpZmljYXRlMB4XDThMDYzMDIxMDQyN1oX
DTI1MDYyOTIxMDkyN1owHTEbBkGA1UEAxMSemVuLWNhLNcnRpZmljYXRlMIIB
TIANBekohkiG9w0RAOFAAOCAQ8AMITBcEKAQFAOTY5Nkiv3s9H66Jii/vCifTx

Cancel Save

14. Enter the CP4D credentials and click on **Test Connection**. It should be successful. Click **Save** to save the credential information in the connection.

15. Now Click on the Virtualized Data Asset ADMIN.joined_customers_sales_table.

16. You should be able to preview the data without being prompted for credentials.

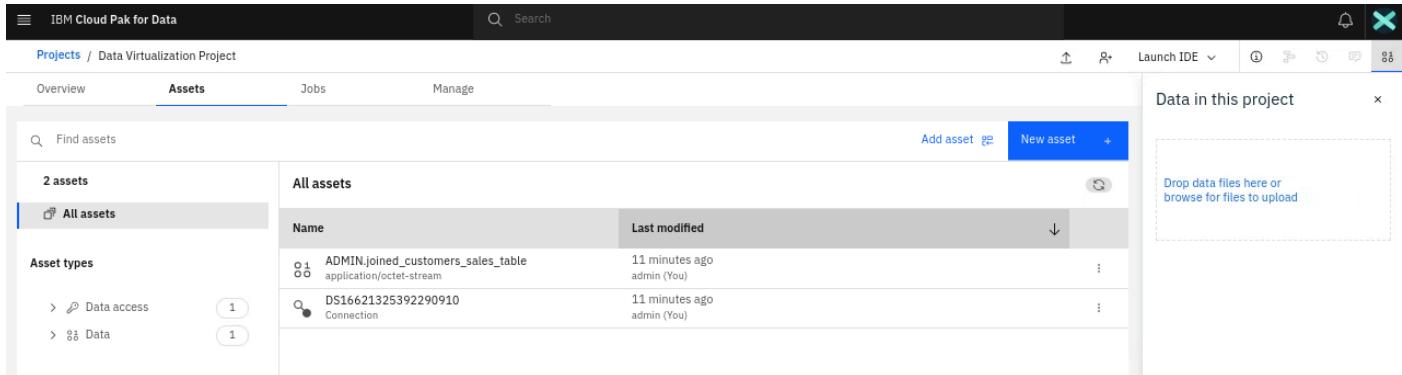
17. Optionally you can click on **Refine** button to explore or refine this data using **Watson Data Refinery**. The part of virtualized data can be saved as CSV file in the project using the Data Refinery Jobs.

Next steps

Now your virtual data is ready to be used. For example, you can these tasks:

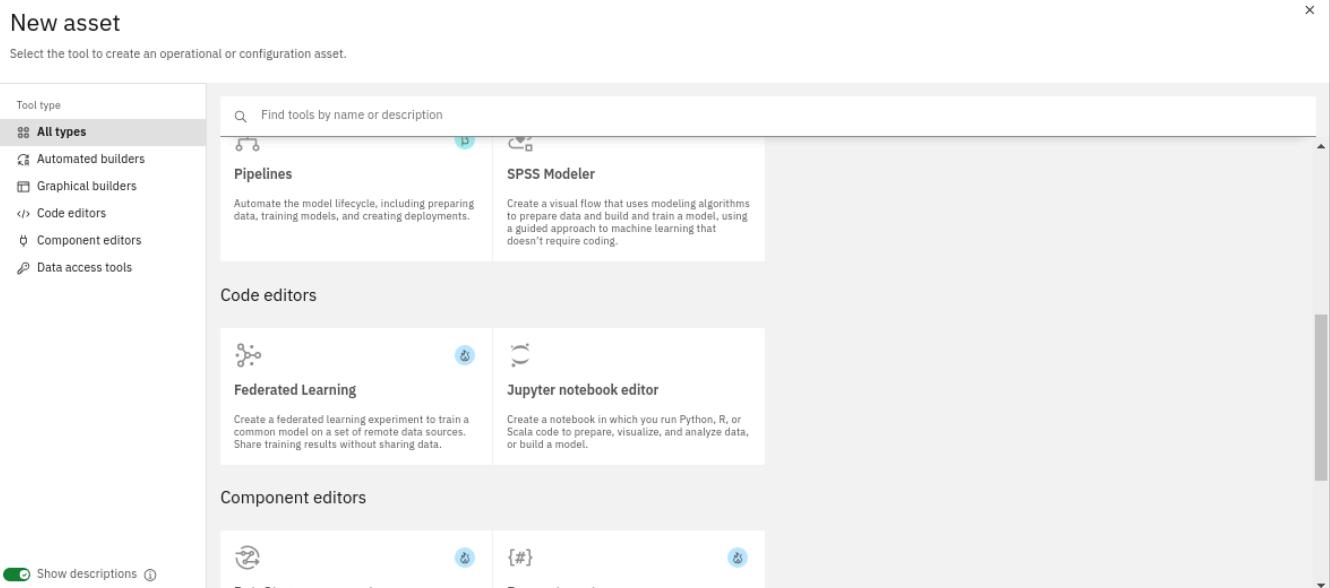
Analyze the data in a notebook

- 1) Under the Data Virtualization Project, In **Assets** Tab, Click on **New Asset +** to add a notebook Asset.



