

# IBM CLOUD PAK FOR DATA 4.5

## DATA VIRTUALIZATION HANDS ON LAB

Sandeep Ved

CUSTOMER SUCCESSS MANAGER , IBM SINGAPORE

## Table of Contents

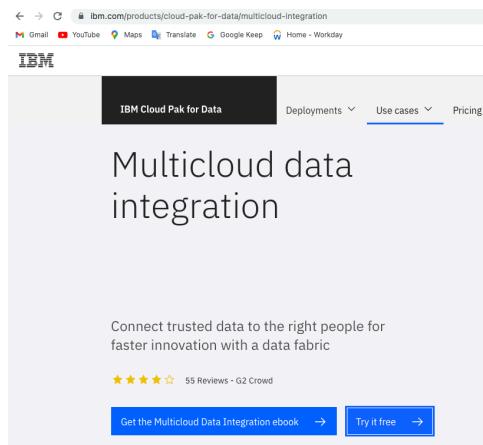
<b>Data Virtualization Lab Prerequisites.....</b>	<b>2</b>
Sign up for CP4D as a Service .....	2
Generate an IBM API key.....	3
Provision the necessary services.....	4
Watson Query (Data Virtualization) – Mandatory .....	4
Watson Studio - Optional.....	5
IBM Cognos Dashboard Embedded – Optional.....	7
<b>Data Virtualization Lab – Using Sample Project &amp; Connections.....</b>	<b>8</b>
Task 1: Use or Create the sample project .....	8
Task 2: Create a Platform assets catalog .....	10
Task 3: Add data connections to the Platform assets catalog .....	11
Task 4: Add data sources to Watson Query.....	13
Task 5: Virtualize data tables .....	14
Task 6: Create a virtual join view by joining virtual tables.....	17
Virtual join view 1: Join MORTGAGE_APPLICANT & MORTGAGE_APPLICATION virtual tables .....	17
Virtual join view 2: Join APPLICANTS_APPLICATIONS_JOINED & CREDIT_SCORE virtual tables .....	22
Task 7: Access the virtual join view in the project .....	24
Task 8: Use Cognos Dashboard (Optional).....	27
Task 9: Use Jupyter Notebook Dashboard (Optional).....	30
<b>Data Virtualization Lab – Using New Project &amp; Connections.....</b>	<b>32</b>
Task 1: Provision the required services .....	32
Task 2: Load data into your Db2 data source.....	35
Task 3: Create a project .....	41
Task 4a: Add a connection to your Db2 data source .....	44
Task 4b: Add a connection to a PostgreSQL data source.....	47
Task 5: Add tables to your virtualized data .....	50
Task 6a: Publish virtualized data to a catalog and project .....	53
Task 6b: Publish virtualized data to a catalog and project .....	57
Task 7: Update IBM API Keys in Project Connection.....	59
Task 7: Next Actions (Optional).....	61

# Data Virtualization Lab Prerequisites

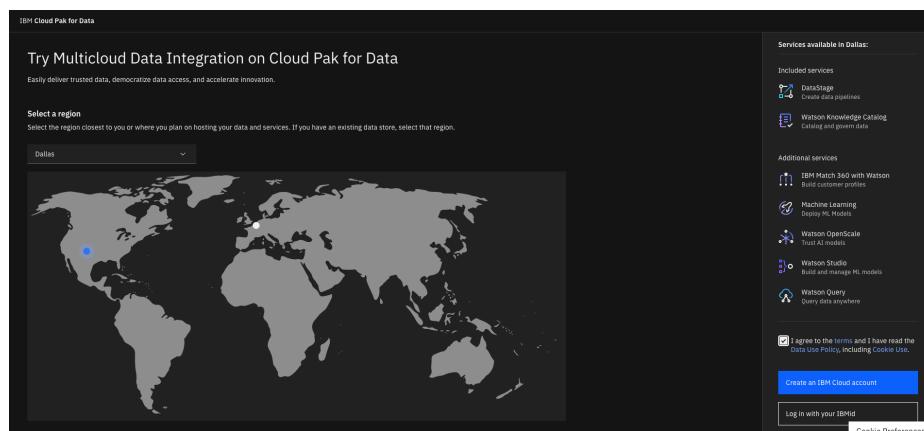
## Sign up for CP4D as a Service

You must sign up for Cloud Pak for Data as a Service and provision the necessary services for the Data Virtualization use case.

- If you have an existing Cloud Pak for Data as a Service account, then you can get started with this tutorial. If you have a Lite plan account, only one user per account can run this tutorial.
- If you don't have a Cloud Pak for Data as a Service account yet, then [sign up for a data fabric trial](#).
- Navigate to <https://www.ibm.com/products/cloud-pak-for-data/multicloud-integration> and click on Try it free.



- Select a region and Agree to terms and condition. Either Create an IBM ID or login with existing IBM Id. You can choose [Dallas](#) to get access to most of the IBM services and plans.



- You will see the IBM CP4D home page.

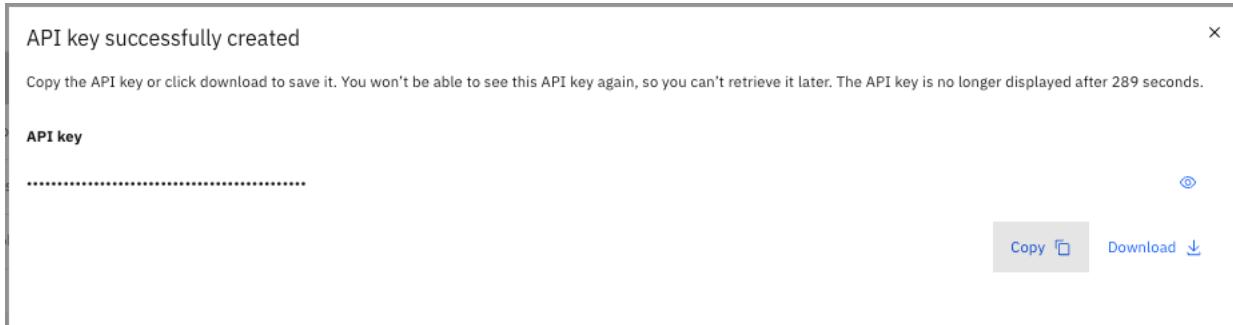
## Generate an IBM API key

You need to provide your personal credentials in the form of an API key to view virtualized assets. If you don't already have a saved API key, then follow these steps to create an API key.

1. Access the [API keys page](#) in the IBM Cloud console. Log in if prompted.
2. On the *API keys* page, click **Create an IBM Cloud API key**.

3. Type a name and description. Click **Create**.

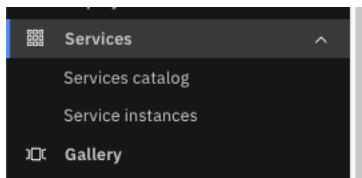
4. **Copy** the API key.



5. Download the API key for future use. With the use of IBM API keys, you are now ready to view the virtual table in the project.

## Provision the necessary services

From the Cloud Pak for Data navigation menu , choose **Services > Service instances**.



Verify if the service instance exists. If not, click on **Add Service** button to provision the necessary service as below.

Name	Group	Location	Product	Plan	Status
CloudObjectStorage	Default	Global	Cloud Object Storage	Lite	Active
DataStage-ia	Default	Dallas	DataStage	Lite	Active
Db2-qb	Default	Dallas	Db2	Lite	Active
IBM Cognos Dashboard Embedded-am	Default	Dallas	IBM Cognos Dashboard Embedded	Lite	Active
WatsonMachineLearning	Default	Dallas	Machine Learning	Lite	Active
WatsonKnowledgeCatalog	Default	Dallas	Watson Knowledge Catalog	Lite	Active
WatsonOpenScale	Default	Dallas	Watson OpenScale	Lite	Active
WatsonQuery-eg	Default	Dallas	Watson Query	Lite	Active
WatsonStudio	Default	Dallas	Watson Studio	Lite	Active

### Watson Query (Data Virtualization) – Mandatory

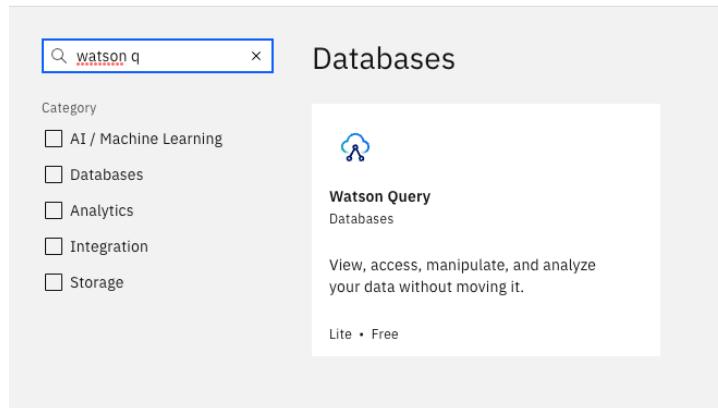
With the [Watson Query](#) service, you can connect to multiple data sources, create and govern virtual assets to segment or combine data from one or more tables, and consume the virtualized data.

- **Connect:** Start by connecting to data sources. You can connect to multiple data sources into a single self-balancing collection of data sources or databases. For more information, see [Adding and connecting to data sources in Watson Query](#) and [Supported data sources in Watson Query](#).
- **Join, create, and govern:** Then, create virtual tables, group tables by schema, associate data with projects, and govern your virtual assets. For more information, see [Creating virtualized objects](#) and [Governing virtual data in Watson Query](#).
- **Consume:** Finally, consume virtual tables in analytics projects, dashboards, data catalogs, and other applications. For more information, see [Dashboard services](#).

## IBM Data Virtualization

1. Check if an existing Watson Query service instance exists. If you need to create a Watson Query service instance, click **Add service**. Select **Watson Query**.

