

# IBM CLOUD PAK FOR DATA 4.5

## DATA VIRTUALIZATION HANDS ON LAB

Sandeep Ved  
CUSTOMER SUCCESS MANAGER ARCHITECT, IBM SINGAPORE

## Table of Contents

<b>Data Virtualization Lab .....</b>	<b>3</b>
Access Your Cluster.....	3
Service Provisioning .....	3
DV and DB2.....	4
PostgreSQL.....	5
Manage Users .....	5
Db2 data loading .....	10
Check DB2 Connection Details.....	15
Setting up Data Virtualization.....	18
Task 1: Create a project.....	18
Task 2: Add a connection to your Db2 data source .....	19
Task 3: Add a connection to a PostgreSQL data source .....	22
Task 4: Add tables to your virtualized data .....	24
Task 5: Join and Publish virtualized data to project .....	27
Next steps.....	34
Analyze the data in a notebook .....	34
Visualize the data with a dashboard .....	36
Appendix.....	40
DB2 Instance Provisioning.....	40

# Data Virtualization Lab

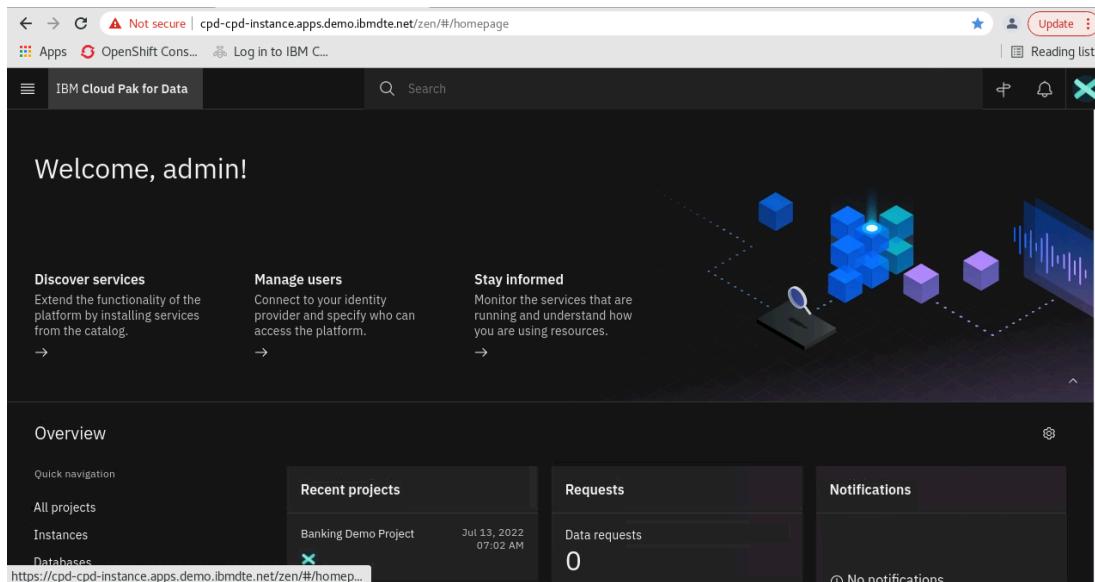
## Access Your Cluster

- 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](https://cpd-cpd-instance.apps.itzocp-664003qtcr-mn9o.selfservice.aws.techzone.ibm.com)
- 3) Login to Web Console using **admin** ID and credentials provided.

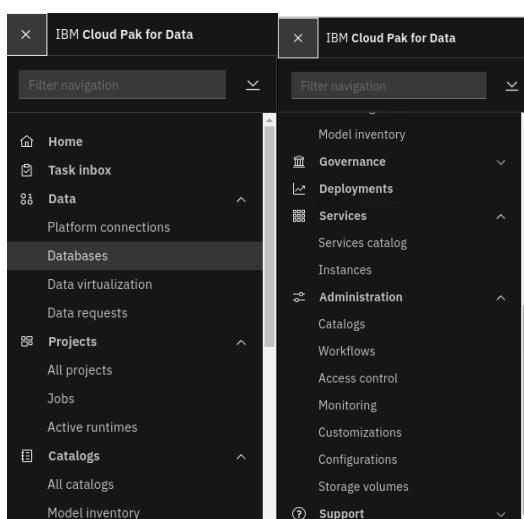
## Service Provisioning

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.



- 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.

The screenshot shows the 'Instances' page in the IBM Cloud Pak for Data interface. At the top, there are filter options for 'Type' and 'Status'. A search bar and a 'New instance +' button are also present. Below the header, a table lists four service instances:

Name	Type	Created by	vCPU requests	Memory requests (GiB)	Users	Status	Created on
data-virtualization	dv	admin	11.50	38.50 Gi	1	<span>Green</span>	13 Sept 2022
data-management-console	dmc	admin	5.20	21.40 Gi	1	<span>Green</span>	8 Sept 2022
Db2-1	Service instance for db2oltp-16625688212...	db2oltp	0.60	4.25 Gi	1	<span>Green</span>	8 Sept 2022
openscale-defaultinstance	aios	admin	-	-	1	<span>Green</span>	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.

The screenshot shows the 'Select a service' page. It includes a breadcrumb trail: 'Instances / Select a service /'. A 'Select a service' heading and a note to 'Select the service for which you want to create a new instance.' Below this, there are three sections: 'AI', 'Analytics', and 'Data Refinery'.

- AI** section: Contains three items: Watson Machine Learning, Watson OpenScale, and Watson Studio. Each item has a brief description and an 'Enabled' status indicator.
- Analytics** section: Contains one item: Data Refinery. It also has an 'Enabled' status indicator.
- Data Refinery** section: This section is partially visible and contains the same information as the Analytics section.

The screenshot shows the 'Instances' section of the IBM Cloud Pak for Data interface. It lists two service categories: 'Dashboards' (1 item) and 'Data governance' (1 item). Each category has a card with a brief description and an 'Enabled' status indicator.

- Cognos Dashboards**: Identify patterns in your data with sophisticated visualizations. No coding needed. Enabled ✓
- Watson Knowledge Catalog**: Organize and govern data. Automatically discover, classify, profile, and protect your data so data scientists can find trusted data fast. Enabled ✓

## PostgreSQL

6) An external PostgreSQL DB database has been already prepared for the lab with the following connection details.

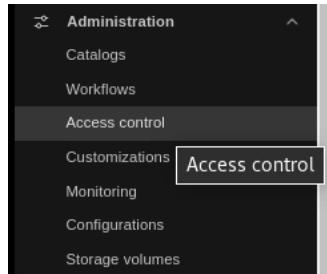
- Database Name: vlyawtap
- Host Name: tiny.db.elephantsql.com
- Port: 5432
- User Name: vlyawtap
- Password: fXt4TyCB\_W0d0LCaCaPF7MbLKWlpti60

## Manage Users

1) Navigate to CP4D Console Home Page.

The screenshot shows the IBM Cloud Pak for Data Console Home Page. The main header says 'Welcome, admin!'. Below it, there are three main sections: 'Discover services', 'Manage users', and 'Stay informed'. The 'Overview' section on the left includes links for 'All projects', 'Instances', 'Databases', 'Data virtualization', 'Support', 'Documentation', 'Community', and 'Diagnostics'. The right side features four cards: 'Recent projects' (zz01jj856, Sep 07, 2022 10:59 PM), 'Requests' (0 Data requests), 'Deployment spaces' (Credit Risk, Sep 08, 2022 11:31 AM), and 'My instances' (data-virtualization, dv). A large sidebar on the right displays a 3D visualization of data cubes and a waveform, with a 'Not Secure' warning at the top.

- 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.

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

Full name (optional)  
Sandeep Ved

Username  
svd

Email (optional)  
sandeep.ved1@ibm.com

Password  
\*\*\*\*\*

Confirm password  
\*\*\*\*\*

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

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

**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](https://www.ibm.com/docs/en/cloud-paks/cp-data/4.5.x?topic=users-predefined-roles-permissions-in-cloud-pak-data) -> <https://www.ibm.com/docs/en/cloud-paks/cp-data/4.5.x?topic=users-predefined-roles-permissions-in-cloud-pak-data>

Add users

Authorize user access to the platform.

The screenshot shows the 'Add users' interface with the 'Roles' tab selected. On the left, there's a sidebar with options: Profile information (radio button), Platform access (radio button), Roles (radio button, selected), and Summary (radio button). The main area is titled 'Roles' with a sub-instruction 'Assign roles to this user or [create a new role](#)'. A search bar says 'Find roles'. Below is a list of roles with checkboxes. The 'Data Engineer' role is checked and highlighted with a blue border. To its right, detailed information is shown for the 'Administrator' role, including its description ('Administrator role'), modification date ('Modified on Aug 30, 2022 7:57 AM'), and a list of 13 permissions and 73 actions. At the bottom are 'Cancel', 'Back', and 'Next' buttons.

- 7) Click Next.

Add users

Authorize user access to the platform.

The screenshot shows the 'Add users' interface with the 'Roles' tab selected. The 'Data Engineer' role is now selected (indicated by a checked checkbox). The right panel shows its details: 'Description' (Data engineer role), 'Modified on' (12 Sept 2022 22:13), and '1 permission, 1 action' (Create projects). At the bottom are 'Cancel', 'Back', and a large blue 'Next' button.

- 8) Click Add on the Summary Window.

Add users

Authorize user access to the platform.

The screenshot shows the 'Summary' window. It contains a summary of the user and role information: 'Username' (sved), 'Name' (Sandeep Ved), 'Email' (sandeep.ved1@ibm.com), 'Role name' (Data Engineer), and 'Description' (Data engineer role). At the bottom are 'Cancel', 'Back', and a large blue 'Add' button.

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

The screenshot shows the 'Access control' section of the IBM Cloud Pak for Data interface. It includes a search bar, a 'LDAP configuration' link, and tabs for 'Users', 'User groups', and 'Roles'. A table lists users with columns for Name, User ID, Username, Created on, and Roles. Two users are listed: 'Sandeep Ved' (User ID 1000331001, Username svd, Roles Data Engineer) and 'admin' (User ID 1000330999, Username admin, Roles Administrator + 2 more). A blue 'Add users +' button is visible.

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

The screenshot shows the 'Instances' page. It lists several instances including 'data-virtualization', 'data-management-console', 'Db2-1', and 'openscale-defaultinstance'. For the 'Db2-1' instance, a context menu is open with options: 'Open', 'Manage access', and 'Delete'. The 'Manage access' option is highlighted with a blue border.

- 11) Review the users with Access to DB2.