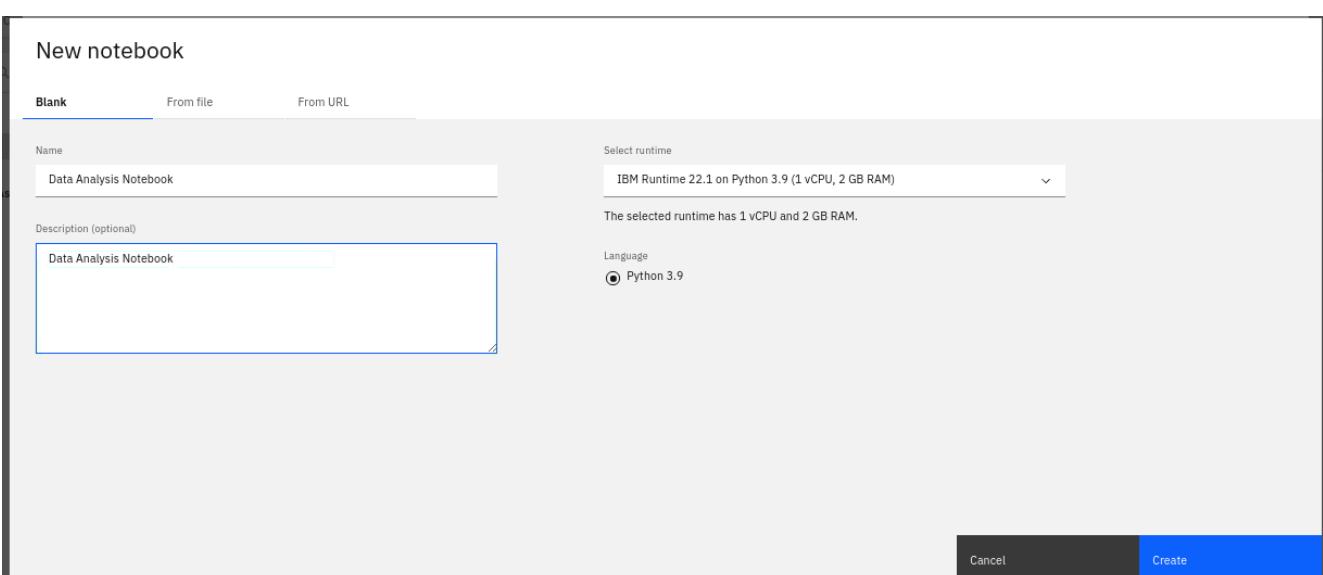
The screenshot shows the 'Assets' tab in the IBM Cloud Pak for Data interface. On the left, there's a sidebar with 'Asset types' and counts for 'Data access' (1) and 'Data' (1). The main area shows a table of assets with columns for 'Name' and 'Last modified'. Two assets are listed: 'ADMIN.joined_customers_sales_table' (modified 11 minutes ago) and 'DS16621325392290910' (modified 11 minutes ago). To the right, there's a dashed box labeled 'Drop data files here or browse for files to upload'.

- 2) Click on **Jupyter notebook Editor**.



The screenshot shows the 'New asset' creation dialog. Under 'Tool type', 'All types' is selected. Under 'Code editors', 'Jupyter notebook editor' is selected, showing its description: 'Create a notebook in which you run Python, R, or Scala code to prepare, visualize, and analyze data, or build a model.' Other options like 'Pipelines' and 'Federated Learning' are also shown. Under 'Component editors', 'DataStage component' and 'Parameter set' are listed. At the bottom, there's a 'Show descriptions' toggle and a 'Create' button.

- 3) Enter the Notebook Name and Credentials and Click **Create**.

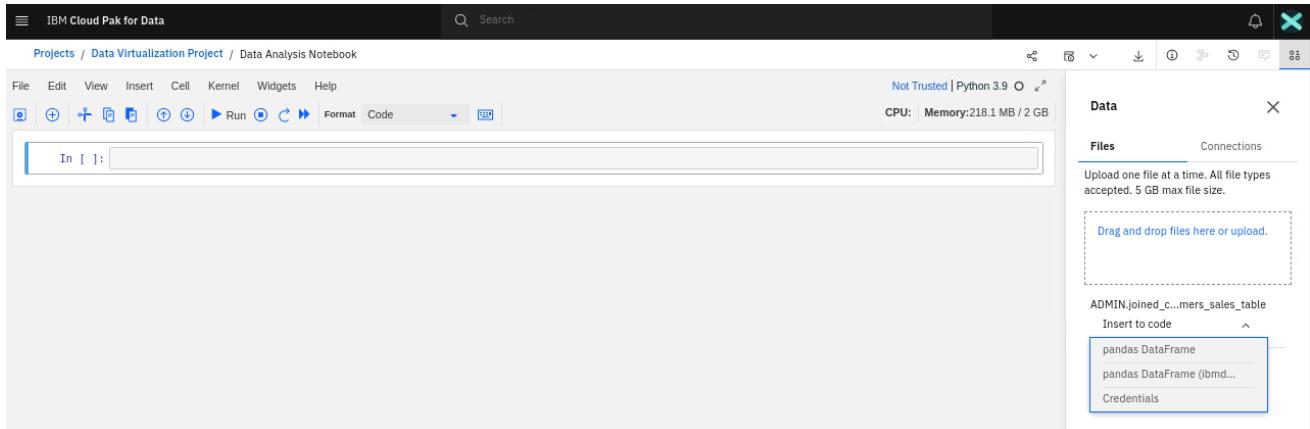


The screenshot shows the 'New notebook' creation dialog. It has tabs for 'Blank', 'From file', and 'From URL', with 'Blank' selected. The 'Name' field contains 'Data Analysis Notebook'. The 'Select runtime' dropdown shows 'IBM Runtime 22.1 on Python 3.9 (1 vCPU, 2 GB RAM)'. Below it, a note says 'The selected runtime has 1 vCPU and 2 GB RAM.' The 'Language' field is set to 'Python 3.9'. There's also a 'Description (optional)' field with 'Data Analysis Notebook' and a large text area below it. At the bottom, there are 'Cancel' and 'Create' buttons, with 'Create' being highlighted.