### Services catalog



2. Select the free **Lite** plan. Click **Create** to provision the service.

This screenshot shows the 'Watson Query' service creation page. The 'Create' tab is selected. It displays the service's purpose as a Cloud Pak for Data as a Service, its minimum configuration, and the fact that instances expire after 30 days. The 'Lite' plan is highlighted as a 'Free dedicated Lite instance'. On the right, a summary panel shows the service is in the 'Tokyo' region, on the 'Watson Query' plan, with the service name 'Watson Query-bg' and resource group 'Default'. At the bottom, there are 'Create' and 'View terms' buttons.

3. Wait while the Watson Query service is provisioned, which might take a few minutes to complete and listed under service instances.

### Watson Studio - Optional

Watson Studio provides the environment and tools for you to collaboratively work on data to solve your business problems. You can choose the tools you need to analyze and visualize data, to cleanse and shape data, to ingest streaming data, or to create and train machine learning models.

The architecture of Watson Studio is centered around the analytics project. Data scientists and business analysts use analytics projects to organize resources and analyze data.

These tools are included with the Watson Studio service:

## IBM Data Virtualization

- Data Refinery: Prepare and visualize data.
- Jupyter notebook editor: Code Jupyter notebooks.
- JupyterLab IDE: Code Jupyter notebooks and Python scripts with Git integration. Other project tools require additional services. See the lists of supplemental and related services.

1. Check if an existing Watson Studio service instance exists. If you need to create a Watson Studio service instance, click **Add service**. Select **Watson Studio**.

2. Select the free **Lite** plan. Click **Create** to provision the service.

3. Wait while the Watson Studio service is provisioned, which might take a few minutes to complete and listed under service instances.

## IBM Cognos Dashboard Embedded – Optional

You can use the dashboard editor in projects to build sophisticated visualizations of your analytics results, and communicate the insights that you've discovered in your data on a dashboard.

**Data format** CSV files with comma (,) delimiter Data in tables in Db2, Db2 on Cloud, Db2 Warehouse, IBM Cloud Databases for PostgreSQL, PostgreSQL, and Microsoft SQL Server

**Important:** *Data sources that are for file storage, for example, IBM Cloud Object Storage or Box, are not supported.*

**Data size** Any; columns in CSV files can only be 128 characters long

1. Check if an existing **IBM Cognos Dashboard Embedded** service instance exists. If you need to create a new instance, click **Add service**.
  1. Select **IBM Cognos Dashboard Embedded**.

The screenshot shows the IBM Cloud Pak for Data interface. At the top, there's a navigation bar with 'IBM Cloud Pak for Data', a search bar, and user account information. Below it is a 'Services catalog' section. A search bar contains the text 'cognos'. To the right of the search bar, the word 'Analytics' is displayed. Underneath, there's a list of categories: 'Category' with options like 'AI / Machine Learning', 'Databases', 'Analytics' (which is selected), 'Integration', and 'Storage'. To the right of this list is a detailed card for 'IBM Cognos Dashboard Embedded' under the 'Analytics' category. The card includes a small icon, the service name, its category, a brief description ('Build and deliver visually stunning dashboards that accelerate your journey to a data driven business.'), and a note that it's 'Lite • Free'.

2. Select the free **Lite** plan. Click **Create** to provision the service.

## IBM Data Virtualization

The screenshot shows the IBM Cloud Pak for Data service catalog. A search bar at the top has 'Search in your workspaces' and a magnifying glass icon. On the right, there are buttons for 'Buy', 'Help', 'Bell', and a user account dropdown for 'SANDEEP VED's Account' with 'Dallas' selected. Below the search bar, a link 'Services catalog /' is visible.

The main content area displays the 'IBM Cognos Dashboard Embedded' service. It includes a 'Create' button and an 'About' tab. A section titled 'Select a region' shows a dropdown menu set to 'Dallas'. Below this is a 'Pricing plan' section stating 'Displayed prices do not include tax. Monthly prices shown are for country or region: United States'. It lists two plans:

Plan	Features	Pricing
Lite	50 sessions/month A session is a 60 minute period where end-users can perform unlimited interactions with an embedded dashboard. Lite plan services are deleted after 30 days of inactivity.	Free
Pay as you go	After 50 sessions Live connection to underlying data Embed dashboards where users are without losing interactivity Smart Creation of Visualizations Interactive exploration of data through filtering and navigation methods	\$0.05 USD/Session

On the right side of the pricing table are 'Create' and 'View terms' buttons. The background of the page is dark grey.

2. Wait while the **IBM Cognos Dashboard Embedded** service is provisioned, which might take a few minutes to complete and listed under service instances.

# Data Virtualization Lab – Using Sample Project & Connections

Use Watson Query to combine data from different data sources and with different types in the given **sample project with existing Connections**. Use SQL syntax and access & combined data without data movement.

## Task 1: Use or Create the sample project

You would already have the sample project “**Multicloud Data Integration**” for this tutorial. So you can skip to next Task. If not or if you deleted the sample project, you can follow these steps to add it.

1. Access the [Multicloud data integration guided tutorial sample project](#) in the gallery. Click **Create project**.

The screenshot shows the Multicloud Data Integration project details page in the gallery. At the top, there is a back arrow, the project name 'Multicloud Data Integration', and a 'Create project' button. The page includes sections for 'Tags' (Data fabric, Guided tutorial), 'Required Services' (0), and 'Modified' (Jul 12, 2022). Below this is a large text area containing project instructions and a detailed diagram of a data pipeline.

**Tutorials included in this project:** Integrate your data by creating a data pipeline with DataStage to transform data. Virtualize external data by using Watson Query to combine data from different data sources.

Golden Bank must comply to a new regulation that prohibits lending to under-qualified mortgage applicants. Your job is to create a data pipeline that delivers concise, pre-processed, and current data on mortgage applicants to the lending department using data from 3 different data sources: DB2 Warehouse, PostgreSQL, and MongoDB.

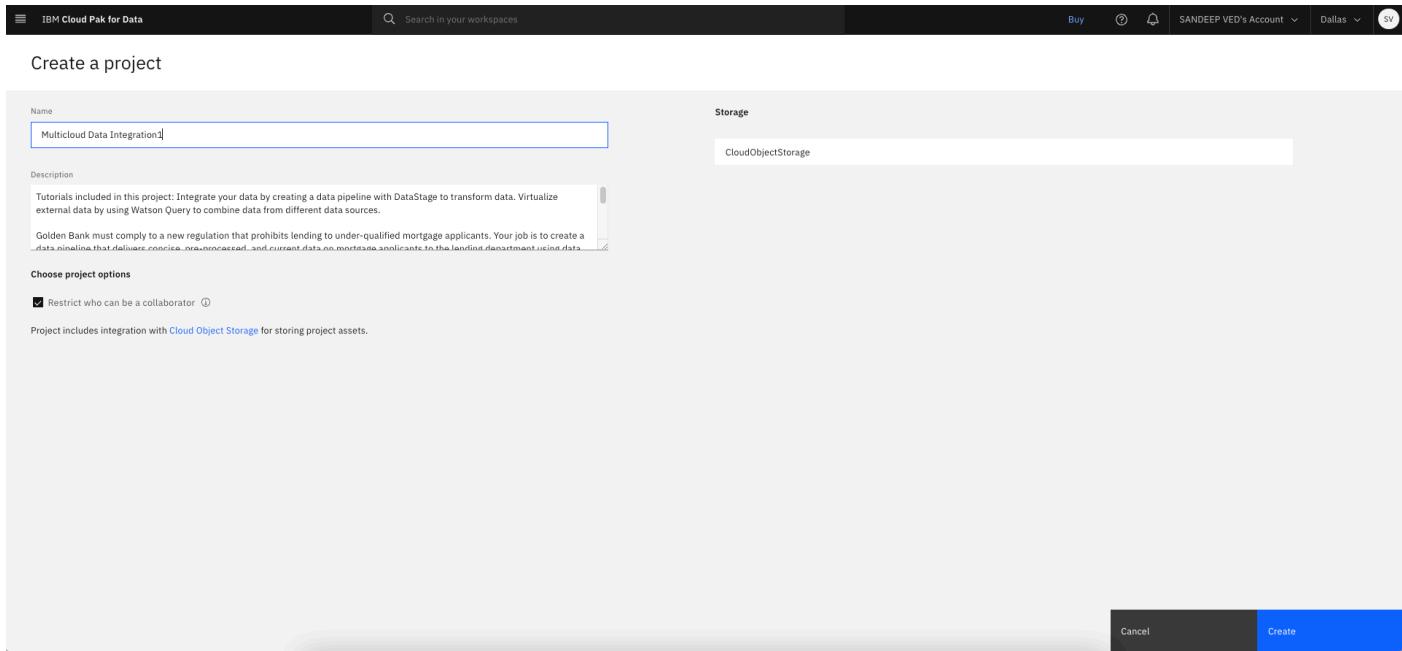
To get started with this tutorial, click **Create project**. In the project, you'll see the list of required services and the tutorial instructions:

- [Integrate data](#)
- [Virtualize external data](#)

The diagram on the right shows a complex data pipeline with multiple stages of data transformation and integration. It includes nodes for DataStage, Watson Query, and various databases like DB2, PostgreSQL, and MongoDB. Arrows indicate the flow of data between these components.

## IBM Data Virtualization

2. Provide the name and description of the project. Associate the project to a Cloud Object Storage instance, select a Cloud Object Storage instance from the list. Click **Create**.



3. Click **View new project** to verify that the project and assets were created successfully.
4. Click the **Assets** tab to see the connections and DataStage flow.

**Note:** You might see a guided tour showing the tutorials that are included with this use case. The links in the guided tour will open these tutorial instructions.

The following image shows the Assets tab in the sample project. You are now ready to start the tutorial.