The screenshot shows the 'Access management: Db2-1' page. It displays a table with columns for Name, Username, and Service role. One user, 'admin', is listed with a service role of 'Admin'. A blue 'Add users +' button is visible.

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

The screenshot shows a modal dialog titled 'Grant access to users and user groups'. It asks to specify users who can access Db2 and their roles. A table lists users and groups, with 'admin' and 'Sandeep Ved' selected. A blue 'Add' button is at the bottom right.

This screenshot shows the 'Access management' section for the 'Db2-1' instance. It lists two users: 'admin' and 'Sandeep Ved'. Both users have the 'Admin' service role assigned. A blue box highlights the 'Add users' button in the top right corner.

Name	Username	Service role
admin	admin	Admin
Sandeep Ved	sved	Admin

**13)** Similarly assign access to new users to the data virtualization instance as well.

This screenshot shows the 'Instances' list for the 'data-virtualization' instance. A context menu is open over the 'Db2-1' row, with the 'Manage access' option highlighted by a blue box.

Name	Type	Created by	vCPU requests	Memory requests (GiB)	Users	Status	Created on
openscale-1 IBM Watson OpenScale	aios	admin	0.00	0.00 Gi	1	Green	20 Sept 2022
data-virtualization	dv	admin	11.50	38.50 Gi	3	Green	13 Sept 2022
data-management-console IBM Db2 Data Management Console	dmc	admin	5.20	21.40 Gi	3	Green	8 Sept 2022
Db2-1 Service instance for db2oltp-16625688212...	db2oltp	admin	0.60	4.25 Gi	3	Green	8 Sept 2022

This screenshot shows the 'Access management' section for the 'data-virtualization' instance. It lists three users: 'admin', 'Satit Pongbundit', and 'Sandeep Ved'. The 'Manage access' button in the top right corner is highlighted by a blue box.

Name	Username	Service role
admin	admin	Admin
Satit Pongbundit	satit.pongbundit	Engineer
Sandeep Ved	sved	Admin

**14)** Similarly assign access to new users to the data management console instance as well.

This screenshot shows the 'Instances' list for the 'data-management-console' instance. A context menu is open over the 'Db2-1' row, with the 'Manage access' option highlighted by a blue box.

Name	Type	Created by	vCPU requests	Memory requests (GiB)	Users	Status	Created on
openscale-1 IBM Watson OpenScale	aios	admin	0.00	0.00 Gi	1	Green	20 Sept 2022
data-virtualization	dv	admin	11.50	38.50 Gi	3	Green	13 Sept 2022
data-management-console IBM Db2 Data Management Console	dmc	admin	5.20	21.40 Gi	3	Green	8 Sept 2022
Db2-1 Service instance for db2oltp-16625688212...	db2oltp	admin	0.60	4.25 Gi	3	Green	8 Sept 2022

Name	Username	Service role
admin	admin	Admin
Sandeep Ved	sved	Admin
Satit Pongbundit	satit.pongbundit	User

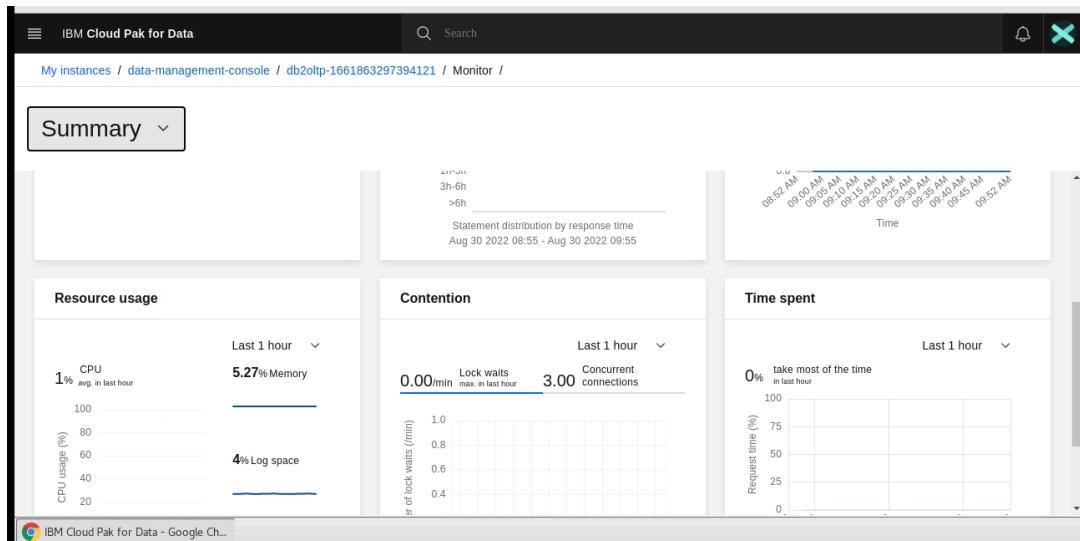
## Db2 data loading

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

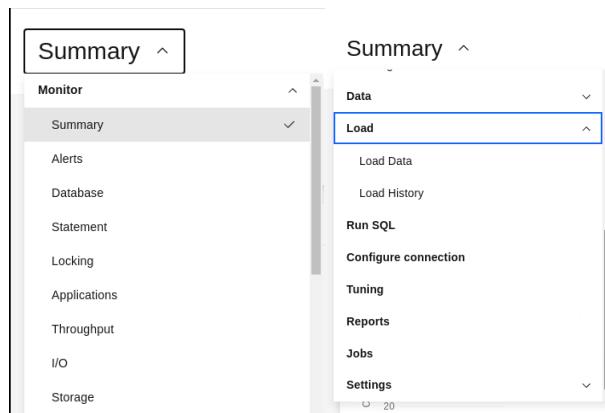
1. Access these two data sets in the GitHub, and download them to your local machine:
  - o **Customers data set -> customers.csv** (To upload data in DB2 Database)
  - o **Sales data set -> sales.csv** (For Reference only, as its already loaded in PostgreSQL DB)
2. Navigate to **Data -> Databases.**

3. Click on **Open Database**

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

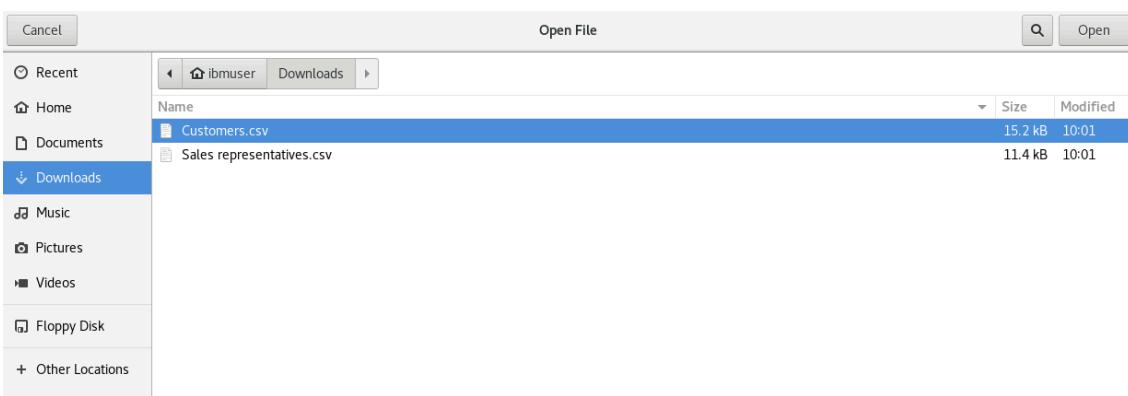


5. Under the Summary Drop down, click on **Load -> Load Data**.

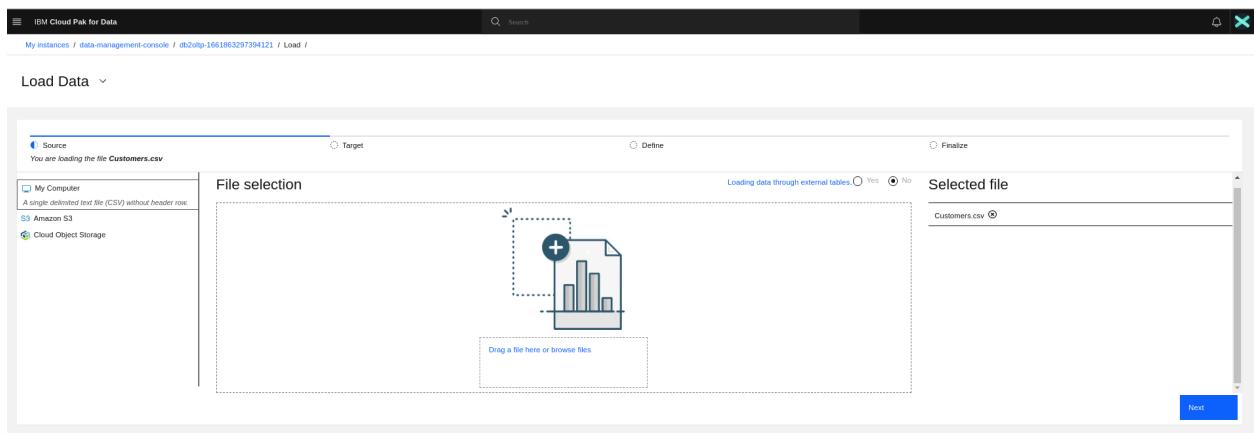


6. The DB2 Load Data Page will appear. Select or Drag the **customers.csv** file from your local machine into the *File Selection* panel in the *Db2* console.