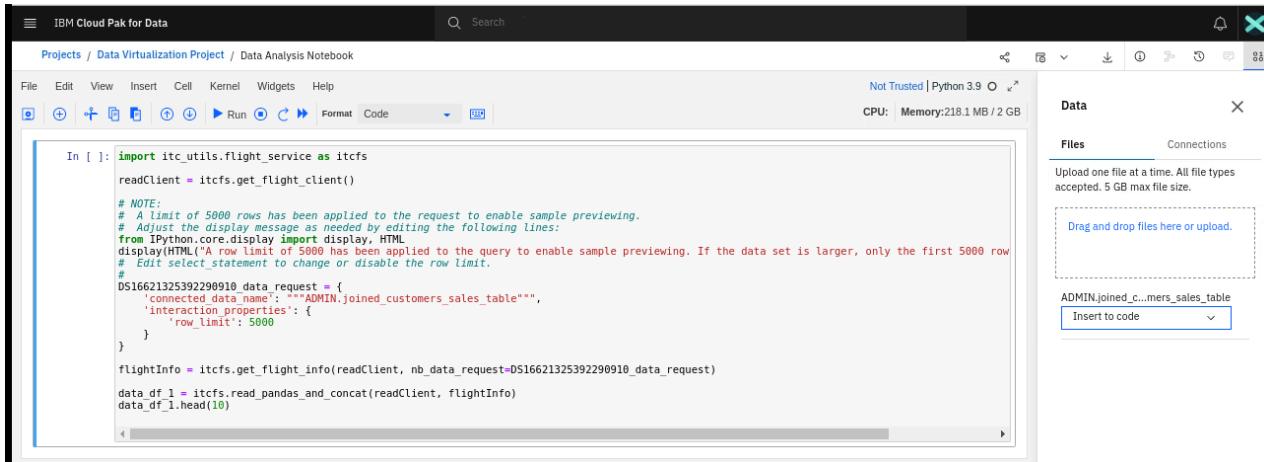
4) Click on  icon to View the Find and Add Data.



5) Under the Files tab, Click **Insert to Code -> pandas Data Frame** For the Virtual Table in the project.



6) The code will be added automatically. Click Run button in the toolbar to execute the code.



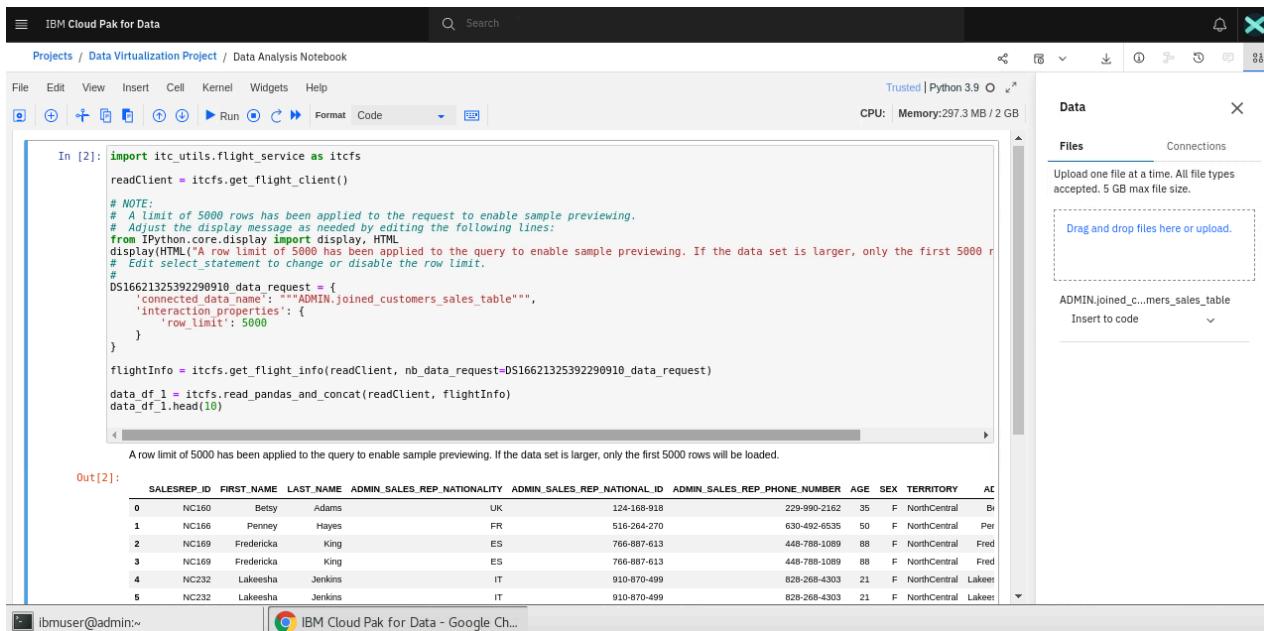
```
In [ ]: import itc_utils.flight_service as itcfs
readClient = itcfs.get_flight_client()

# NOTE:
# A limit of 5000 rows has been applied to the request to enable sample previewing.
# Adjust the display message as needed by editing the following lines:
from IPython.core.display import display, HTML
display(HTML("A row limit of 5000 has been applied to the query to enable sample previewing. If the data set is larger, only the first 5000 row"))
# Edit select_statement to change or disable the row limit.

DS16621325392290910_data_request = {
    'connected_data_name': "'ADMIN.joined_customers_sales_table'", 
    'interaction_properties': {
        'row_limit': 5000
    }
}

flightInfo = itcfs.get_flight_info(readClient, nb_data_request=DS16621325392290910_data_request)
data_df_1 = itcfs.read_pandas_and_concat(readClient, flightInfo)
data_df_1.head(10)
```

7) You should be able to access the virtualized data in the Watson Studio Notebook.



```
In [2]: import itc_utils.flight_service as itcfs
readClient = itcfs.get_flight_client()

# NOTE:
# A limit of 5000 rows has been applied to the request to enable sample previewing.
# Adjust the display message as needed by editing the following lines:
from IPython.core.display import display, HTML
display(HTML("A row limit of 5000 has been applied to the query to enable sample previewing. If the data set is larger, only the first 5000 row"))
# Edit select_statement to change or disable the row limit.

DS16621325392290910_data_request = {
    'connected_data_name': "'ADMIN.joined_customers_sales_table'", 
    'interaction_properties': {
        'row_limit': 5000
    }
}

flightInfo = itcfs.get_flight_info(readClient, nb_data_request=DS16621325392290910_data_request)
data_df_1 = itcfs.read_pandas_and_concat(readClient, flightInfo)
data_df_1.head(10)
```

A row limit of 5000 has been applied to the query to enable sample previewing. If the data set is larger, only the first 5000 rows will be loaded.