The screenshot shows the 'Assets' tab in the 'Multicloud Data Integration' project. The left sidebar shows 'Overview', 'Assets' (selected), 'Jobs', and 'Manage'. Under 'Assets', there's a search bar and a table titled 'All assets' with columns 'Name' and 'Last modified'. The table lists four assets:

- Multicloud Data Integration - DataStage flow (Now, Service)
- Data Fabric Trial - MongoDB Connection (1 minute ago, Service)
- Data Fabric Trial - Db2 Warehouse Connection (1 minute ago, Service)
- Data Fabric Trial - Databases for PostgreSQL Connection (1 minute ago, Service)

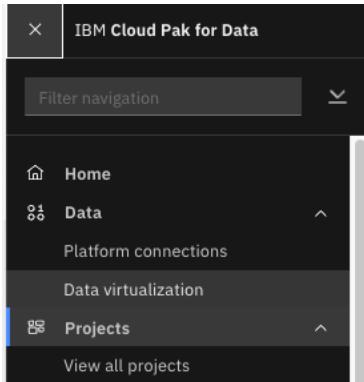
A sidebar on the left shows 'Asset types' with 'Data access' (3 items) and 'Flows' (1 item). On the right, there's a 'Data in this project' section with a 'Drop data files here or browse for files to upload' area. At the bottom, there are pagination controls for 'Items per page: 20' and '1–4 of 4 items'.

**Tip:** If you encounter a guided tour while you are completing this tutorial in the Cloud Pak for Data as a Service user interface, click **Maybe later** or close out the tour window.

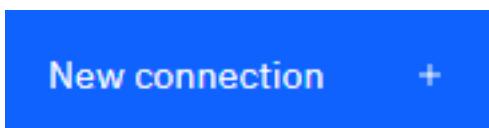
## Task 2: Create a Platform assets catalog

You can add connections to external data sources at either the platform level or the service level. When you add the connections at the platform level by using the Platform assets catalog, you can easily include those connections in projects, catalogs, and Watson Query data sources. Follow these steps to create a Platform assets catalog.

- From the Cloud Pak for Data navigation menu , choose **Data > Platform connections**.



- If you see existing connections, then you already have a Platform assets catalog, and you can skip to Next Task. If you don't see any connections, but you see an option to create a new connection, then also you can skip to Next Task.



If you don't have a Platform assets catalog, click **Create catalog**.

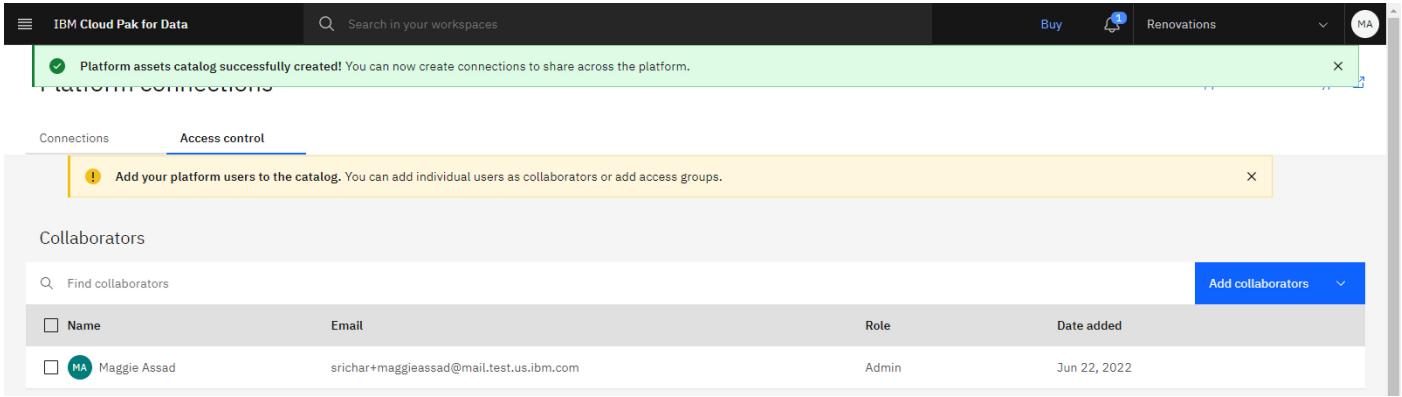
The screenshot shows the 'Platform connections' page. At the top, there's a blue header bar with the text 'New connection' and a '+' sign. Below this, the main title is 'Platform connections'. There's a large, light-grey rectangular area containing a small icon of a cube and a white square. Below the icon, the text reads 'Share connections in the Platform assets catalog' and 'To share platform-level connections with users on the platform, create the Platform assets catalog. Then, add your platform users and your platform level connections to the catalog.' A blue link 'Learn more' is present. At the bottom of this area is a blue button labeled 'Create catalog' with a gear icon next to it. The rest of the page is mostly empty white space.

- Select a Cloud Object Storage from the list.
- Accept the default value for *Duplicate asset handling*.
- Click **Create**. The *Platform connections* page displays.

The following image shows the Access control tab in the Platform assets catalog. On this tab, you can add collaborators. On the Connections tab, you can create connections. Since the

## IBM Data Virtualization

sample project includes the connections, you can add the connections for the external data sources to this catalog from the sample project.



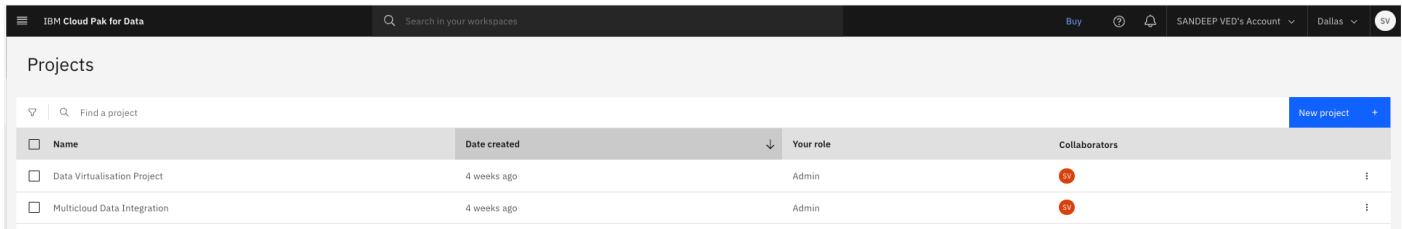
The screenshot shows the IBM Cloud Pak for Data interface. At the top, there's a navigation bar with 'IBM Cloud Pak for Data', a search bar, and various icons including 'Buy', 'Notifications' (with 1 notification), 'Renovations', and a user profile icon. A green banner at the top indicates that the 'Platform assets catalog successfully created! You can now create connections to share across the platform.' Below this, there are two tabs: 'Connections' and 'Access control'. The 'Access control' tab is selected, showing a message: 'Add your platform users to the catalog. You can add individual users as collaborators or add access groups.' Under the 'Collaborators' section, there's a table with columns: Name, Email, Role, and Date added. One row is listed: Maggie Assad (srchar+maggieassad@mail.test.us.ibm.com) with Admin role and Jun 22, 2022 date.

## Task 3: Add data connections to the Platform assets catalog

The Multicloud data integration sample project includes three connections to external data sources. Next, you add those connections to the Platform assets catalog, and then you can make these connections available in Watson Query. Follow these steps to publish the connections from the sample project to the Platform assets catalog.

1. From the Cloud Pak for Data navigation menu , choose **Projects > View all projects**.
2. Click the **Multicloud data integration** project.

**Note:** You might see a guided tour showing the tutorials that are included with this use case. The links in the guided tour will open these tutorial instructions. If you already have the tutorial instructions open in a new window, then close out this guided tour.



The screenshot shows the 'Projects' page in the IBM Cloud Pak for Data interface. The top navigation bar is visible with 'IBM Cloud Pak for Data', a search bar, and user account information ('SANDEEP VED's Account', 'Dallas'). Below the navigation is a table titled 'Projects' with columns: Name, Date created, Your role, and Collaborators. Two projects are listed: 'Data Virtualisation Project' (created 4 weeks ago, Admin role, 20 collaborators) and 'Multicloud Data Integration' (created 4 weeks ago, Admin role, 59 collaborators). A 'New project +' button is located in the top right corner of the table header.

3. Click the **Assets** tab.
4. Under **Asset types**, click **Data access > Connection**.
5. Select the following connection assets:
  - o **Data Fabric Trial - Db2 Warehouse**
  - o **Data Fabric Trial - MongoDB**
  - o **Data Fabric Trial - Databases for PostgreSQL**
6. Click **Publish to catalog**.

## IBM Data Virtualization

The screenshot shows the 'Assets' tab in the IBM Cloud Pak for Data interface. On the left, there's a sidebar with categories like 'Data access', 'Connections', 'Data', 'Flows', 'Visualizations', and 'Source Code'. The 'Connections' section is currently selected. In the main area, there's a table of connections with columns for 'Name', 'Created by', and 'Last modified'. One connection, 'Data Fabric Trial - Databases for PostgreSQL', has a context menu open with the 'Publish to catalog' option highlighted.

### 7. For the Target, select **Platform assets catalog**. Click **Publish**.

This screenshot shows the 'Publish to catalog' dialog box. It includes fields for 'Target' (set to 'Platform assets catalog'), 'Selected assets (1)' (listing a single asset), 'Tags' (empty), 'Description' (placeholder text), 'Duplicate action' (radio buttons for Update, Overwrite, Add duplicate, Preserve, with 'Update' selected), 'Privacy' (radio buttons for Public, Private, with 'Public' selected), and 'Cancel' and 'Publish' buttons at the bottom.

### 8. From the Cloud Pak for Data navigation menu ☰ choose **Data > Platform connections** to see the three connections that are published to the catalog.