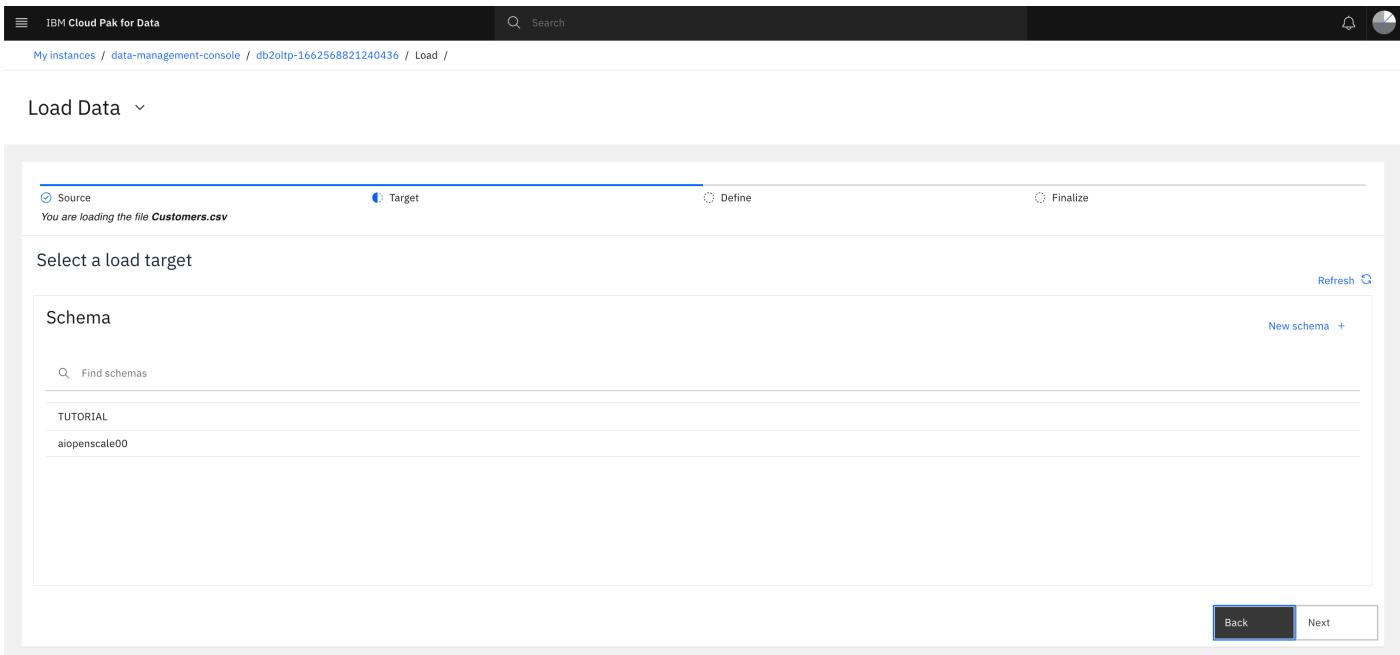
The screenshot shows the 'Load Data' configuration page. The 'Source' tab is selected. In the 'File selection' panel, there is a 'Drag a file here or browse files' area with a plus sign icon. At the bottom right of the panel, there is a 'Next' button.



**7. Click Next.**



**8. Click New Schema+.**



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

The screenshot shows the 'Load Data' interface in the IBM Cloud Pak for Data Practicum. The top navigation bar includes 'IBM Cloud Pak for Data', a search bar, and a notification icon. The URL in the address bar is 'My instances / data-management-console / db2oltp-1662568821240436 / Load /'. The main area is titled 'Load Data' with a dropdown menu. Below it, there are four tabs: 'Source' (selected), 'Target', 'Define', and 'Finalize'. A message says 'You are loading the file Customers.csv'. The 'Select a load target' section shows a 'Schema' list with 'IBMADMIN', 'TUTORIAL', and 'aiopenscale00'. To the right, a 'Create a new schema' panel has 'IBMADMIN' in a text input field and a 'Create' button. At the bottom are 'Back' and 'Next' buttons.

## 10. Click New table +.

The screenshot shows the 'Load Data' interface in the IBM Cloud Pak for Data Practicum. The top navigation bar includes 'IBM Cloud Pak for Data', a search bar, and a notification icon. The URL in the address bar is 'My instances / data-management-console / db2oltp-1662568821240436 / Load /'. The main area is titled 'Load Data' with a dropdown menu. Below it, there are four tabs: 'Source' (selected), 'Target', 'Define', and 'Finalize'. A message says 'You are loading the file Customers.csv'. The 'Select a load target' section shows a 'Schema' list with 'IBMADMIN', 'TUTORIAL', and 'aiopenscale00'. To the right, a 'Table' section shows a search bar 'Find tables in IBMADMIN' and a message 'No entries found.' Below are 'Back' and 'Next' buttons.

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

Select a load target

**Schema**

- IBADMIN (selected)
- TUTORIAL
- aipayscale00

**Table**

No entries found.

**Create a new table**

Customers

**Back** **Next**

12.Click **Next** to continue.

Select a load target

**Schema**

- IBADMIN (selected)
- TUTORIAL
- aipayscale00

**Table**

CUSTOMERS (selected)

**Back** **Next**

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

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	724-454-2453	90.3
2	Renee Mullins	Columbus	OH	US	45101	Rene.K.mullins@dodgeit.com	229-990-2162	0
3	Allen Schmidt	Abeta	PG	IT	6040	Allen.M.schmidt@spambob.com	289-202-8553	304
4	Robert May	Houston	TX	US	77001	Robert.C.may@spambob.com	630-492-6555	304
5	Joe Cruz	Harrisonburg	VA	US	24210	Joe.D.cruz@spambob.com	446-788-1089	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		88480	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

**Back** **Next**

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

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

## 16. Table details will appear.

CUST_ID	CUSTNAME	CITY	STATE	COUNTRY_CODE	POSTAL_CODE	EMAIL_ADDRESS	PHONE_NUMBER	YTD_SALES	SALESREP_ID	NATIONALITY	NATIONAL_ID	CREDITCARD_TYPE	CREDITCARD_NUMBER	
1	Michael Golden	Abbadia Alpina	TO	IT	10060	Michael.E.golden@is	724-454-8453	90.30	SE133	IT	152-374-114	American Express	10-24	
2	Renee Mullins	Columbus	OH	US	45101	Rene.K.mullins@dod	229-990-2162	0.00	NC169	UK	124-168-918	5148585346736100	American Express	10-24
3	Allen Schmidt	Albeto	PG	IT	60340	Allen.M.schmidt@isp	288-202-8653	304.00	RP385	U.S.	346-450-984	34107893021430	Master Card	10-24
4	Robert May	Houston	TX	US	79601	Robert.C.may@span	630-492-6535	304.00	SE337	U.S.	516-264-270	4861418511939110	VISA	10-24
5	Joe Cruz	Harrisonburg	VA	US	24210	Joe.D.cruz@span	448-788-1085	180.30	WE355	ES	766-887-613	49149125367230	VISA	10-24
6	Rebecca White	Acate	RG	IT	97011	Rebecca.white@spa	828-288-4303	52.00	SC325	IT	910-870-499	578693646984100	American Express	10-24
7	Anthony Johnson	Achstetten	DE		88480	Anthony.H.johnson@	907-111-5490	354.00	WE349	IT	595-442-524	419775925770030	VISA	11-24
8	Alberto Fabian	Acquafona	BL	IT	32043	Alberto.F.fabian@gi	113-785-7672	352.00	SW211	CA	270-598-741	58039342795960	American Express	11-24
9	Ronald Gordon	Acquafrica Del Capo	LE	IT	73040	Ronald.J.gordon@gi	994-960-8008	0.00	NW313	IT	862-654-322	5612076703910780	American Express	11-24
10	Christopher Marcelli	Acquareola	SA	IT	84080	Christopher.V.marcelli@gi	717-217-7230	180.30	NC160	FR	173-526-693	595263933628150	American Express	11-24
11	Chris Green	ACTON	UK		SY9 6CX	Chris.M.green@poo	933-753-6160	5669.40	RP385	U.S.	321-821-746	2781411612400590	Discover	11-24
12	Norbert Crawford	Lensing	MI	US	49301	Norbert.T.crawford@	983-926-6221	6101.00	SW280	CA	626-436-952	2380938407157940	Discover	11-24

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.

About this database	
Database name	BLUDB01
Database type	db2oltp
Database software version	11.5.7.0-cn6-x86_64
Processor	x86-64
Deployment id	db2oltp-1662568821240436
Created on	8 Sept 2022 00:40
Status	Available

Access information		
JDBC Connection URL	jdbc:db2://<CLUSTER_ACCESSIBLE_IP>:32757/BLUDB01:user=-;password=<password>;securityMechanism=9;encryptionAlgorithm=2	
JDBC Connection URL (SSL)	jdbc:db2://<CLUSTER_ACCESSIBLE_IP>:31189/BLUDB01:user=-;password=<password>;securityMechanism=9;sslConnection=true;encryptionAlgorithm=2	

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

The screenshot shows the 'Authorization' section of the IBM Cloud Pak for Data interface. On the left, there's a navigation sidebar with categories like Monitor, Data, Views, Indexes, Aliases, MQTs, Schemas, Sequences, Storage objects, Application objects, Authorization, Workloads, Load, Run SQL, Configure connection, Tuning, and Reports. The 'Authorization' section is expanded. At the top of this section, there are buttons for 'Add role', 'Remove role', 'Grant multiples', and 'Revoke multiples'. Below these buttons is a search bar labeled 'Find name'. The main area displays a table with columns 'Type' and 'Name'. The types listed are USER, GROUP, and ROLE. There are multiple entries for each type, such as 'USER' entries for 'NULLID', 'SQLJ', 'SYSCAT', 'SYSFUN', 'SYSIBM', 'SYSIBADM', 'SYSIBINTERNAL', 'SYSIBMTS', 'SYSPROC', 'SYSPUBLIC', 'SYSSTAT', and 'SYSTOOLS'.

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

The screenshot shows the 'Run SQL' interface. On the left, there's a tree view of 'Data objects' including IBMADMIN, IBMCONSOLE, IBM\_RTMON, NULLID, SQLJ, SYSCAT, SYSFUN, SYSIBM, SYSIBADM, SYSIBINTERNAL, SYSIBMTS, SYSPROC, SYSPUBLIC, SYSSTAT, and SYSTOOLS. The main area is a script editor titled '\* Untitled - 1' with a toolbar above it. The toolbar includes icons for file operations, a syntax assistant toggle, and a 'Run all' button. The script editor contains a single line of SQL: 'grant dbadm on database to user sved;'. Below the script editor is a 'History' tab showing the execution of the grant command.

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 again. The left sidebar shows the same list of data objects. The main area now contains two lines of SQL: 'grant dbadm on database to user sved;' and 'grant secadm on database to user sved;'. The 'Run all' button is highlighted in blue. Below the script editor is a 'History' tab showing the execution of the two grant commands. The history table has columns for 'Script', 'Date', 'Status', and 'Runtime'. It lists three rows corresponding to the executed grant statements.

# 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. Refer this link for more details -> <https://www.ibm.com/docs/en/cloud-paks/cpdata/4.5.x?topic=data-virtualizing>

## 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.

Name	Date created	Your role	Collaborators
Banking Demo Project	2 months ago	Admin	
Datastage Demo Assets	2 months ago	Admin	

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

**Create an empty project**

Add the data you want to prepare, analyze, or model. Choose tools based on how you want to work: write code, create a flow on a graphical canvas, or automatically build models.