SALESREP_ID	FIRST_NAME	LAST_NAME	ADMIN.SALES.REP.NATIONALITY	ADMIN.SALES.REP.NATIONAL_ID	ADMIN.SALES.REP.PHONE_NUMBER	AGE	SEX	TERRITORY	AC	
0	NC160	Betsy	Adams	UK	124-168-918	229-990-2162	35	F	NorthCentral	Bi
1	NC166	Penny	Hayes	FR	516-264-270	630-492-6535	50	F	NorthCentral	Per
2	NC169	Frederick	King	ES	766-887-613	448-789-1089	88	F	NorthCentral	Fred
3	NC169	Frederick	King	ES	766-887-613	448-789-1089	88	F	NorthCentral	Fred
4	NC232	Lakeisha	Jenkins	IT	910-870-499	828-268-4303	21	F	NorthCentral	Lakeer
5	NC232	Lakeisha	Jenkins	IT	910-870-499	828-268-4303	21	F	NorthCentral	Lakeer

Visualize the data with a dashboard

- Under the Data Virtualization Project, In **Assets** Tab, Click on **New Asset +** to add a new Dashboard.

The screenshot shows the 'Assets' tab in the IBM Cloud Pak for Data interface. On the left, there's a sidebar with 'Asset types': Data access (1), Data (1), and Source Code (1). The main area shows a table titled 'All assets' with columns 'Name' and 'Last modified'. Three assets are listed:

Name	Last modified
Data Analysis Notebook Notebook	1 minute ago admin (You)
ADMIN.joined_customers_sales_table application/octet-stream	50 minutes ago admin (You)
DS16621325392290910 Connection	50 minutes ago admin (You)

To the right, there's a panel titled 'Data in this project' with a message: 'Drop data files here or browse for files to upload'.

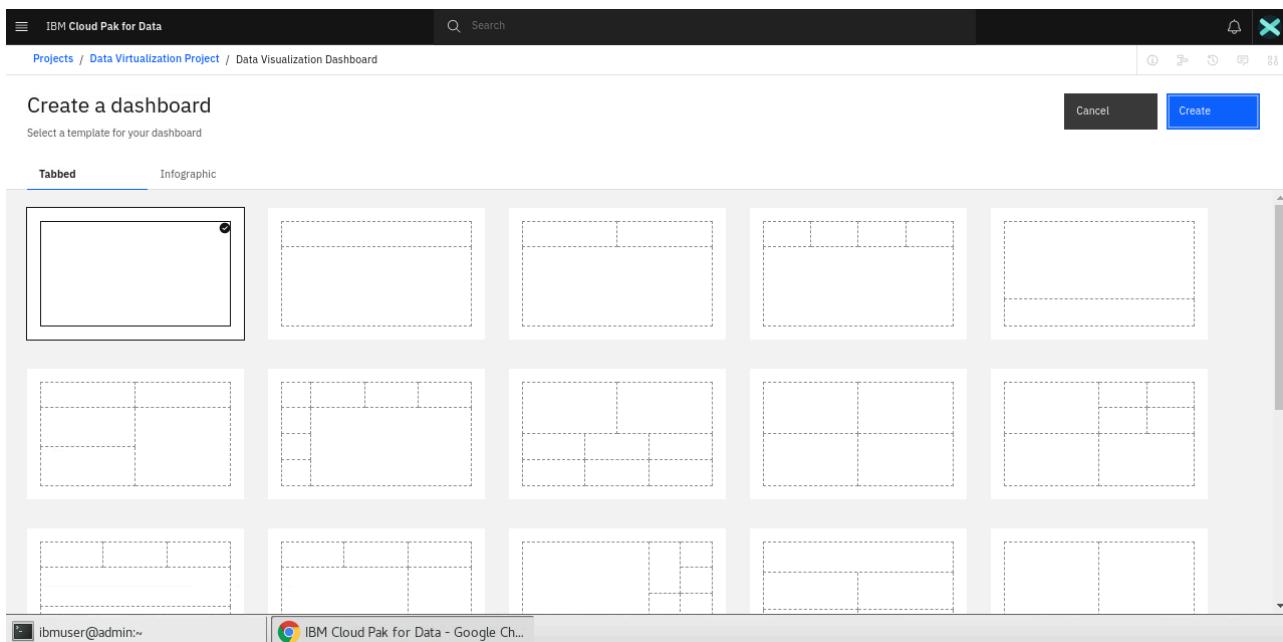
- Click on **Dashboard Editor**.

The screenshot shows the 'New asset' dialog for creating a new asset. The 'Tool type' dropdown is set to 'All types'. Under 'Automated builders', there are two options: 'AutoAI' and 'Metadata enrichment'. Under 'Graphical builders', there are four options: 'Dashboard editor', 'Data Refinery', 'DataStage', and 'Decision Optimization'. The 'Dashboard editor' option is highlighted.