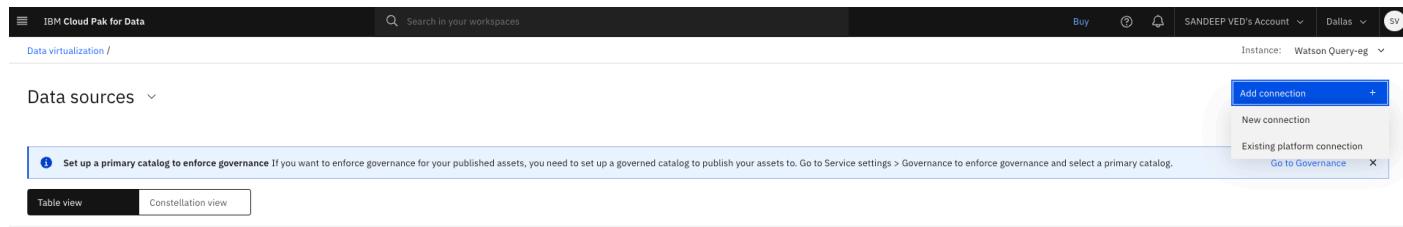
The following image shows the Platform assets catalog with three connections. You are now ready to add data sources to Watson Query.

This screenshot shows the 'Platform connections' page. It features a 'Connections' tab and an 'Access control' tab. Below the tabs, it says 'Connected data sources' and 'Manage existing data source connections or create new connections that can be used across the platform.' There's a 'Find connections' search bar and a 'New connection' button. The main area displays a table of connections with columns for 'Name', 'Type', 'Created by', 'Modified by', and 'Last updated'. Three connections are listed: 'Data Fabric Trial - Mongo DB' (Type: Databases for MongoDB), 'Data Fabric Trial - Databases for Postgre' (Type: Databases for PostgreSQL), and 'Fabric Trial - Db2 Warehouse' (Type: Db2 Warehouse). All connections were created and modified by 'Maggie Assad' on June 22, 2022.

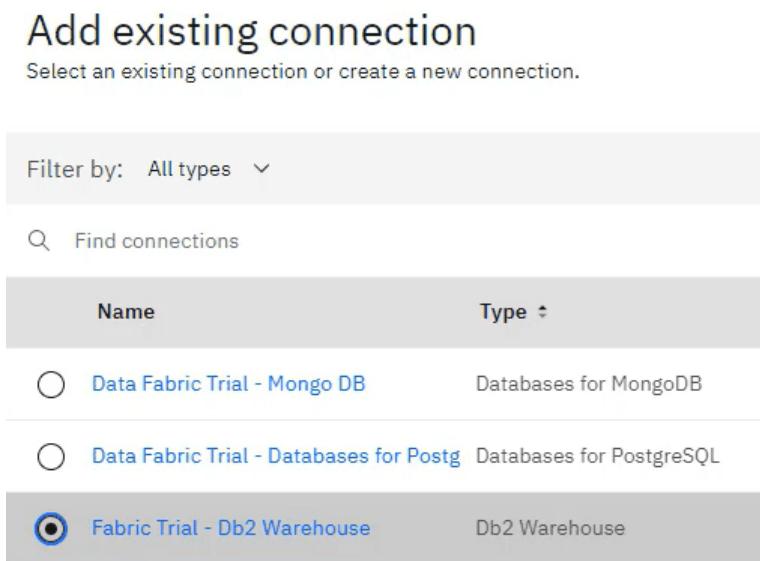
## Task 4: Add data sources to Watson Query

Now you can add these external data sources from the Platform assets catalog to Watson Query. Follow these steps to add the data sources:

1. From the Cloud Pak for Data navigation menu , choose **Data > Data virtualization**.
2. If you see a notification to *Set up a primary catalog to enforce governance*, you can safely close this notification. Setting up a primary catalog is optional.
3. On the *Data sources* page, in the *Table view*, click **Add connection > Existing connection**.



4. Select **Data Fabric Trial - Db2 Warehouse**. Click **Add**.



Name	Type
Data Fabric Trial - Mongo DB	Databases for MongoDB
Data Fabric Trial - Databases for Postg	Databases for PostgreSQL
<b>Fabric Trial - Db2 Warehouse</b>	Db2 Warehouse

5. Repeat these steps to add the **Data Fabric Trial - Mongo DB** and **Data Fabric Trial - Databases for PostgreSQL** connections.

The following image shows the Watson Query data sources. You are now ready to create a virtual table from data stored in those external data sources.

IBM Data Virtualization

Name	Endpoint	Type	Status
Data Fabric Trial - Databases for PostgreSQL	39c6f420-703f-487a-99261a7950626f0e5.blijti...	PostgreSQL	Active
Data Fabric Trial - Mongo DB	f8af5a63-e4d1-44e6-b92c-da4e1d3a6e7d-0.01...	Mongo DB	Active
Fabric Trial - Db2 Warehouse	db2w-rugbyab.us-south.db2w.cloud.ibm.com: 50...	Db2 Family	Active

## Task 5: Virtualize data tables

You want to virtualize the ***MORTGAGE\_APPLICATION***, ***MORTGAGE\_APPLICANT***, and ***CREDIT\_SCORES*** tables. Later, you can join the first two virtual tables with the third table to create a new virtual join view. Follow these steps to virtualize the data tables:

1. From the *Data sources* menu, click **Virtualization > Virtualize**.



2. Wait while the tables load, which might take up to 30 seconds. When you see *Available tables*, then all of the tables loaded.

Table	Business terms	Schema	Connection name
AUDIT		AUDIT	Data Fabric Trial - Db2 ...
TEMPLATE_CONFIG		IBMCORSOLE	Data Fabric Trial - Db2 ...

3. On the *Tables* tab, filter the tables based on the following criteria:
  1. *Source type*: **Db2 Family** and **PostgreSQL**
  2. *Database*: **Data Fabric Trial - Db2 Warehouse** and **Data Fabric Trial - Databases for PostgreSQL**

3. Schema: **BANKING**
4. Select the **MORTGAGE\_APPLICATION**, **MORTGAGE\_APPLICANT**, and **CREDIT\_SCORE** tables to virtualize. You can hover over a table name to see the full name. Careful **not** to select the tables **MORTGAGE\_APPLICATIONS** and **MORTGAGE\_APPLICANTS**.
5. Click **Add to cart**.

Table	Business terms	Schema	Connection	Hostname: Port	Database	Columns
<input type="checkbox"/> MORTGAGE_APPLICAT...	-	BANKING	Data Fabric Trial - Db2 ...	db2w-rugbyab.us-sout...	BLUDB	10
<input type="checkbox"/> MORTGAGE_APPLICA...	-	BANKING	Data Fabric Trial - Db2 ...	db2w-rugbyab.us-sout...	BLUDB	14
<input checked="" type="checkbox"/> MORTGAGE_APPLICAT...	-	BANKING	Data Fabric Trial - Db2 ...	db2w-rugbyab.us-sout...	BLUDB	12
<input type="checkbox"/> MORTGAGE_APPLICATION	-	BANKING	Data Fabric Trial - Data...	39c6f420-703f-487a...	3RDPARTY	2
<input checked="" type="checkbox"/> CREDIT_SCORE	-	BANKING	Data Fabric Trial - Data...	39c6f420-703f-487a...	3RDPARTY	9
<input type="checkbox"/> CREDIT_SCORE_EXCE...	-	BANKING	Data Fabric Trial - Data...	39c6f420-703f-487a...	3RDPARTY	9
<input checked="" type="checkbox"/> MORTGAGE_APPLICANT	-	BANKING	Data Fabric Trial - Db2 ...	db2w-rugbyab.us-sout...	BLUDB	14

6. Click **View cart** to view your selection. From here, you can edit the table and schema names, or remove a selection from your cart.

Table	Business terms	Schema	Connection	Hostname: Port	Database	Columns
<input type="checkbox"/> MORTGAGE_APPLICAT...	-	BANKING	Data Fabric Trial - Db2 ...	db2w-rugbyab.us-sout...	BLUDB	10
<input type="checkbox"/> MORTGAGE_APPLICA...	-	BANKING	Data Fabric Trial - Db2 ...	db2w-rugbyab.us-sout...	BLUDB	14
<input checked="" type="checkbox"/> MORTGAGE_APPLICAT...	-	BANKING	Data Fabric Trial - Db2 ...	db2w-rugbyab.us-sout...	BLUDB	12
<input type="checkbox"/> MORTGAGE_CANDIDATE	-	BANKING	Data Fabric Trial - Data...	39c6f420-703f-487a...	3RDPARTY	2
<input type="checkbox"/> CREDIT_SCORE	-	BANKING	Data Fabric Trial - Data...	39c6f420-703f-487a...	3RDPARTY	9
<input type="checkbox"/> CREDIT_SCORE_EXCE...	-	BANKING	Data Fabric Trial - Data...	39c6f420-703f-487a...	3RDPARTY	9
<input checked="" type="checkbox"/> MORTGAGE_APPLICANT	-	BANKING	Data Fabric Trial - Db2 ...	db2w-rugbyab.us-sout...	BLUDB	14

7. Assign the virtualized tables to **Virtualized data**. Alternatively, you can assign the virtualized tables to **Project**, and select the **Multicloud data integration** project. This action will add the virtual tables to the assigned project and make them available on the *Virtualized data* page. Click **Virtualize**.