**USE TO**

**Create an empty project**

- Prepare and visualize data
- Analyze data in notebooks
- Train models

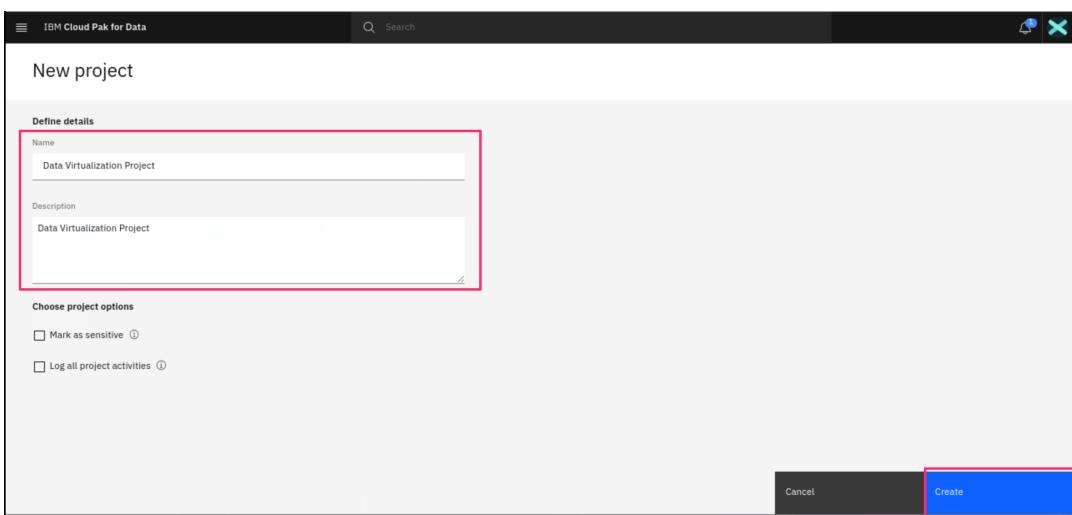
**Create a project from a file**

Get started fast by loading existing assets. Choose a project file from your system.

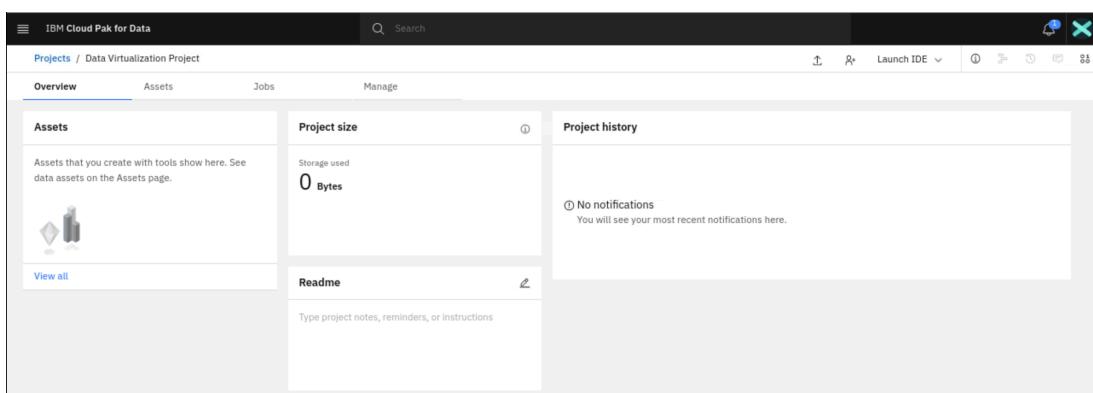
**USE TO**

- Learn by example
- Build on existing work
- Run tutorials

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



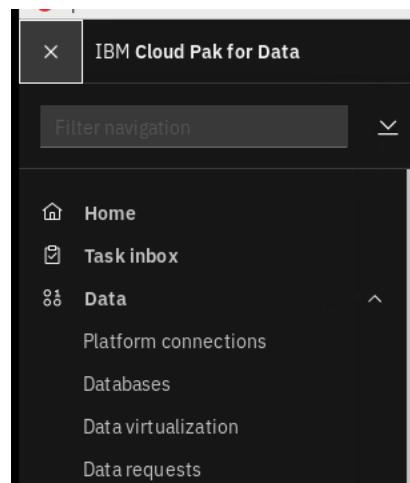
6. The Project will be created successfully.



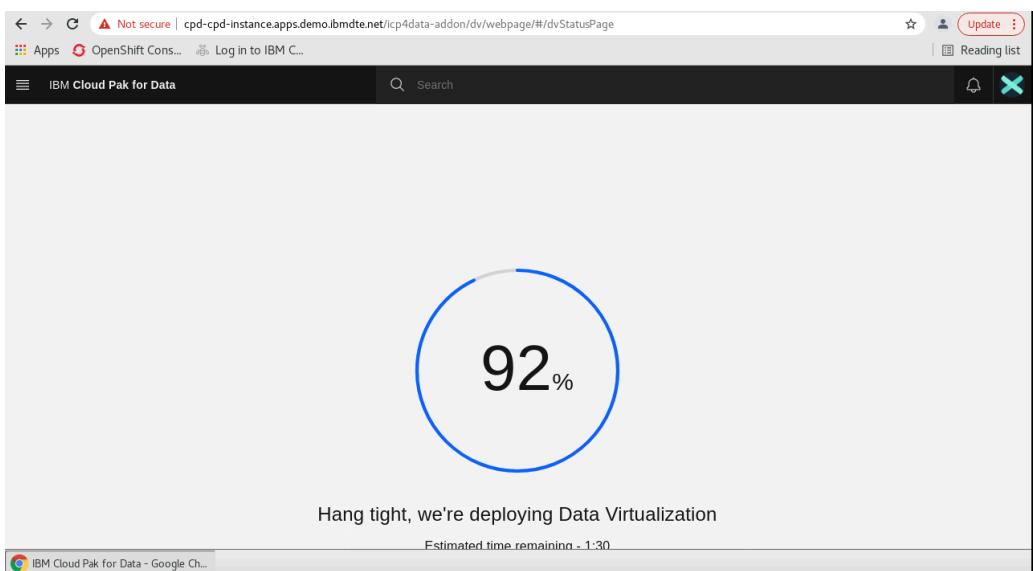
## Task 2: Add a connection to your Db2 data source

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.

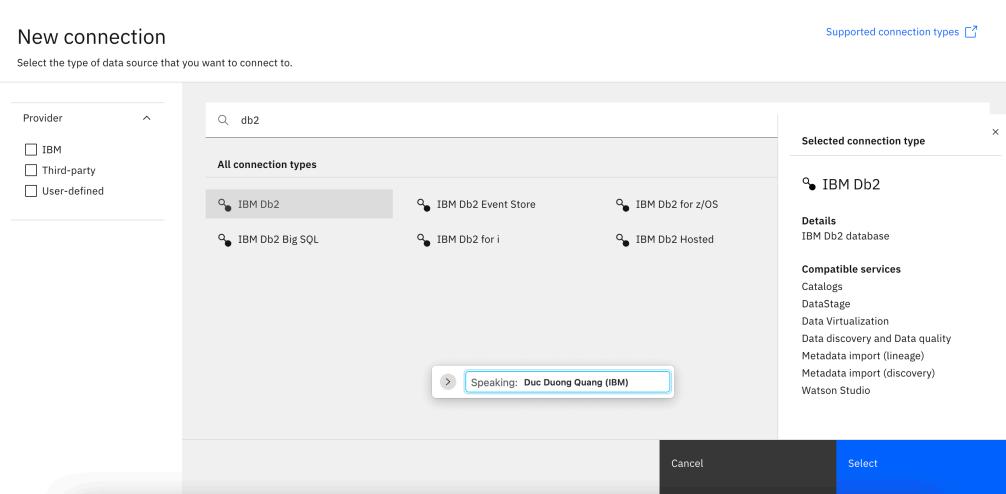


3. After initial initialization, the Data Virtualization Data Sources Page appears.

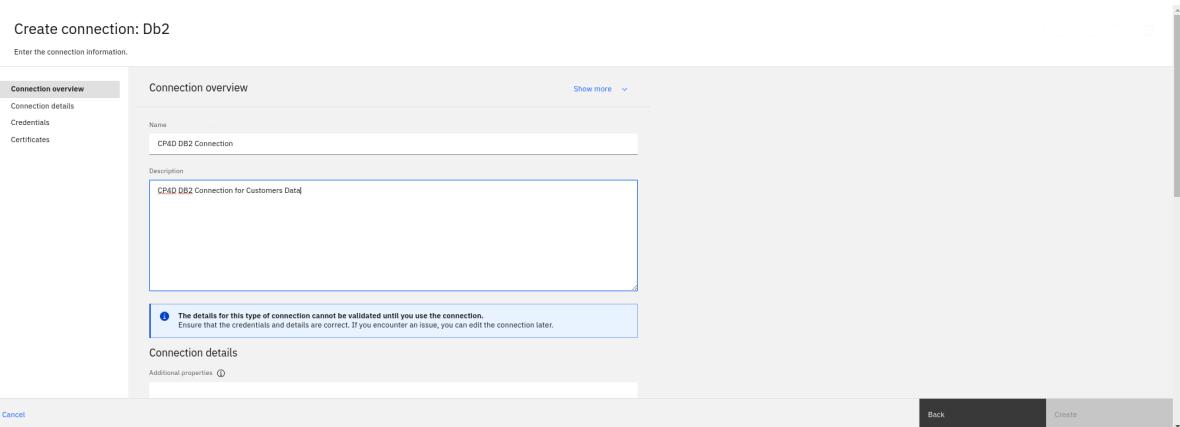
Name	Endpoint	Type	Status	Listed tables

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

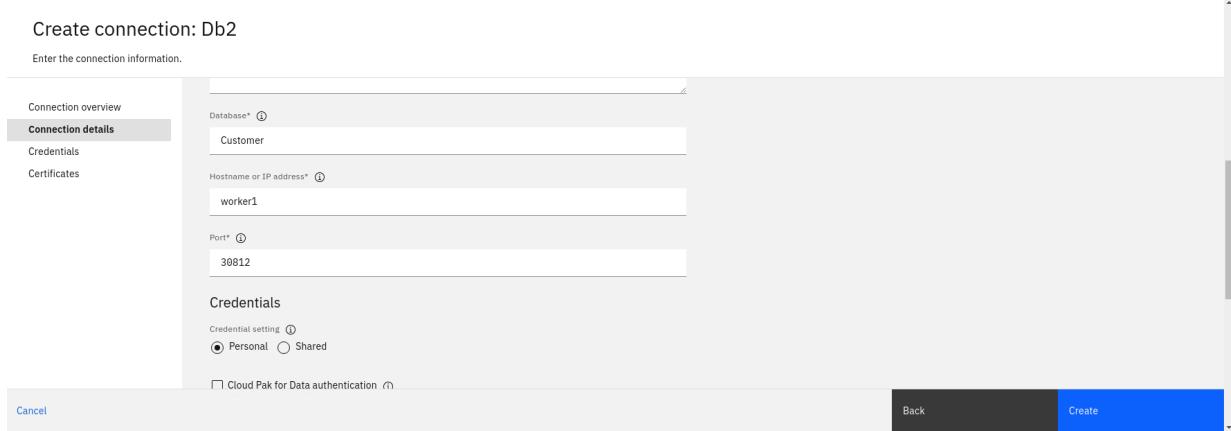
5. Select **IBM Db2**. Click **Select**.