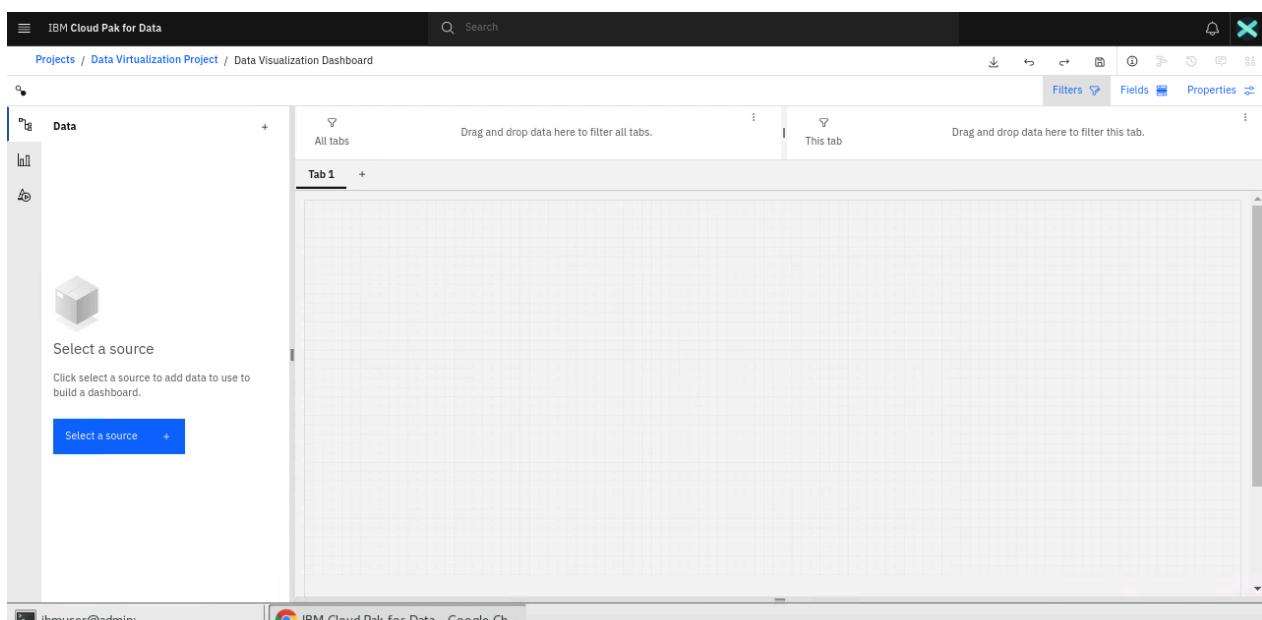
- Enter the name and description and click Create.

The screenshot shows the 'Create asset' dialog for a dashboard. The 'Name' field contains 'Data Visualization Dashboard'. The 'Description (Optional)' field also contains 'Data Visualization Dashboard'. At the bottom right, there are 'Cancel' and 'Create' buttons, with 'Create' being the active button.

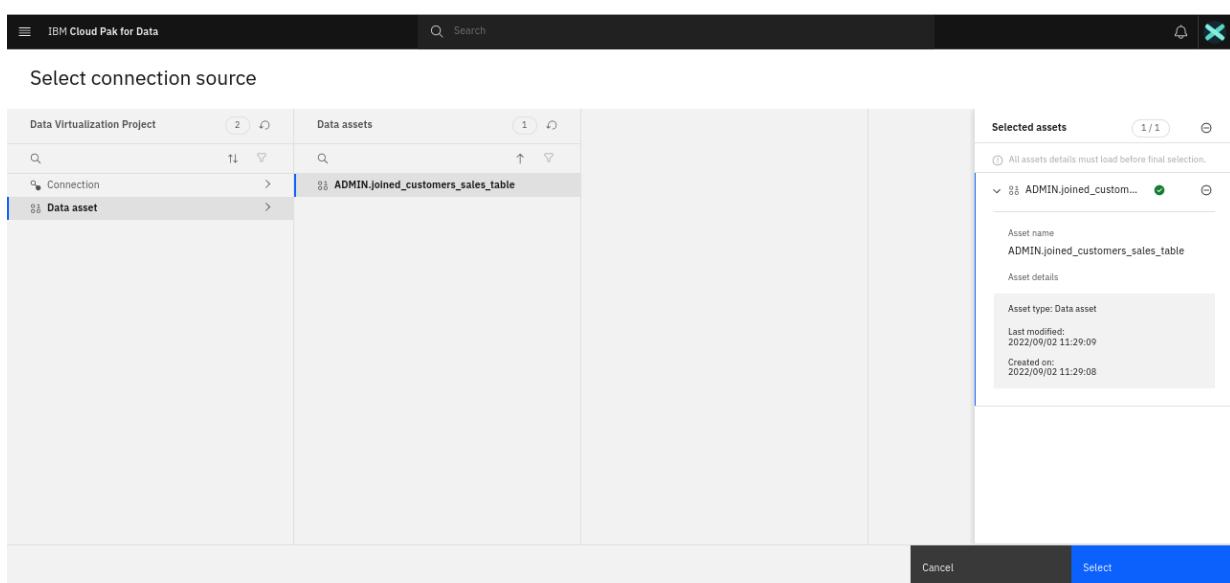
4) Click Create to create an empty tabbed Dashboard.



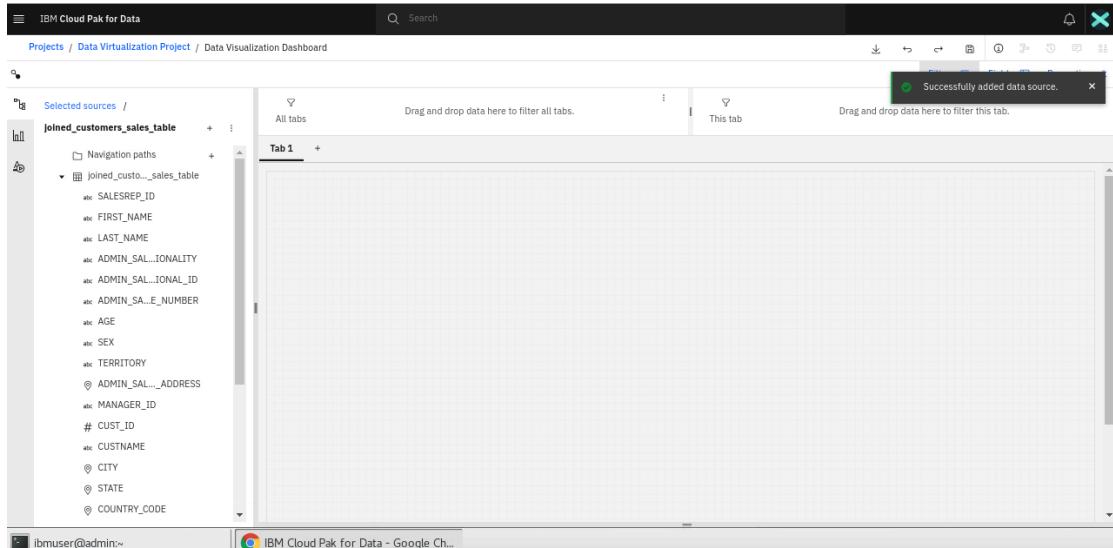
5) Click Select a source + button in the Data Pane to select a data source.