IBM Data Virtualization

Review cart and virtualize tables

Assign to: Project (radio button) Virtualized data (radio button)

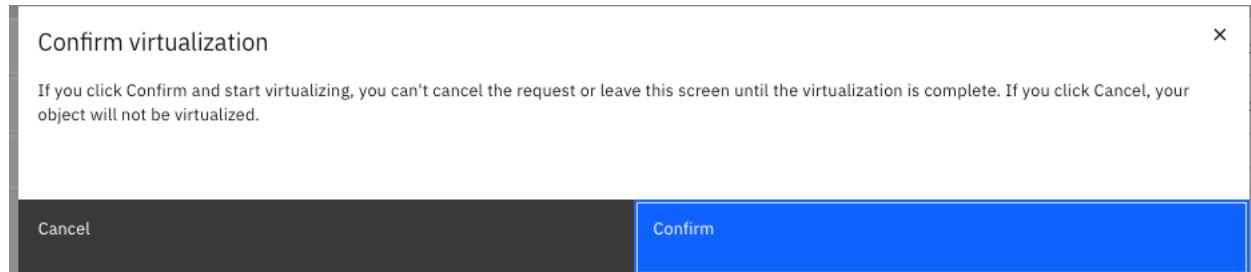
Publish to: Catalog (checkbox)

Select a catalog:

Objects to be virtualized:

Table	Schema	Source schema	Connections	Databases/File Path	Hostname: Port	Grouped tables	
MORTGAGE_APPLICATION	DV_IBMID_6650024CT3	X v	BANKING	Data Fabric Trial - Db2 ...	BLUDB	db2w-rugbyab.us-sout...	1
CREDIT_SCORE	DV_IBMID_6650024CT3	X v	BANKING	Data Fabric Trial - Data...	3RDPARTY	39c6f420-703f-487a-...	1
MORTGAGE_APPLICANT	DV_IBMID_6650024CT3	X v	BANKING	Data Fabric Trial - Db2 ...	BLUDB	db2w-rugbyab.us-sout...	1

## 8. Click **Confirm** to begin virtualizing the tables.



## 9. When virtualization is complete, click **Go to virtualized data** to see your newly created table.

Virtualize objects

Table	Schema	Virtualization status
MORTGAGE_APPLICATION	DV_IBMID_6650024CT3	<span style="color: green;">✓ Success</span>
CREDIT_SCORE	DV_IBMID_6650024CT3	<span style="color: green;">✓ Success</span>
MORTGAGE_APPLICANT	DV_IBMID_6650024CT3	<span style="color: green;">✓ Success</span>

Virtualize more data

Go to virtualized data

The following image shows the *Virtualized data* page. You are now ready to create a virtual table by joining these virtual tables.

Virtualized data

Table	Schema name	Created on
MORTGAGE_APPLICANT	DV_IBMID_663002GN1Q	Jun 22, 2022 3:26 PM
MORTGAGE_APPLICATION	DV_IBMID_663002GN1Q	Jun 22, 2022 3:26 PM
CREDIT_SCORE	DV_IBMID_663002GN1Q	Jun 22, 2022 3:26 PM

## Task 6: Create a virtual join view by joining virtual tables

You want to create a virtual join view by joining the MORTGAGE\_APPLICANT and MORTGAGE\_APPLICATION virtual tables. Then, you want to join the resulting virtual object with the CREDIT\_SCORE virtual table to create a second virtual join view.

### Virtual join view 1: Join MORTGAGE\_APPLICANT & MORTGAGE\_APPLICATION virtual tables

Follow these steps to create the first virtual join view:

1. From the *Virtualized data* page, select the **MORTGAGE\_APPLICATION** and **MORTGAGE\_APPLICANT** tables to virtualize.
2. Make note of the schema name. You will need that name later to run an SQL query.
3. Click **Join**.

Virtualized data

Table	Schema name	Created on
MORTGAGE_APPLICATION	DV_IBMID_6650024CT3	Aug 23, 2022 1:53 PM
MORTGAGE_APPLICANT	DV_IBMID_6650024CT3	Aug 23, 2022 1:53 PM
CREDIT_SCORE	DV_IBMID_6650024CT3	Aug 23, 2022 1:53 PM

4. In the list of columns for *MORTGAGE\_APPLICATION* table, drag to connect the **ID** column with the **ID** column in the *MORTGAGE\_APPLICANT* table.

### Join virtual objects

[Back](#)

Table 1: MORTGAGE\_APPLICANT

Find columns

Column name	Data type
ID	INTEGER
NAME	VARCHAR

Table 2: MORTGAGE\_APPLICATION

Find columns

Column name	Data type
ID	INTEGER
INCOME	INTEGER

## IBM Data Virtualization

5. Click **Preview** to see a preview of the joined tables.

Table: MORTGAGE\_APPLICATION | Schema: DV\_IBMID\_6650024CT3

<input checked="" type="checkbox"/> Column name	Data type
<input checked="" type="checkbox"/> ID	INTEGER
<input checked="" type="checkbox"/> INCOME	INTEGER
<input checked="" type="checkbox"/> APPLIEDONLINE	VARCHAR
<input checked="" type="checkbox"/> RESIDENCE	VARCHAR
<input checked="" type="checkbox"/> YRS_AT_CURRENT_ADDRESS	SMALLINT
<input checked="" type="checkbox"/> YRS_WITH_CURRENT_EMPLOYER	SMALLINT
<input checked="" type="checkbox"/> NUMBER_OF_CARDS	SMALLINT
<input checked="" type="checkbox"/> CREDITCARD_DEBT	SMALLINT
<input checked="" type="checkbox"/> LOANS	SMALLINT
<input checked="" type="checkbox"/> LOAN_AMOUNT	SMALLINT
<input checked="" type="checkbox"/> SALEPRICE	INTEGER
<input checked="" type="checkbox"/> LOCATION	VARCHAR

Table: MORTGAGE\_APPLICANT | Schema: DV\_IBMID\_6650024CT3

<input checked="" type="checkbox"/> Column name	Data type
<input checked="" type="checkbox"/> ID	INTEGER
<input checked="" type="checkbox"/> NAME	VARCHAR
<input checked="" type="checkbox"/> STREET_ADDRESS	VARCHAR
<input checked="" type="checkbox"/> CITY	VARCHAR
<input checked="" type="checkbox"/> STATE	VARCHAR
<input checked="" type="checkbox"/> STATE_CODE	VARCHAR
<input checked="" type="checkbox"/> ZIP_CODE	VARCHAR
<input checked="" type="checkbox"/> EMAIL_ADDRESS	VARCHAR
<input checked="" type="checkbox"/> PHONE_NUMBER	VARCHAR
<input checked="" type="checkbox"/> GENDER	VARCHAR
<input checked="" type="checkbox"/> SOCIAL_SECURITY_NUMBER	VARCHAR
<input checked="" type="checkbox"/> EDUCATION	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.

MORTGAGE_APPLICATION	MORTGAGE_APPLICANT
ID	ID

6. Close the preview window.

Table: MORTGAGE\_APPLICATION | Schema: DV\_IBMID\_6650024CT3

<input checked="" type="checkbox"/> Column name	New join preview									
<input checked="" type="checkbox"/> ID	ID	INCOME	APPLIEDONLINE	RESIDENCE	YRS_AT_CURRENT_ADDRESS	YRS_WITH_CURRENT_EMPL...	NUMBER_OF_CARDS	CREDITCARD_DEBT	LOANS	LOAN_AMOL
<input checked="" type="checkbox"/> INCOME	101593	50445	NO	Public Housing	10	8	1	237	1	9310
<input checked="" type="checkbox"/> APPLIEDONLINE	100534	44413	NO	Private Renting	13	16	2	885	0	4665
<input checked="" type="checkbox"/> RESIDENCE	100558	45419	NO	Private Renting	1	2	2	231	0	9200
<input checked="" type="checkbox"/> YRS_AT_CURRENT_ADDRESS	100745	45313	NO	Owner Occupier	13	3	2	341	1	4065
<input checked="" type="checkbox"/> YRS_WITH_CURRENT_EMPLOYER	101330	45217	NO	Private Renting	19	8	1	732	0	5860
<input checked="" type="checkbox"/> NUMBER_OF_CARDS	100752	44395	NO	Owner Occupier	20	16	2	1299	1	8750
<input checked="" type="checkbox"/> CREDITCARD_DEBT										
<input checked="" type="checkbox"/> LOANS		SMALLINT								
<input checked="" type="checkbox"/> LOAN_AMOUNT										
<input checked="" type="checkbox"/> SALEPRICE		INTEGER								
<input checked="" type="checkbox"/> LOCATION		VARCHAR								

Table: MORTGAGE\_APPLICANT | Schema: DV\_IBMID\_6650024CT3

<input checked="" type="checkbox"/> Column name	Data type
<input checked="" type="checkbox"/> PHONE_NUMBER	VARCHAR
<input checked="" type="checkbox"/> GENDER	VARCHAR
<input checked="" type="checkbox"/> SOCIAL_SECURITY_NUMBER	VARCHAR
<input checked="" type="checkbox"/> EDUCATION	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.

MORTGAGE_APPLICATION	MORTGAGE_APPLICANT
ID	ID

7. Click **Open in SQL editor**.

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.

MORTGAGE_APPLICATION	MORTGAGE_APPLICANT
ID	ID

8. Click **Yes, continue** at the notice that you are not able to return to the join canvas. The SQL Editor lets you run queries on the data set. In this case, you want to preview what records the data set will contain when you filter on California applicants.

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

- a) Delete the existing query.
- b) Copy and paste the following SELECT statement for the new query. Replace <your schema> with the schema name that you noted earlier.

```
SELECT * FROM <your-schema>.MORTGAGE_APPLICANT WHERE STATE_CODE LIKE 'CA'
```

- c) Your query looks similar to *below screenshot*. Click **Run all**.

IBM Cloud Pak for Data [Buy](#) [?](#) [Bell](#) [Help](#) [SV](#)

Search in your workspaces

Data virtualization / Virtualized data / Instance: Watson Query-eg

SQL editor [Back](#)

\* Untitled - 1 [X](#) [+](#)

[New](#) [Save](#) [Copy](#) [Paste](#) [Find](#) [Replace](#) [Format](#) [Delete](#) [Help](#) [Syntax assistant](#) [Run all](#)

```
1 SELECT * FROM DV_IBMID_6650024CT3.MORTGAGE_APPLICANT WHERE STATE_CODE LIKE 'CA'
```

- d) Click on the query in the history tab to view the results to see all of the applicants from the state of California.