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



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



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

### Create connection: Db2

Enter the connection information.

Connection overview Connection details <b>Credentials</b> Certificates	<input checked="" type="radio"/> Personal <input type="radio"/> Shared <input type="checkbox"/> Cloud Pak for Data authentication Username* <input type="text" value="admin"/> Password* <input type="password"/> Username and Password Security Mechanism <input type="text" value="Default"/> Username and Password Encryption Algorithm <input type="text" value="Default"/> Certificates <input type="checkbox"/> Port is SSL-enabled
<a href="#">Cancel</a> <a href="#">Back</a> <a href="#" style="background-color: blue; color: white; padding: 5px;">Create</a>	

### 9. Click Skip

Add to a remote connector (optional)

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

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 +](#)

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

Data sources

Tell IBM what you think. Let us know what you think about IBM Cloud Pak for Data - Data Virtualization service. It only takes a few minutes, and your feedback will help us improve our product.  
[IBM Privacy Policy](#)

No thanks [Provide feedback](#)

Table view [Constellation view](#)

Connections: 1 | Remote connectors: 0

Find by Name, Endpoint, Type

Name	Endpoint	Type	Status	Listed tables
CP4D DB2 Connection	worker1: 30812	Db2 Family	Active	318 / 318
Database	Port	Username	Remote connector	Description
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.

- 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' section of the IBM Cloud Pak for Data management console. A single connection, 'CP4D DB2 Connection', is listed. It has the endpoint 'worker1: 30812', type 'Db2 Family', and status 'Active'. The 'Listed tables' column shows '318 / 318'. Below the table view, there's a 'Constellation view' button. On the right, a modal window titled 'Add connection' is open, showing options for 'New connection', 'Existing platform connection', and 'Remote data source'. There's also a feedback survey at the top of the page.

## 3. Select PostgreSQL and click Select.

The screenshot shows the 'New connection' dialog. In the search bar, 'postgre' is typed, and the 'PostgreSQL' provider is selected. To the right, details about the 'PostgreSQL' connection type are shown, including 'Details' (PostgreSQL database) and 'Compatible services' (Catalogs, DataStage, Data Virtualization, Metadata Import). At the bottom, there are 'Cancel' and 'Select' buttons, with 'Select' being highlighted.

## 4. Enter the Connection Name and Description.

The screenshot shows the 'Create connection: PostgreSQL' dialog. Under 'Connection overview', the 'Name' field is set to 'PostgreSQL on ElephantSQL' and the 'Description' field contains the text 'Added by Sandeep Ved - Testing'. The left sidebar lists 'Connection details', 'Credentials', and 'Certificates'. At the bottom, there are 'Cancel', 'Back', and 'Create' buttons, with 'Create' being highlighted.

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

- Database Name: vlyawtап
- 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	Database* ⓘ
Connection details	vlyawtap
Credentials	Hostname or IP address* ⓘ
Certificates	tiny.db.elephantsql.com
	Port* ⓘ
	5432
Credentials	Credential setting ⓘ
	<input checked="" type="radio"/> Personal <input type="radio"/> Shared
	Username* ⓘ
	vlyawtap
	Password* ⓘ
	..... <input type="button" value="Show password"/>
Certificates	<input type="checkbox"/> Port is SSL-enabled ⓘ

[Cancel](#) [Back](#) [Create](#)

## 6. Click **Skip**.

My instances / data-management-console / dv-1649188282576777 /

Add to a remote connector (optional)

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

Remote connector	Hostname	Description	Port	Username
				

You haven't set up any remote connectors yet.

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

## 7. The new data source will be added.

IBM Cloud Pak for Data

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

Data sources

Table view Constellation view

Connections: 2 | Remote connectors: 0

Find by Name, Endpoint, Type

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.

1. From the *Data sources* menu, select **Virtualization > Virtualized Data**.

The screenshot shows the 'Data sources' section of the 'Virtualization' tab. The left sidebar lists 'Virtualization', 'Data sources', 'Virtualize', 'Virtualized data', 'Cache management', 'Monitor', 'Data', 'Run SQL', and 'Administration'. The right panel shows two entries under 'Endpoint': 'tiny.db.elephantsql.com: 5432' and 'worker1: 30812'.

## 2. Click on Add Virtual Objects.

The screenshot shows the 'Virtualized data' section. It includes a search bar, a filter dropdown set to 'All types', and a table header with columns for 'Table', 'Schema name', 'Created on', and 'Statistics last collected on'. A message at the top states: 'You don't have any virtualized objects yet.' Below it, a note says: 'Add connections to your data sources, then you can create virtual tables and views. Learn more'. A blue 'Add virtual objects' button is located at the bottom left.

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

The screenshot shows the 'Virtualize' section with the 'Tables' tab selected. The 'Connections' tab is open, showing a list with 'PostgreSQL db ElephantSQL...' and 'CP4D DB2 Connection (318)'. The 'PostgreSQL db ElephantSQL...' entry is selected. A message in the center says: '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.

The screenshot shows the 'Virtualize' section with the 'Tables' tab selected. The 'Connections' tab is open, showing 'IBMADMIN (1)' selected. Under 'IBMADMIN (1)', the 'CUSTOMERS' table is selected. A message in the center says: '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.

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**.

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**.

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.

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

Review cart and virtualize tables

Assign to:

- Data request
- Project
- Virtualized data

Publish to:

- Catalog

Default Catalog

Table	Schema	Source schema	Connections	Databases/File Path	Hostname: Port	Grouped tables
CUSTOMERS	ADMIN	IBADMIN	CP4D DB2 Connection	Customer	worker1: 30812	1
SALES_REP	ADMIN	sample	PostgreSQL on Elepha...	vfyawtap	tinydb.elephantsql.co...	1

## 9. Click **Confirm**.

## Confirm virtualization

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.

Do not show this message again

Cancel

Confirm

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

The screenshot shows a modal window titled "Virtualize objects". It displays two rows of data:

Table	Schema	Virtualization status	Publish status
CUSTOMERS	ADMIN	Success	Success
SALES_REP	ADMIN	Success	Success

At the bottom, there are two buttons: "Virtualize more data" (disabled) and "Go to virtualized data" (highlighted with a blue background).

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

The screenshot shows the "Virtualized data" screen. Under the heading "1 item selected", it lists two items:

Table	Schema name	Created on	Statistics last collected on
SALES_REP	ADMIN	Sep 2, 2022 11:21 AM	Not collected
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**.

The screenshot shows the "Virtualized data" screen. Under the heading "2 items selected", it lists the two tables selected for joining:

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

2. Connect the **SALESREP\_ID** columns in the two tables. Click the link **Open in SQL Editor** to see the actual query used to join the tables and create the view.

Table 1: SALES\_REP

Column name	Data type
SALESREP_ID	CLOB
FIRST_NAME	CLOB
LAST_NAME	CLOB
NATIONALITY	CLOB
NATIONAL_ID	CLOB
PHONE_NUMBER	CLOB
AGE	CLOB
SEX	CLOB
TERRITORY	CLOB

Table 2: CUSTOMERS

Column name	Data type
STATE	VARCHAR
COUNTRY_CODE	VARCHAR
POSTAL_CODE	VARCHAR
EMAIL_ADDRESS	VARCHAR
PHONE_NUMBER	VARCHAR
YTD_SALES	DECIMAL
SALESREP_ID	CLOB
NATIONALITY	VARCHAR
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. Click Continue.

#### Open SQL editor

You can run a CREATE VIEW SQL statement in the SQL editor to join your virtual objects. If you create the view in the SQL editor, do not return to the Join virtual objects page and retry your join in the user interface.

Back Continue

### 4. If any Query Customization is required, then we can change here and click on Run All button. For now, Click Back. Click Preview on the Previous screen to see the join data preview.