6) Select the Virtual Table. Click Select.

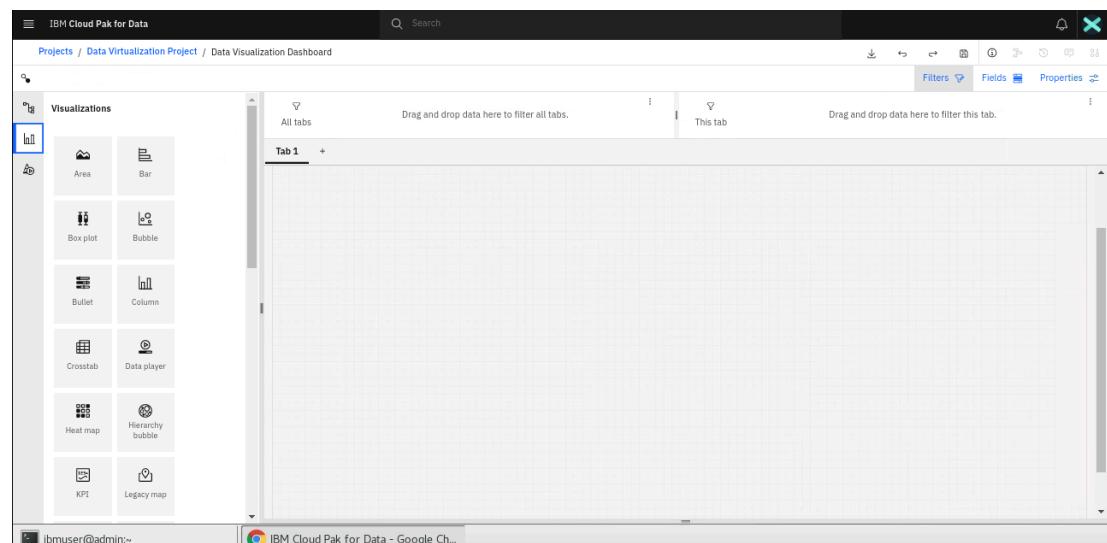


- 7) The Data source will be added successfully. Click on **Visualization** Tab  to see the types of Visualization to be used.



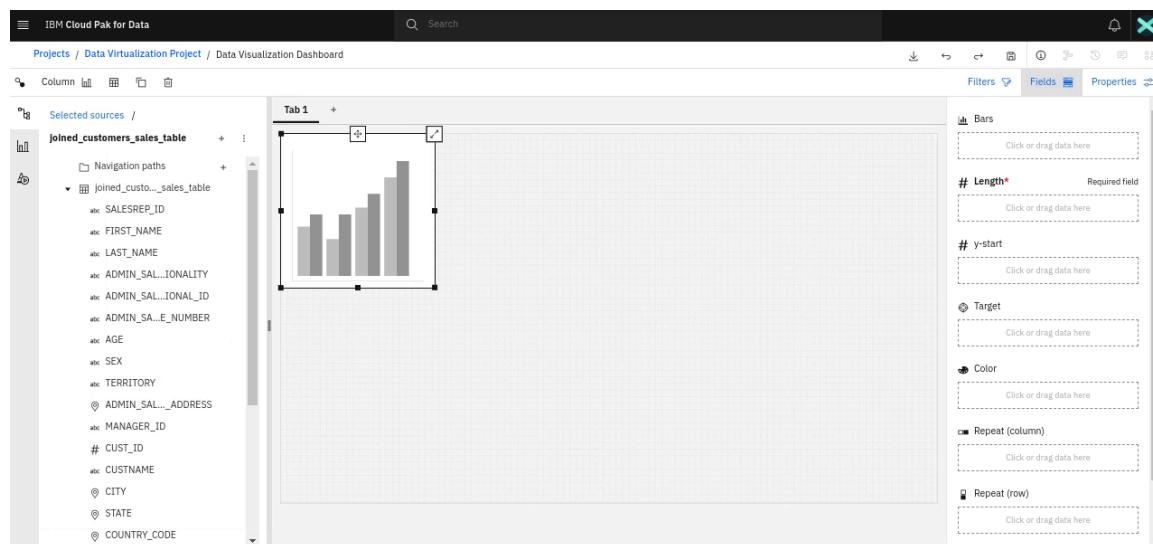
The screenshot shows the IBM Cloud Pak for Data interface. On the left, the 'Selected sources' panel displays a tree view of a joined table named 'joined_customers_sales_table'. This table includes fields from 'joined_customers_table' and 'sales_table', such as SALESREP_ID, FIRST_NAME, LAST_NAME, ADMIN_SAL...IONALITY, ADMIN_SAL...IONAL_ID, ADMIN_SA...E_NUMBER, AGE, SEX, TERRITORY, ADMIN_SAL...ADDRESS, MANAGER_ID, CUST_ID, CUSTNAME, CITY, STATE, and COUNTRY_CODE. The main workspace on the right is titled 'Tab 1' and contains a large, empty grid area. At the top right of the workspace, there is a green success message box that says 'Successfully added data source.'

- 8) Click on the required visualization to add it to the dashboard.