IBM Data Virtualization

Script	Date	Status	Runtime
Untitled - 1	Aug 23, 2022 2:01:11 PM	✓ 1	0.542 s
SELECT * FROM DV_IBMID_6650024CT3.MORTGAGE_APPLICANT WHERE STATE_CODE LIKE 'CA'		✓	0.542 s

ID	NAME	STREET_ADDRESS	CITY	STATE	STATE_CODE	ZIP_CODE	EMAIL_ADDRESS
100679	Madelle Augie	1420 Beaumont Avenue	Beaumont	California	CA	92223	maugie7s@home.com

Items per page: 50 | 1–50 of 69 items | 1 of 2 pages | ▶ | ▷

e) Click **Back** to close the SQL editor.

9. Now that you previewed the data set filtered on California applicants, you will add this **filter criteria** to the virtual join view. For the *MORTGAGE\_APPLICANT* table, copy and paste the following statement for the filter criteria. Replace *<your schema>* with the schema name that you noted earlier.

```
"<your-schema>".MORTGAGE_APPLICANT"."STATE_CODE"='CA'
```

Your filter criteria looks similar to below screenshot.

Open in SQL editor ↗

**Join keys**      **Filters**

Programmatically filter data using simple clause statements like the one below.  
 "<SCHEMA>".<TABLE>".<COLUMN>" = <VALUE>

**MORTGAGE\_APPLICATION**

```
"DV_IBMID_6650024CT3"."MORTGAGE_APPLICATION"."ID" = <VALUE>
```

419 rows meet the filter criteria

**MORTGAGE\_APPLICANT**

```
"DV_IBMID_6650024CT3"."MORTGAGE_APPLICANT"."STATE_CODE"='CA'
```

69 rows meet the filter criteria

10.Click **Next**.

11.You can edit the column names to differentiate between columns with the same name in both tables. In this case, keep the default column names, and click **Next**.

## IBM Data Virtualization

The screenshot shows a table with 9 columns and 8 rows of data. The columns are labeled: ID, INCOME, APPLIEDONLINE, RESIDENCE, YRS\_AT\_CURRENT\_ADDRESS, YRS\_WITH\_CURRENT\_EM, NUMBER\_OF\_CARDS, CREDITCARD\_DEBT, and LOAN. The data includes various values such as 45153, YES, Owner Occupier, 2, 13, 2, 804, 1, etc.

ID	INCOME	APPLIEDONLINE	RESIDENCE	YRS_AT_CURRENT_ADDRESS	YRS_WITH_CURRENT_EM	NUMBER_OF_CARDS	CREDITCARD_DEBT	LOAN
101505	45153	YES	Owner Occupier	2	13	2	804	1
101698	58785	YES	Private Renting	31	16	2	1111	0
101549	44460	NO	Owner Occupier	4	16	1	3467	1
100536	45450	NO	Owner Occupier	15	24	1	4217	0
100620	44705	NO	Owner Occupier	0	8	1	230	0
100415	53791	NO	Owner Occupier	1	15	1	1844	1

12. On the *Assign and review* page, for the *View name*, type:

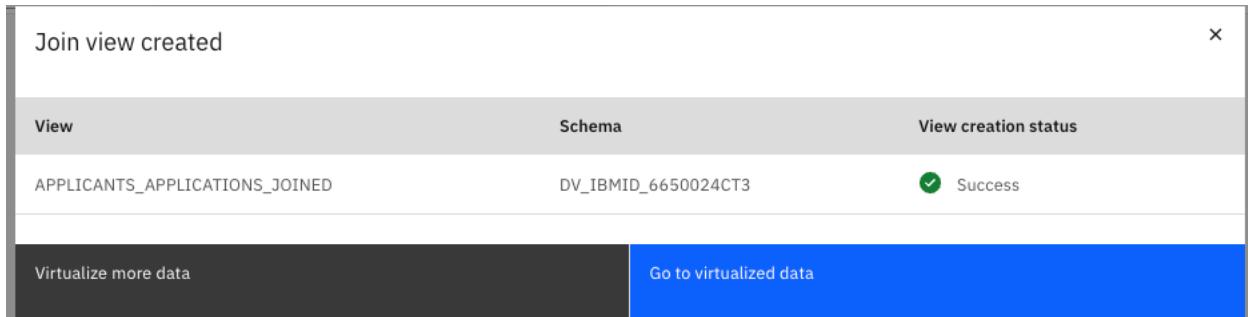
APPLICANTS\_APPLICATIONS\_JOINED

13. For now, assign the virtualized tables to **Virtualized data**. Later, you create a virtual object and assign that to the Multicloud data integration project.

14. Click **Create view**.

The screenshot shows the 'Assign and review' page. Under 'Assign to', the 'Virtualized data' radio button is selected. Under 'Publish to', the 'Catalog' checkbox is checked. In the 'View name' field, 'APPLICANTS\_APPLICATIONS\_JOINED' is typed. In the 'Schema name' field, 'DV\_IBMID\_6650024CT3' is typed. A success message 'Join view created' is displayed at the top.

15. When virtualization is complete, click **Go to virtualized data** to see your newly created join view.



The following image shows the *Virtualized data* page. You are now ready to create a second virtual join view.

## IBM Data Virtualization

The screenshot shows the IBM Data Virtualization interface. At the top, there's a navigation bar with 'IBM Cloud Pak for Data', a search bar, and links for 'Buy', 'Renovations', and a user profile. Below the navigation is a breadcrumb trail: 'Data Virtualization / Watson Query-qi'. The main area is titled 'Virtualized data' with a dropdown arrow. A search bar at the top says 'Find virtual objects'. Below it, a filter section says 'Filter by: All types'. A table lists four virtual objects:

<input type="checkbox"/> Table	Schema name	Created on
<input type="checkbox"/> APPLICANTS_APPLICATIONS_JOINED	DV_IBMID_663002GN1Q	Jun 22, 2022 3:44 PM
<input type="checkbox"/> MORTGAGE_APPLICANT	DV_IBMID_663002GN1Q	Jun 22, 2022 3:26 PM
<input type="checkbox"/> MORTGAGE_APPLICATION	DV_IBMID_663002GN1Q	Jun 22, 2022 3:26 PM
<input type="checkbox"/> CREDIT_SCORE	DV_IBMID_663002GN1Q	Jun 22, 2022 3:26 PM

A blue button in the top right corner of the table area says 'Add virtual objects +'.

### Virtual join view 2: Join APPLICANTS\_APPLICATIONS\_JOINED & CREDIT\_SCORE virtual tables

Follow these steps to create the second virtual join view:

1. From the *Virtualized data* page, select the **APPLICANTS\_APPLICATIONS\_JOINED** and **CREDIT\_SCORE** tables to virtualize.
2. Click **Join**.

The screenshot shows the 'Virtualized data' page with a list of selected virtual objects. A blue header bar indicates '2 items selected'. The table shows the following data:

<input checked="" type="checkbox"/> Table	Schema name	Created on
<input checked="" type="checkbox"/> APPLICANTS_APPLICATIONS_JOINED	DV_IBMID_6650024CT3	Aug 23, 2022 2:06 PM
<input type="checkbox"/> MORTGAGE_APPLICATION	DV_IBMID_6650024CT3	Aug 23, 2022 1:53 PM
<input type="checkbox"/> MORTGAGE_APPLICANT	DV_IBMID_6650024CT3	Aug 23, 2022 1:53 PM
<input checked="" type="checkbox"/> CREDIT_SCORE	DV_IBMID_6650024CT3	Aug 23, 2022 1:53 PM

At the top right of the table area, there are buttons for 'Manage access', 'Join', 'Assign', 'Publish to catalog', and 'Cancel'.

3. In the list of columns for *APPLICANTS\_APPLICATIONS\_JOINED* table, drag to connect the **EMAIL\_ADDRESS** column with the **EMAIL\_ADDRESS** column in the *CREDIT\_SCORE* table.
4. Click **Preview** to see a preview of the joined tables.

## IBM Data Virtualization

Table: APPLICANTS\_APPLICATIONS\_JOINED | Schema: DV\_IBMID\_6650024CT3

<input checked="" type="checkbox"/> Column name	Data type
STREET_ADDRESS	VARCHAR
CITY	VARCHAR
STATE	VARCHAR
STATE_CODE	VARCHAR
ZIP_CODE	VARCHAR
EMAIL_ADDRESS	VARCHAR
PHONE_NUMBER	VARCHAR
GENDER	VARCHAR
SOCIAL_SECURITY_NUMBER	VARCHAR
EDUCATION	VARCHAR
EMPLOYMENT_STATUS	VARCHAR
MARITAL_STATUS	VARCHAR

Table: CREDIT\_SCORE | Schema: DV\_IBMID\_6650024CT3

<input checked="" type="checkbox"/> Column name	Data type
ID	VARCHAR
NAME	VARCHAR
STREET_ADDRESS	VARCHAR
CITY	VARCHAR
STATE	VARCHAR
STATE_CODE	VARCHAR
ZIP_CODE	VARCHAR
EMAIL_ADDRESS	VARCHAR
CREDIT_SCORE	VARCHAR

### 5. Close the preview window and Click **Next**.

Table: APPLICANTS\_APPLICATIONS\_JOINED | Schema: DV\_IBMID\_6650024CT3

<input checked="" type="checkbox"/> Column name	Data type
STREET_ADDRESS	VARCHAR
CITY	VARCHAR
STATE	VARCHAR
STATE_CODE	VARCHAR
ZIP_CODE	VARCHAR
EMAIL_ADDRESS	VARCHAR
PHONE_NUMBER	VARCHAR
GENDER	VARCHAR
SOCIAL_SECURITY_NUMBER	VARCHAR
EDUCATION	VARCHAR
EMPLOYMENT_STATUS	VARCHAR
MARITAL_STATUS	VARCHAR