SQL editor

\* Untitled - 1 Back

```

1 CREATE VIEW view_name
2   AS SELECT "ADMIN"."CUSTOMERS"."CUST_ID" AS "ADMIN_CUSTOMERS_CUST_ID", "ADMIN"."CUSTOMERS"."CUSTNAME" AS "ADMIN_CUSTOMERS_CUSTNAME",
3   "ADMIN"."CUSTOMERS"."CITY" AS "ADMIN_CUSTOMERS_CITY", "ADMIN"."CUSTOMERS"."STATE" AS "ADMIN_CUSTOMERS_STATE",
4   "ADMIN"."CUSTOMERS"."COUNTRY_CODE" AS "ADMIN_CUSTOMERS_COUNTRY_CODE", "ADMIN"."CUSTOMERS"."POSTAL_CODE" AS "ADMIN_CUSTOMERS_POSTAL_CODE",
5   "ADMIN"."CUSTOMERS"."EMAIL_ADDRESS" AS "ADMIN_CUSTOMERS_EMAIL_ADDRESS", "ADMIN"."CUSTOMERS"."PHONE_NUMBER" AS "ADMIN_CUSTOMERS_PHONE_NUMBER",
6   "ADMIN"."CUSTOMERS"."YTD_SALES" AS "ADMIN_CUSTOMERS_YTD_SALES", "ADMIN"."CUSTOMERS".SALESREP_ID AS "ADMIN_CUSTOMERS_SALESREP_ID",
7   "ADMIN"."CUSTOMERS"."NATIONALITY" AS "ADMIN_CUSTOMERS_NATIONALITY", "ADMIN"."CUSTOMERS".NATIONAL_ID AS "ADMIN_CUSTOMERS_NATIONAL_ID",
8   "ADMIN"."CUSTOMERS"."CREDITCARD_NUMBER" AS "ADMIN_CUSTOMERS_CREDITCARD_NUMBER", "ADMIN"."CUSTOMERS"."CREDITCARD_TYPE" AS "ADMIN_CUSTOMERS_CREDITCARD_TYPE",
9   "ADMIN"."CUSTOMERS"."CREDITCARD_EXP" AS "ADMIN_CUSTOMERS_CREDITCARD_EXP", "ADMIN"."CUSTOMERS".CREDITCARD_CVV AS "ADMIN_CUSTOMERS_CREDITCARD_CVV",
10  "ADMIN"."SALES REP"."FIRST_NAME" AS "ADMIN_SALES REP_FIRST_NAME", "ADMIN".SALES REP".LAST_NAME" AS "ADMIN_SALES REP_LAST_NAME",
11  "ADMIN".SALES REP".NATIONALITY" AS "ADMIN_SALES REP_NATIONALITY", "ADMIN".SALES REP".NATIONAL_ID" AS "ADMIN_SALES REP_NATIONAL_ID",
12  "ADMIN".SALES REP".PHONE_NUMBER" AS "ADMIN_SALES REP_PHONE_NUMBER", "ADMIN".SALES REP".AGE" AS "ADMIN_SALES REP_AGE",
13  "ADMIN".SALES REP".SEX" AS "ADMIN_SALES REP_SEX", "ADMIN".SALES REP".TERRITORY" AS "ADMIN_SALES REP_TERRITORY",
14  "ADMIN".SALES REP".EMAIL_ADDRESS" AS "ADMIN_SALES REP_EMAIL_ADDRESS", "ADMIN".SALES REP".MANAGER_ID" AS "ADMIN_SALES REP_MANAGER_ID"
15  FROM "ADMIN".CUSTOMERS, "ADMIN".SALES REP"
16  WHERE "ADMIN".CUSTOMERS.SALESREP_ID="ADMIN".SALES REP".SALESREP_ID

```

### 5. Close the Preview Popup.

The screenshot shows the 'Join virtual objects' step in the data management console. A tooltip is displayed over the 'TERRITORY' column in the preview table, indicating that multiple rows from the 'CUSTOMERS' table are being joined. The preview table contains data from both the 'SALES\_REP' and 'CUSTOMERS' tables.

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

6. **Filters** tab can be used to filter some of the data from this view. For now leave it as default and click **Next**.

The screenshot shows the 'Join virtual objects' interface with the 'Filters' tab selected. It displays filter clauses for the 'CUSTOMERS' and 'SALES\_REP' tables.

Column name	Data type
COUNTRY_CODE	VARCHAR
POSTAL_CODE	VARCHAR
EMAIL_ADDRESS	VARCHAR
PHONE_NUMBER	VARCHAR
YTD_SALES	DECIMAL
SALESREP_ID	VARCHAR
NATIONALITY	VARCHAR
NATIONAL_ID	VARCHAR
CREDITCARD_NUMBER	BIGINT

Column name	Data type
SALESREP_ID	CLOB
FIRST_NAME	CLOB
LAST_NAME	CLOB
NATIONALITY	CLOB
NATIONAL_ID	CLOB
PHONE_NUMBER	CLOB
AGE	CLOB
SEX	CLOB
TERRITORY	CLOB

**Open in SQL editor**

Join keys      **Filters**

Programmatically filter data using simple clause statements like the one below.

```
<SCHEMA>"<TABLE>"<COLUMN>" = <VALUE>
```

**CUSTOMERS**

```
"ADMIN"."CUSTOMERS"."CUST_ID" = <VALUE>
```

100 rows meet the filter criteria

**SALES\_REP**

```
"ADMIN"."SALES_REP"."SALESREP_ID" = <VALUE>
```

7. 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

8. 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 (1)  Project (1)  Virtualized data

Data Virtualization Project

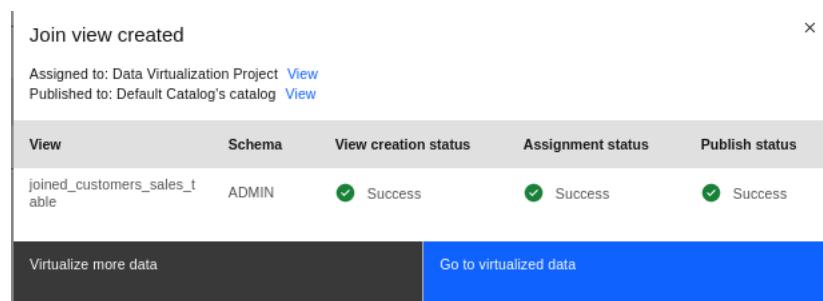
Publish to

Catalog (1)

Default Catalog

View name	Schema name
joined_customers_sales_table	ADMIN

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



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

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

11.Click on the project created earlier.

Name	Date created	Your role	Collaborators
Data Virtualization Project	3 days ago	Admin	1
Banking Demo Project	2 months ago	Admin	1
Datastage Demo Assets	2 months ago	Admin	1

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

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

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

14.Click on **Credentials** in the left navigation pane menu.

Review the connection information.

**Connection overview**

Name: DS16621325392290910

Description: Connection description

**Connection details**

Database\*: BIGSQL

Cancel Save

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

Review the connection information.

**Credentials**

Credential setting: Personal

Authentication method\*

Port is SSL-enabled

SSL certificate

```
-----BEGIN CERTIFICATE-----
MIIDBzCCAegAwIBAgIRAljyOHEkIpGPu/kYybPFWFUwDQYJKoZIhvcNAQELBQA
wHTEBMBkGA1UEAxMSemVulWnhLWNicnRpZmljYXRIMBAXDjTyMDYzMDiMDQyN1oX
DTI1MDY0TlxMDkyN1owHTEBMBkGA1UEAxMSemVulWnhLWNicnRpZmljYXRIMIIB
IANBeknnkiG9wvRA0FFAAOCAB8AMTBCeKCAQFA0TOY5Nkuw3s9H66lII/vCfrx
-----END CERTIFICATE-----
```

Cancel Save

16. Chose User Name and Password.

Review the connection information.

**Credentials**

Credential setting: Personal

Authentication method\*: API key

SSL certificate:

```
-----BEGIN CERTIFICATE-----
MIIDBzCCAf+gAwIBAgIRAIyOHEkIpGPu/kYybPFWFUwDQYJKoZIhvNaQELBQA
...
-----END CERTIFICATE-----
```

Cancel Save

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

Review the connection information.

**Credentials**

Credential setting: Personal

Authentication method\*: Username and password

Username\*: admin

Password\*: [REDACTED]

Certificates

Port is SSL-enabled

SSL certificate:

The test was successful.  
Click Save to update the connection information.

Cancel Save

18. Now Click on the Virtualized Data Asset ADMIN.joined\_customers\_sales\_table.

Projects / Data Virtualization Project

Overview Assets Jobs Manage

Find assets

2 assets

All assets

Name	Last modified
ADMIN.joined_customers_sales_table	11 minutes ago admin (You)
DS16621325392290910	11 minutes ago admin (You)

Add asset New asset

Asset types

Data access: 1

Data: 1

Data in this project

Drop data files here or browse for files to upload

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

The screenshot shows the 'Preview' tab selected in the top navigation bar. The main area displays a table with 26 columns and 10 rows of data. The columns include 'SALESRE...', 'FIRST\_N...', 'LAST\_N...', 'ADMIN\_SALESREP\_NATIONA...', 'ADMIN\_SALESREP\_NATIONA...', 'ADMIN\_SALESPHONE\_NU...', 'AGE', and 'SEX'. The data is as follows:

SALESRE... Club	FIRST_N... Club	LAST_N... Club	ADMIN_SALESREP_NATIONA... Club	ADMIN_SALESREP_NATIONA... Club	ADMIN_SALESPHONE_NU... Club	AGE Club	SEX Club
NC160	Betsy	Adams	UK	124-168-918	229-990-2162	35	F
NC166	Penney	Hayes	FR	516-264-270	630-492-6535	50	F
NC169	Fredericka	King	ES	766-887-613	448-788-1089	88	F
NC169	Fredericka	King	ES	766-887-613	448-788-1089	88	F
NC232	Lakeesha	Jenkins	IT	910-870-499	828-268-4303	21	F
NC232	Lakeesha	Jenkins	IT	910-870-499	828-268-4303	21	F
NC292	Rosalyn	Inman	CA	145-878-802	267-929-9454	35	F
NC364	Teresa	Underhill	ES	173-526-693	717-217-7230	21	F
NC424	Devon	Quigley	CA	321-821-746	933-753-6160	80	M
NC424	Devon	Quigley	CA	321-821-746	933-753-6160	80	M
NE172	Nevada	Brown	IT	446-132-118	546-811-6192	69	M

20. 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 selected in the top navigation bar. The main area displays a table with two assets listed: 'ADMIN.joined\_customers\_sales\_table' and 'DS16621325392290910'. The table columns are 'Name' and 'Last modified'. Both assets were modified 11 minutes ago by 'admin (You)'. On the left sidebar, there are sections for 'Asset types' (Data access, Data), 'All assets' (2 assets), and 'Find assets'.

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

### New asset

Select the tool to create an operational or configuration asset.

The screenshot shows the 'New asset' interface with a search bar at the top. On the left, a sidebar lists 'Tool type' categories: 'All types' (selected), 'Automated builders', 'Graphical builders', 'Code editors', 'Component editors', and 'Data access tools'. The main area displays three tool categories:

- Pipelines**: Automate the model lifecycle, including preparing data, training models, and creating deployments. It includes the 'SPSS Modeler' tool.
- Code editors**: Create a visual flow that uses modeling algorithms to prepare data and build and train a model, using a guided approach to machine learning that doesn't require coding. It includes the 'Federated Learning' and 'Jupyter notebook editor' tools.
- Component editors**: Create a federated learning experiment to train a common model on a set of remote data sources. Share training results without sharing data. It includes the 'DataStage component' and 'Parameter set' tools.

A 'Show descriptions' toggle switch is visible on the left.

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

The screenshot shows the 'New notebook' dialog. At the top, there are three tabs: 'Blank' (selected), 'From file', and 'From URL'. Below the tabs, there are fields for 'Name' (set to 'Data Analysis Notebook') and 'Description (optional)' (set to 'Data Analysis Notebook'). To the right, there is a 'Select runtime' dropdown set to 'IBM Runtime 22.1 on Python 3.9 (1 vCPU, 2 GB RAM)', with a note below it stating 'The selected runtime has 1 vCPU and 2 GB RAM.' Under 'Language', 'Python 3.9' is selected. At the bottom right are 'Cancel' and 'Create' buttons, with 'Create' being highlighted in blue.