The screenshot shows the 'Visualizations' panel on the left side of the dashboard. It lists several chart types: Area, Bar, Box plot, Bubble, Bullet, Column, Crosstab, Data player, Heat map, Hierarchy bubble, KPI, and Legacy map. Each chart type is represented by a small icon. The main workspace on the right is titled 'Tab 1' and is currently empty.

- 9) Configure the chart to show the required visualization. Eg. X Axis/Y Axis fields.
Alternatively, New charts can also be created by directly dragging and dropping a field from the virtual table to the dashboard.



The screenshot shows a bar chart on 'Tab 1'. The Y-axis is labeled '# Length*' and the X-axis is labeled '# SALESREP_ID'. The chart displays several bars of increasing height corresponding to different sales rep IDs. To the right of the chart, the 'Fields' panel is open, showing configuration options for the chart: 'Bars', 'Length*', 'y-start', 'Target', 'Color', 'Repeat (column)', and 'Repeat (row)'. Each option has a 'Click or drag data here' placeholder.

Appendix

Below steps are for reference only.

DB2 Instance Provisioning

1. If you have a *Db2* service listed, then there is no need to provision another instance. Otherwise, follow these steps to procure a DB2 instance.
2. Under the services selection page, select **DB2**. Note that the service status is Available. For rest, the status is enabled.

The screenshot shows the 'Instances / Select a service' page. It lists three data sources: 'Data Virtualization' (Enabled), 'Db2' (Available), and 'Db2 Data Management Console' (Enabled). The 'Db2' card is highlighted with a red box. Below this, there is a 'Developer tools' section with 'RStudio Server with R3.6' listed.

3. Click New Instance.

The screenshot shows the 'Instances / Select a service / Db2' page. The 'Db2' service is selected. On the right, there is a large 'New instance' button highlighted with a red box. The left sidebar shows details for the Db2 service, including its type (Database), version (11.5.7.0-cn5-x86_64), provider (IBM), category (Data sources), and related links (Docs).

4. Provide the DB Details like DB Name (Max 8 Chars). Scroll Down.

IBM Cloud Pak for Data

Info
You must grant additional privileges to enable validation of the cluster resources such as CPU, memory, and node resources availability (including node labels/taints). See https://www.ibm.com/docs/SSQNUZ_4.0/svc-db2/cpumemprivileges.html for details.

Create a database

Configure

Database name: Customer

Number of nodes: 1

CPU per node for Db2: 2.1 (selected), 6.1, 2.1

Memory per node for Db2: 5.5 (selected), 18, 5.5

Deploy database on dedicated nodes

Namespace: cpd-instance

Continue with defaults

Cancel Previous Next →

5. Select Single Location for all Data Option and click Next.

IBM Cloud Pak for Data

Info
You must grant additional privileges to enable validation of the cluster resources such as CPU, memory, and node resources availability (including node labels/taints). See https://www.ibm.com/docs/SSQNUZ_4.0/svc-db2/cpumemprivileges.html for details.

Create a database

Type

Configure

Advanced configuration

Storage

Finalize

Memory per node for Db2: 5.5 (selected), 18, 5.5

Deploy database on dedicated nodes

Namespace: cpd-instance

Storage structure

Select a storage option: Single location for all data, Separate locations for all data

Select a supported sector size: 4K sector size

Continue with defaults

Cancel Previous Next →

6. Click Next on the advanced configuration.

IBM Cloud Pak for Data

My data: Databases / Create a database

Create a database

Type

Configure

Advanced configuration

Storage

Finalize

Advanced configuration

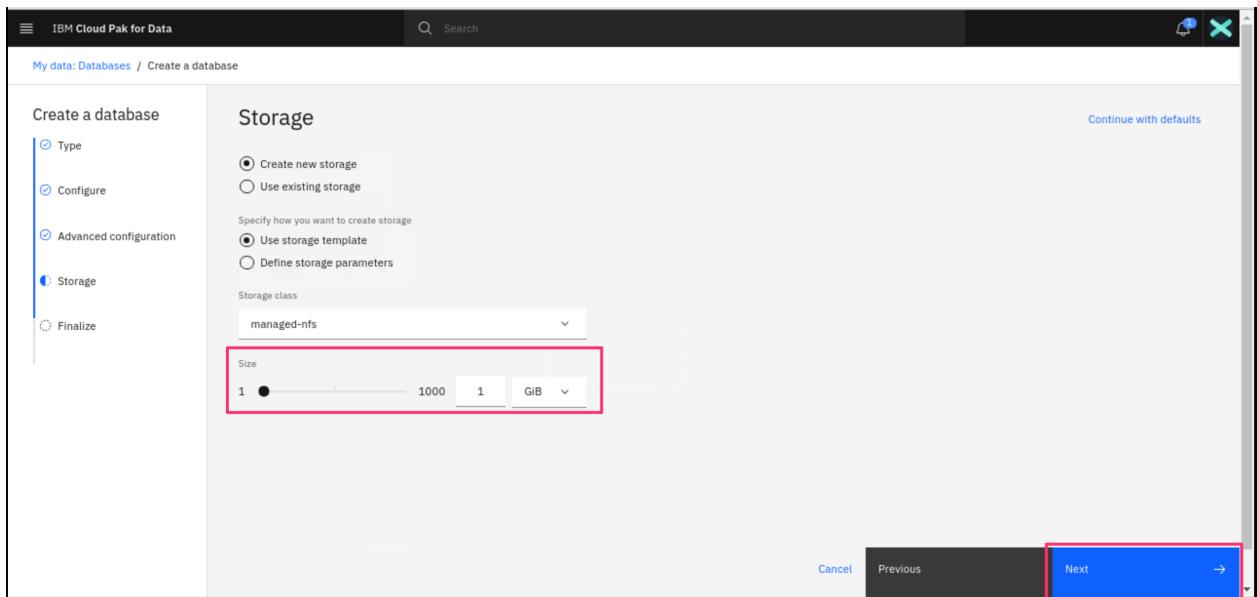
Oracle compatibility:

Page size: 16384

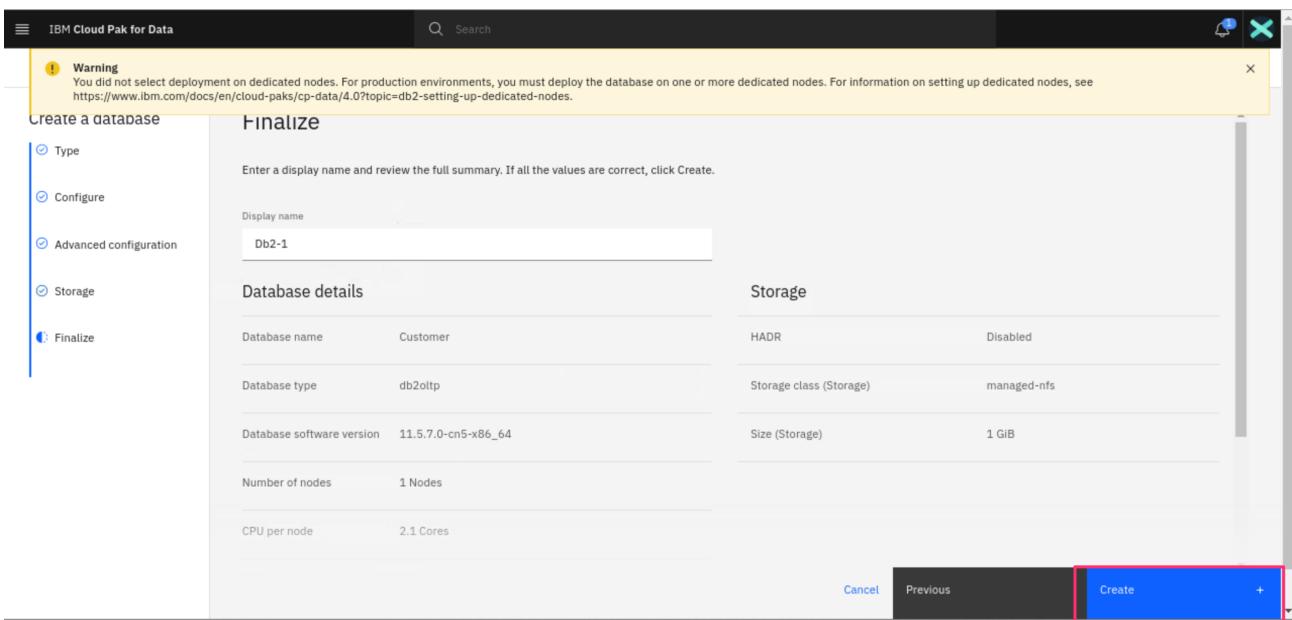
Continue with defaults

Cancel Previous Next →

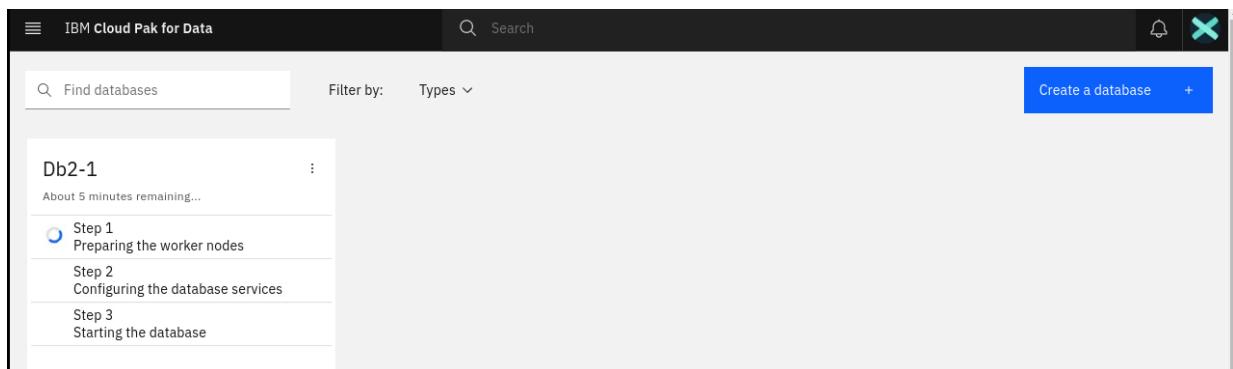
7. Configure the Storage Size and Click Next.



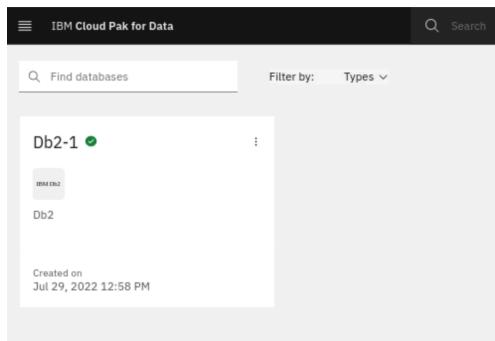
8. Click **Create** on the summary screen.



9. It will start provisioning the service.



10. The service will be provisioned in 10-20 min depends on the configuration chosen.



11. Verify that the services are provisioned on your *Service instances* page.

Instances								
Filter by:		Type	Status					
Find instances						New instance	+	
Name	Type	Created by	vCPU requests	Memory requests (GiB)	Users	Status	Created on	
Db2-1 Service instance for db2oltp-16591138850024...	db2oltp	admin	2.20	5.75 Gi	1	Green	Jul 29, 2022	
ds-px-default The default DataStage runtime instance	datastage	admin	2.50	6.00 Gi	1	Green	Jul 12, 2022	
data-virtualization	dv	admin	11.50	38.50 Gi	1	Green	Jul 6, 2022	
data-management-console Data Management Console	dmc	admin	4.20	10.62 Gi	1	Green	Jul 6, 2022	
openscale-defaultinstance IBM Watson OpenScale	aios	admin	0.00	0.00 Gi	1	Green	Jul 1, 2022	