Table: CREDIT\_SCORE | Schema: DV\_IBMID\_6650024CT3

<input checked="" type="checkbox"/> Column name	Data type
ID	VARCHAR
INCOME	VARCHAR
APPLIEDONLINE	VARCHAR
RESIDENCE	VARCHAR
YRS_AT_CURRENT_ADDRESS	VARCHAR
YRS_WITH_CURRENT_EMPL...	VARCHAR
NUMBER_OF_CARDS	VARCHAR
CREDITCARD_DEBT	VARCHAR
LOANS	VARCHAR
LOAN_AMOU...	VARCHAR
EMAIL_ADDRESS	VARCHAR

### 6. Accept the default column names, and click **Next**.

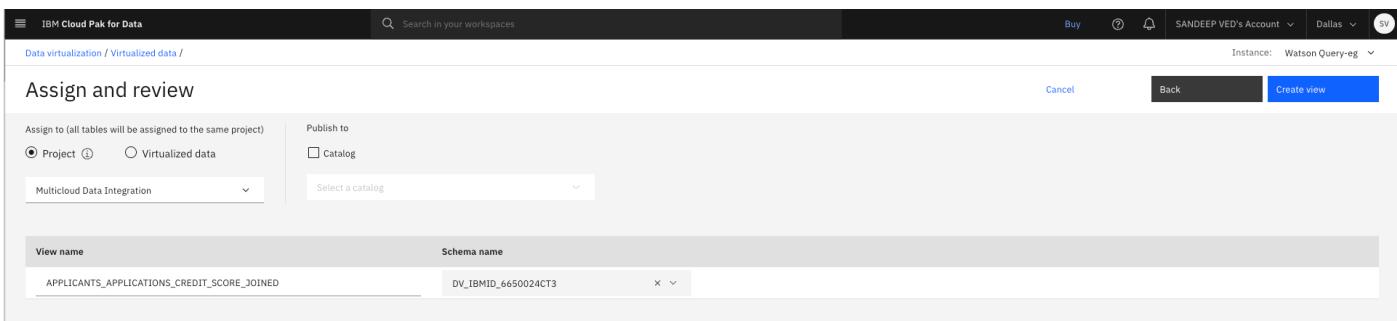
Table: APPLICANTS\_APPLICATIONS\_CREDIT\_SCORE\_JOINED

ID	INCOME	APPLIEDONLINE	RESIDENCE	YRS_AT_CURRENT_ADDRESS	YRS_WITH_CURRENT_EMPL...	NUMBER_OF_CARDS	CREDITCARD_DEBT	LOAN_AMOU...
100676	44597	NO	Owner Occupier	10	12	2	468	0
100377	51945	YES	Public Housing	19	0	1	25	1
101626	52887	NO	Public Housing	10	9	2	1339	0
100521	44394	YES	Owner Occupier	20	17	1	2041	0
100622	49600	NO	Private Renting	5	9	2	3366	1
100794	53704	NO	Owner Occupier	21	17	2	490	1

### 7. On the *Assign and review* page, for the *View name*, type APPLICANTS\_APPLICATIONS\_CREDIT\_SCORE\_JOINED

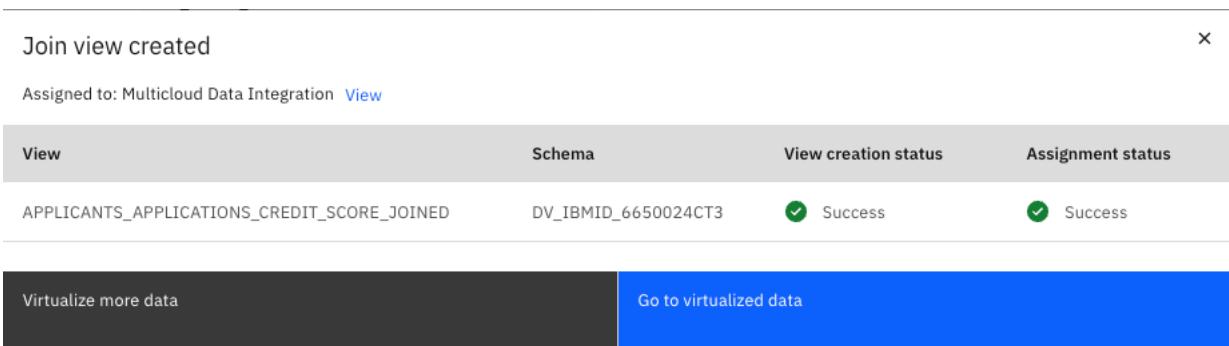
## IBM Data Virtualization

8. Assign the virtualized view to **Project**, and select the **Multicloud data integration** project.
9. Click **Create view**.



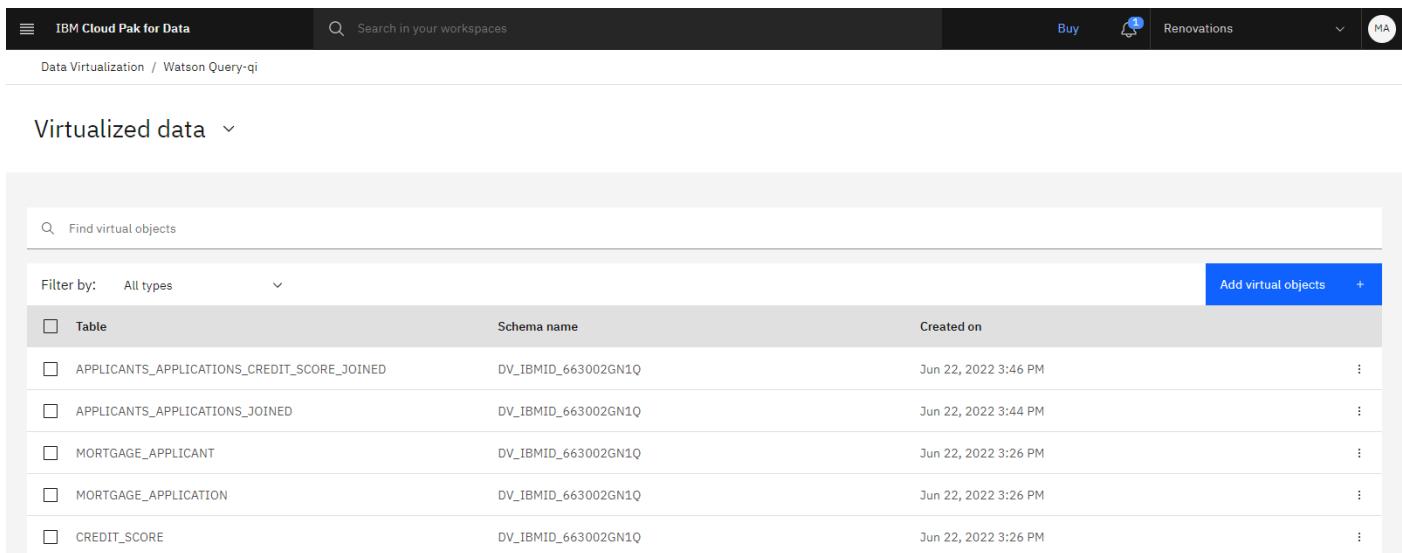
The screenshot shows the 'Assign and review' step in the IBM Cloud Pak for Data interface. It displays the 'View name' as 'APPLICANTS\_APPLICATIONS\_CREDIT\_SCORE\_JOINED' and the 'Schema name' as 'DV\_IBMID\_6650024CT3'. The 'Assignment status' is marked as 'Success'.

10. When virtualization is complete, click **Go to virtualized data** to see your newly created join view.



The screenshot shows a confirmation dialog titled 'Join view created'. It displays the view 'APPLICANTS\_APPLICATIONS\_CREDIT\_SCORE\_JOINED' assigned to schema 'DV\_IBMID\_6650024CT3' with a success status. Below the dialog are buttons for 'Virtualize more data' and 'Go to virtualized data'.

The following image shows the *Virtualized data* page. You are now ready to work with the virtual data in your analytics project.



The screenshot shows the 'Virtualized data' page in the IBM Cloud Pak for Data interface. It lists several virtual objects, including tables like 'APPLICANTS\_APPLICATIONS\_CREDIT\_SCORE\_JOINED' and 'MORTGAGE\_APPLICANT', along with their schema names and creation dates.

Table	Schema name	Created on
APPLICANTS_APPLICATIONS_CREDIT_SCORE_JOINED	DV_IBMID_663002GN1Q	Jun 22, 2022 3:46 PM
APPLICANTS_APPLICATIONS_JOINED	DV_IBMID_663002GN1Q	Jun 22, 2022 3:44 PM
MORTGAGE_APPLICANT	DV_IBMID_663002GN1Q	Jun 22, 2022 3:26 PM
MORTGAGE_APPLICATION	DV_IBMID_663002GN1Q	Jun 22, 2022 3:26 PM
CREDIT_SCORE	DV_IBMID_663002GN1Q	Jun 22, 2022 3:26 PM

## Task 7: Access the virtual join view in the project

The virtual table was added to your project along with a connection to Watson Query. Follow these steps to open the project to see the virtual data and the connection information that is required to access the virtual data.

1. Switch back to Cloud Pak for Data. From the navigation menu ☰, choose **Projects > View all projects**.
2. Open the **Multicloud data integration** project.

Name	Date created	Your role	Collaborators
Data Virtualisation Project	4 weeks ago	Admin	<span>(1)</span>
Multicloud Data Integration	4 weeks ago	Admin	<span>(1)</span>

3. Click the **Assets** tab.

4. Click on the new DV Connection Object. Click on Credentials Navigation Menu in the left.

5. Choose the Authentication Method as API Keys and enter the IBM API Keys. Click **Save**.

## IBM Data Virtualization

**Edit connection: IBM Data Virtualization**

Review the connection information.

**Credentials**

Authentication method\* ⓘ

API key

API key\* ⓘ

.....