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

The screenshot shows the Jupyter Notebook interface. The top navigation bar includes 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', 'Widgets', and 'Help'. The toolbar below the menu bar includes icons for cell creation, run, format, and code. The main area shows a code cell starting with 'In [ ]:'.

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

The screenshot shows the IBM Cloud Pak for Data Data Analysis Notebook. The 'Data' panel on the right is open, showing a file named 'ADMIN.joined\_customers\_sales\_table' selected. The notebook interface includes a toolbar with various icons for file operations, a code editor with an 'In [ ]:' cell, and a status bar indicating 'Not Trusted | Python 3.9' and 'CPU: Memory:218.1 MB / 2 GB'.

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

The screenshot shows the execution of a Python script in the 'In [ ]:' cell. The script imports 'itc\_utils.flight\_service' and reads data from 'ADMIN.joined\_customers\_sales\_table'. It then displays a warning about a row limit of 5000. The output shows the execution of the script and the resulting DataFrame.

```

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.

The screenshot shows the execution of the same Python script in the 'In [2:]' cell. The output shows the resulting DataFrame 'data\_df\_1' with 6 rows. The columns include SALESREP\_ID, FIRST\_NAME, LAST\_NAME, ADMIN\_SALES REP NATIONALITY, ADMIN\_SALES REP NATIONAL ID, ADMIN\_SALES REP PHONE NUMBER, AGE, SEX, TERRITORY, and AL.

```

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 r
# 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)

Out[2]:
   SALESREP_ID FIRST_NAME LAST_NAME ADMIN_SALES REP NATIONALITY ADMIN_SALES REP NATIONAL ID ADMIN_SALES REP PHONE NUMBER AGE SEX TERRITORY AL
0           NC160      Betsy     Adams          UK            124-168-918  229-990-2162   35   F NorthCentral   Bi
1           NC166     Penney     Hayes          FR            516-264-270  630-492-6535   50   F NorthCentral   Per
2           NC169  Fredericka     King          ES            766-887-613  448-788-1089   88   F NorthCentral  Fred
3           NC169  Fredericka     King          ES            766-887-613  448-788-1089   88   F NorthCentral  Fred
4           NC232     Lakeshia  Jenkins          IT            910-870-499  828-268-4303   21   F NorthCentral Lakeer
5           NC232     Lakeshia  Jenkins          IT            910-870-499  828-268-4303   21   F NorthCentral Lakeer

```

## Visualize the data with a dashboard

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

The screenshot shows the 'Assets' tab selected in the navigation bar. On the left, there's a sidebar with sections for '3 assets' (All assets), 'Asset types' (Data access, Data, Source Code), and a search bar. The main area displays a table titled 'All assets' with columns for 'Name' and 'Last modified'. Three assets are listed: 'Data Analysis Notebook' (Notebook), 'ADMIN.joined\_customers\_sales\_table' (application/octet-stream), and 'DS16621325392290910' (Connection). A right-hand panel titled 'Data in this project' contains a dashed box for file uploads.

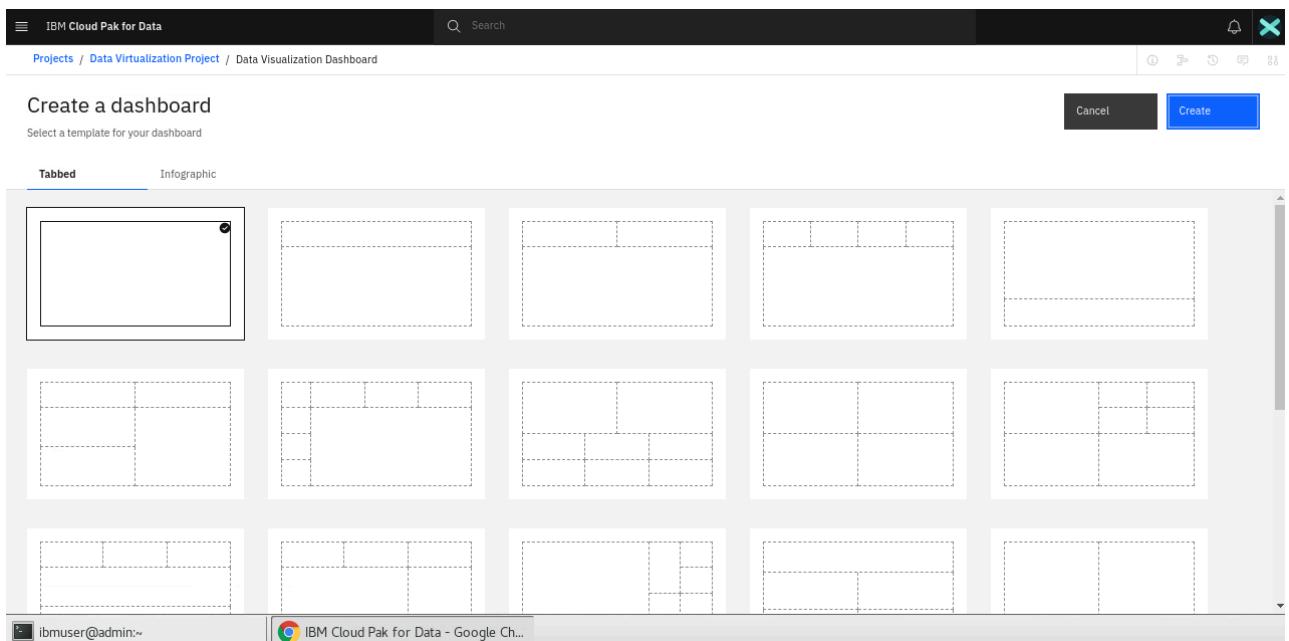
## 2) Click on Dashboard Editor.

This screenshot shows the 'New asset' dialog. In the 'Tool type' section, 'All types' is selected. Under 'Automated builders', two options are shown: 'AutoAI' (analyze tabular data) and 'Metadata enrichment' (enrich imported asset metadata). Other tool types like 'Graphical builders', 'Code editors', 'Component editors', and 'Data access tools' are also listed but not selected.

## 3) Enter the name and description and click Create.

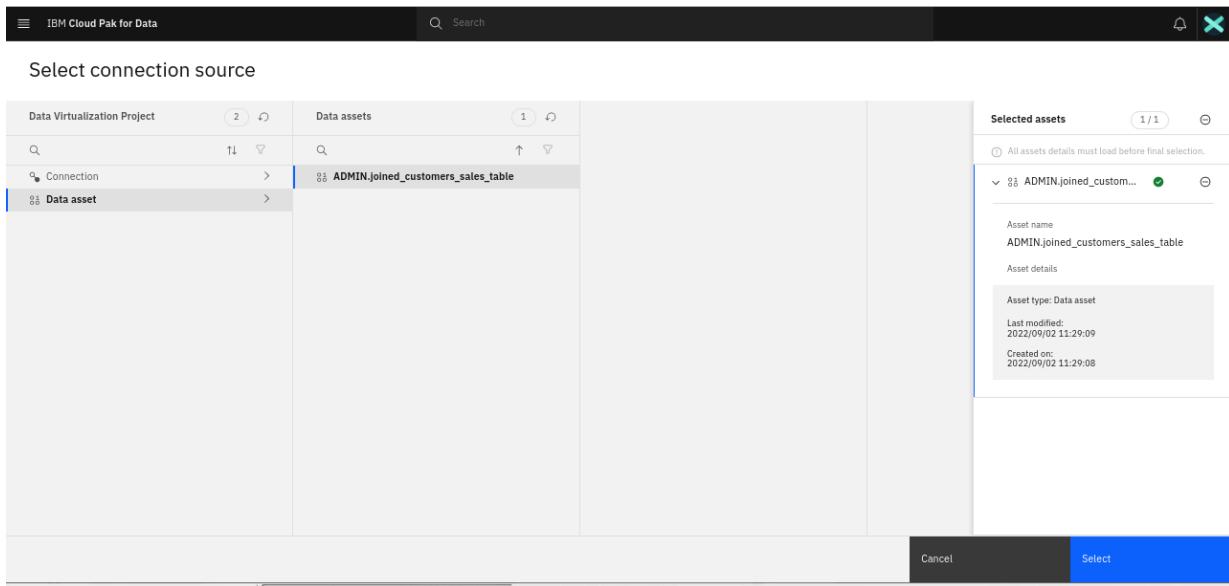
This screenshot shows the 'Create a dashboard' dialog. In the 'Define details' section, the 'Name' field is filled with 'Data Visualization Dashboard'. The 'Description (Optional)' field contains the text 'Data Visualization Dashboard'. At the bottom right, there are 'Cancel' and 'Create' buttons.

## 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**.

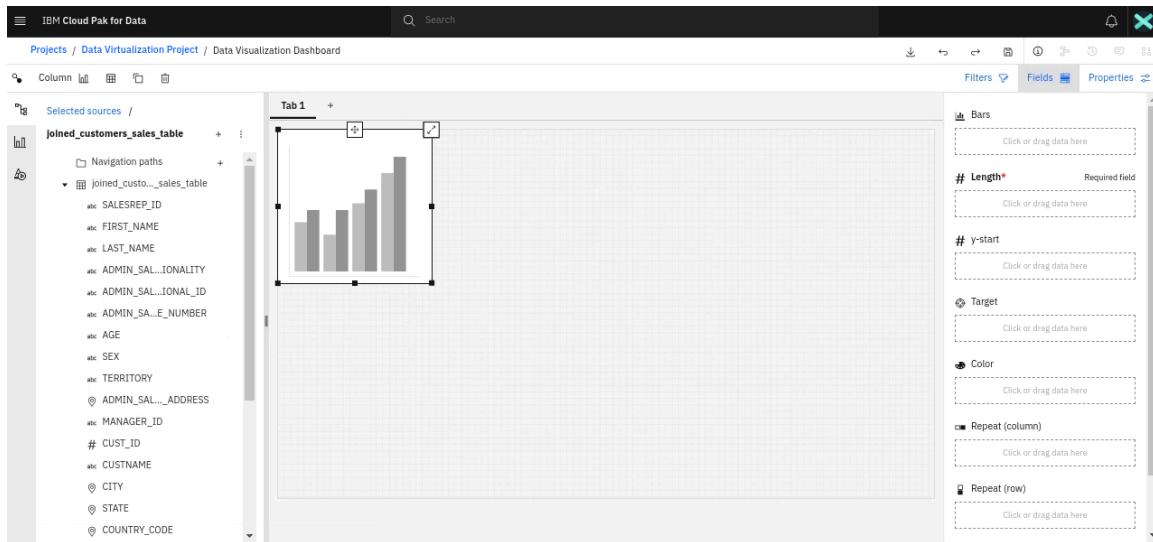


- 7) The Data source will be added successfully. Click on **Visualization Tab**  to see the types of Visualization to be used.

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

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.



## 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.

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

**Data sources** 3 items

- Data Virtualization** IBM  
Query many data sources as one.  
Enabled ✓
- Db2** IBM  
Relational database that delivers advanced data management and analytics capabilities for transactional and warehousing workloads.  
Available ⓘ Enabled ✓
- RStudio Server with R3.6** Partner  
Optional development environment for

**Developer tools** 1 item

3. Click **New Instance**.

Db2

Type Database

Version 11.5.7.0-cn5-x86\_64

Provider IBM

Category Data sources

Related links Docs

**Summary**

Relational database that delivers advanced data management and analytics capabilities for transactional and warehousing workloads.

New instance +

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

Create a database

**Configure**

Database name Customer

Number of nodes 1

CPU per node for Db2 2.1

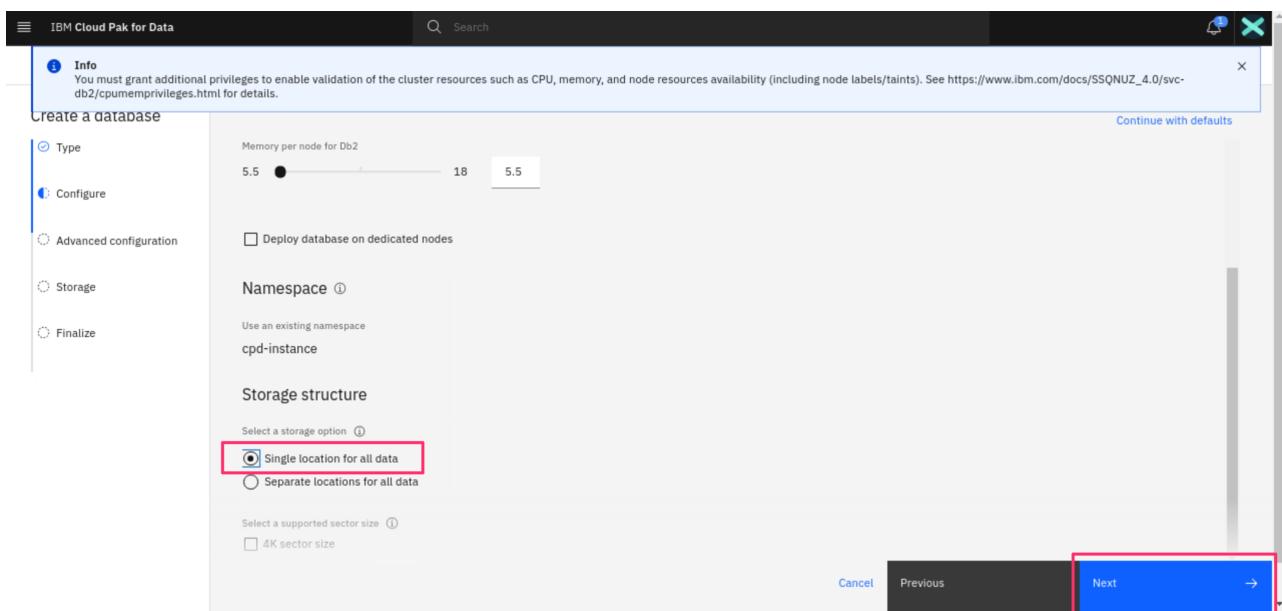
Memory per node for Db2 5.5

Deploy database on dedicated nodes

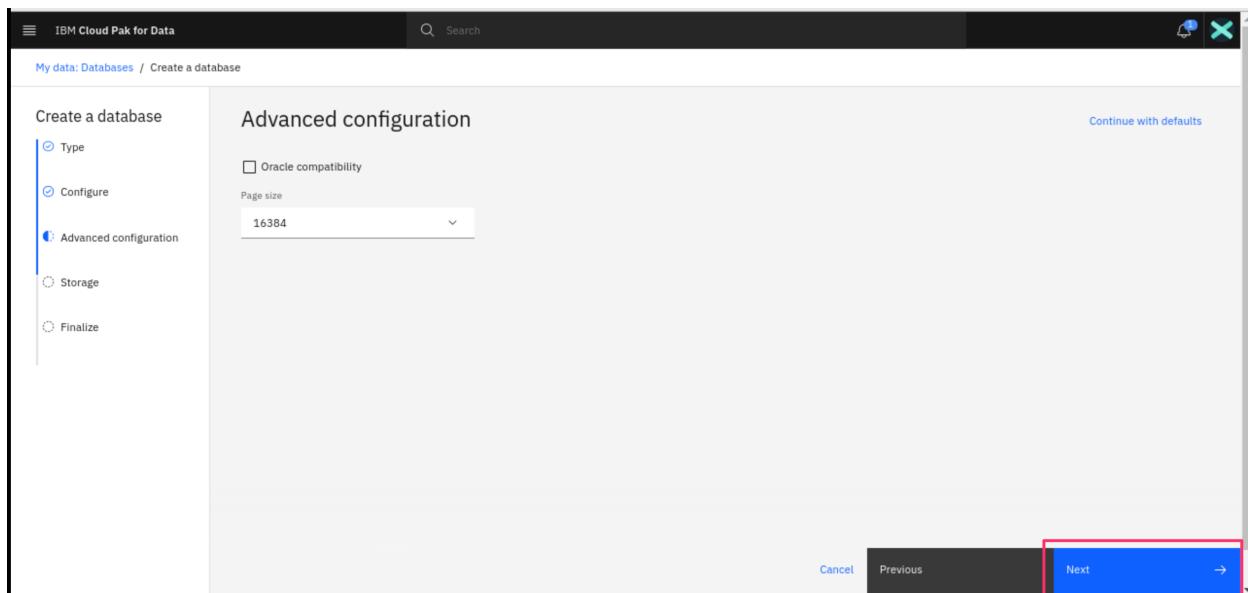
Namespace

Cancel Previous Next

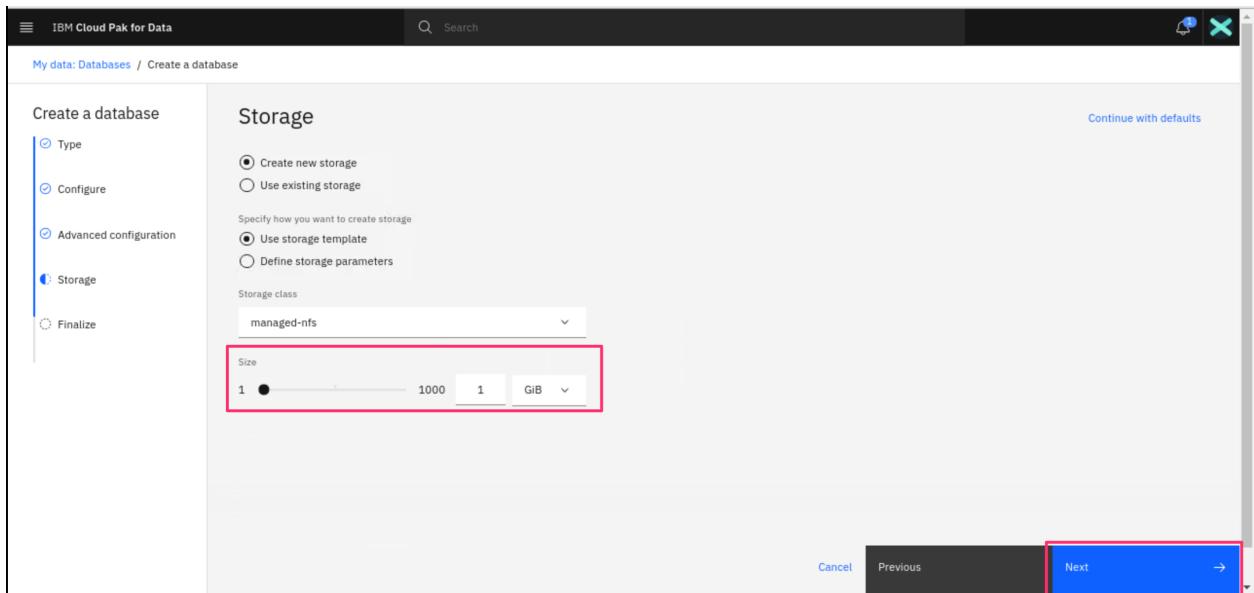
#### 5. Select **Single Location for all Data** Option and click **Next**.



6. Click **Next** on the advanced configuration.



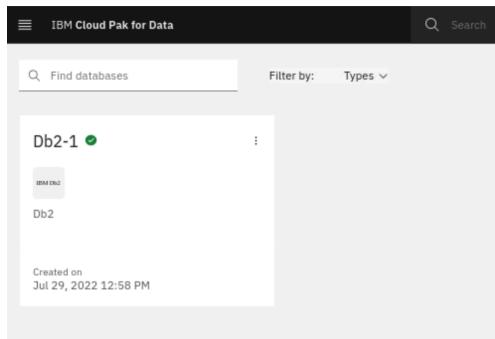
7. Configure the Storage Size and Click **Next**.



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								
Name	Type	Created by	vCPU requests	Memory requests (GiB)	Users	Status	Created on	
<a href="#">Db2-1</a> Service instance for db2oltp-16591138850024...	db2oltp	admin	2.20	5.75 Gi	1	<span>Green</span>	Jul 29, 2022	<span>⋮</span>
<a href="#">ds-px-default</a> The default DataStage runtime instance	datastage	admin	2.50	6.00 Gi	1	<span>Green</span>	Jul 12, 2022	<span>⋮</span>
<a href="#">data-virtualization</a>	dv	admin	11.50	38.50 Gi	1	<span>Green</span>	Jul 6, 2022	<span>⋮</span>
<a href="#">data-management-console</a> Data Management Console	dmc	admin	4.20	10.62 Gi	1	<span>Green</span>	Jul 6, 2022	<span>⋮</span>
<a href="#">openscale-defaultinstance</a> IBM Watson OpenScale	aios	admin	0.00	0.00 Gi	1	<span>Green</span>	Jul 1, 2022	<span>⋮</span>