Port is SSL-enabled ⓘ

SSL certificate ⓘ

**Private connectivity**

Satellite Link ⓘ

Secure Gateway ⓘ

Cancel Save

## 6. You will land back to Asset Tab. Click on Data Sub Menu.

IBM Cloud Pak for Data

Projects / Multicloud Data Integration

Overview Assets Jobs Manage

Find assets

Import asset ⓘ New asset +

13 assets

All assets

Asset types

- Data access 4
- Connections 4
- Data** 6
- Flows 2
- Source Code 1

Name	Last modified
DV_IBMID_6650024CT3.APPLICANTS_APPLICATIONS_CREDIT_SCORE_JOINED	9 minutes ago SANDEEP VED (You)
Customers.csv	4 weeks ago SANDEEP VED (You)
Sales representatives.csv	4 weeks ago SANDEEP VED (You)
MORTGAGE_APPLICANTS_INTEREST_RATES_CSV_shaped	4 weeks ago SANDEEP VED (You)
MORTGAGE_APPLICANTS_INTEREST_RATES.CSV	4 weeks ago Service
MORTGAGE_INTEREST_RATES.CSV	4 weeks ago Service

Data in this project

Drop data files here or browse for files to upload

## 7. Open any of the virtualized data. For example, click the **APPLICANTS\_APPLICATIONS\_CREDIT\_SCORE\_JOINED** data asset to view it.

IBM Cloud Pak for Data

Projects / Multicloud Data Integration / DV\_IBMID\_6650024CT3.APPLIC...

Preview asset Profile Visualization

Schema: 33 Columns

The preview includes only a limited set of columns and rows. ⓘ

DV_IBMID_6650024CT3_APPLICANTS_APPLICATIONS_JOINED...	INCOME	APPLIEDON...	RESIDENT...	YRS_AT_CURRENT_ADD...	YRS_WITH_CURRENT_EMPL...	NUMBER_OF_CA...	CREDITCARD...	LOANS
100415	53791	NO	Owner Occupier	1	15	1	1844	1
100507	43633	NO	Public Housing	14	16	1	6100	0
100789	43151	YES	Private Renting	6	9	1	681	1
101716	45217	YES	Owner Occupier	17	8	2	445	1
100389	59036	NO	Owner Occupier	7	6	1	495	1
101672	54777	NO	Public Housing	26	17	2	2814	1
100473	43750	YES	Private Renting	13	6	1	988	1
100381	53115	NO	Owner Occupier	12	1	2	674	0
101789	43151	YES	Private Renting	6	9	1	681	1
101458	46645	YES	Owner Occupier	19	4	1	884	0
101549	44460	NO	Owner Occupier	4	16	1	3467	1
101468	45970	YES	Owner Occupier	11	7	1	579	0
100792	44230	NO	Owner Occupier	9	22	2	3106	0
101551	43707	NO	Owner Occupier	8	9	1	1131	0
100310	43972	NO	Private Renting	0	13	1	340	1
100529	50075	NO	Private Renting	4	18	2	3768	0
100442	44271	NO	Private Renting	17	20	1	284	1
100705	49485	NO	Private Renting	24	10	1	563	1
100670	56470	NO	Private Renting	31	15	2	390	0
100679	45584	YES	Private Renting	2	19	1	1256	0
101626	52887	NO	Public Housing	10	9	2	1339	0
100521	44394	YES	Owner Occupier	20	17	1	2041	0

Information

Data asset

DV\_IBMID\_6650024CT3.APPLICANTS\_APPLICATIONS\_CREDIT\_SCORE\_JOINED

Description

No description available for this asset

Tags

connected-data

Creator

SANDEEP VED

Usage

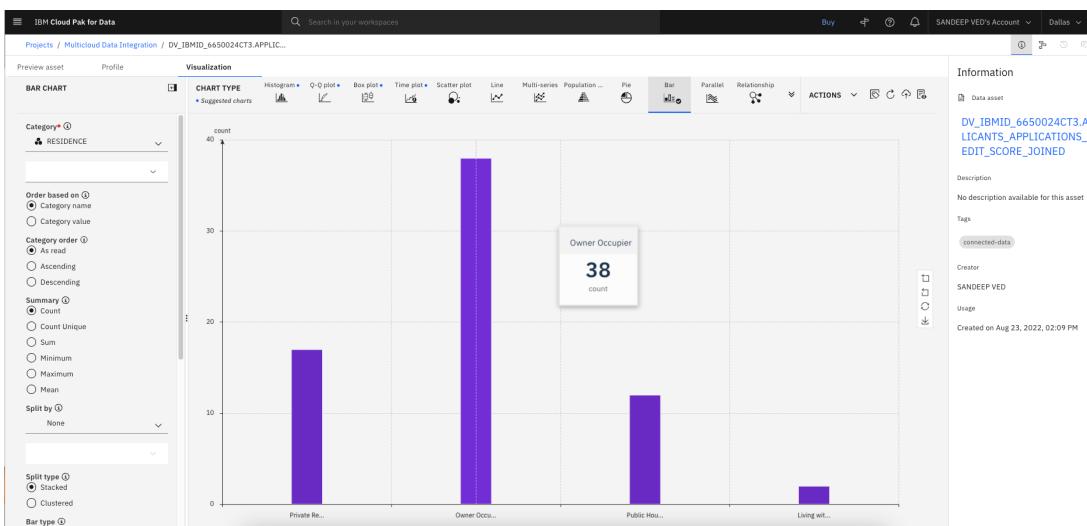
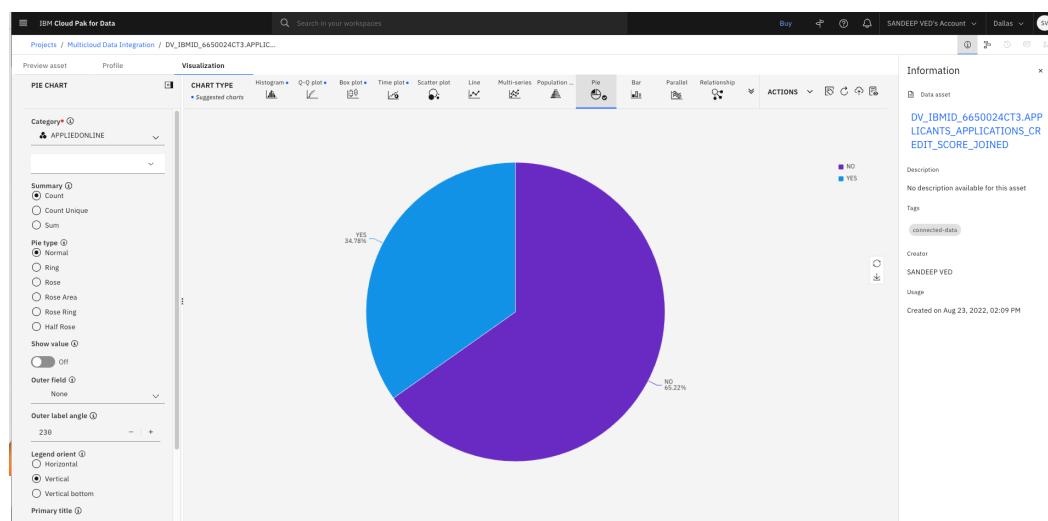
Created on Aug 23, 2022, 02:09 PM

## IBM Data Virtualization

8. Scroll through the data asset to see all of the applicants from the state of California.
9. Click on Visualization Tab.

The screenshot shows the 'Visualization' tab selected in the top navigation bar. A central message says: "Choose a chart above or select columns below, and then choose a chart. If you select columns, suggested charts will be indicated with a dot next to the chart name." Below this is a section titled "Columns to visualize" with a "Add another column" button. At the bottom left is a "SELECTED COLUMNS" panel showing "0" items. On the right is an "Information" panel for the data asset "DV\_IBMID\_6650024CT3.APPLICANTS\_APPLICATIONS\_CREDIT\_SCORE\_JOINED". It includes fields for Description (empty), Tags (connected-data), Creator (SANDEEP VED), and Usage (Created on Aug 23, 2022, 02:09 PM).

10. Select a column and type of graph to visualize the data.



## Task 8: Use Cognos Dashboard (Optional)

If you want to use Cognos Dashboards in the Project, you will need to associate the IBM Cognos Dashboard Embedded Service with the project as below.

- 1) Under the Project, Navigate to **Manage** Tab. Click on **Services & Integrations**. Click on **Associate Service** Button.

The screenshot shows the IBM Cloud Pak for Data interface. The top navigation bar includes 'Buy', 'SANDEEP VED's Account', 'Dallas', and a user icon. Below the navigation is a search bar 'Search in your workspaces'. The main menu has tabs for 'Overview', 'Assets', 'Jobs', and 'Manage', with 'Manage' being the active tab. On the left, a sidebar lists 'Project', 'General', 'Access control', 'Environments', 'Resource usage', and 'Services & integrations', with 'Services & integrations' being the active item. The central content area is titled 'Services & integrations' and 'IBM services'. It contains a sub-section 'Associate IBM Cloud services with this project to add tools, compute environments, or other capabilities. Learn more.' Below this is a search bar 'Find services' and a table with columns 'Name' and 'Service type'. A large blue button 'Associate service +' is located on the right side of the screen.

- 2) Select **IBM Cognos Dashboard Embedded-am** Service and click **Associate**.

The screenshot shows the 'Associate service' dialog box. At the top, it says 'Associate service' and 'Choose an existing or add a new service to associate with your project.' Below this is a filter bar with 'Filter by: Resource Groups', 'Locations' (set to 'None'), and a search input 'co'. The main table lists services with columns 'Name', 'Type', 'Plan', 'Location', 'Status', and 'Group'. One row is selected: 'IBM Cognos Dashboard Embedded-am' (Type: IBM Cognos Dashboard Embedded, Plan: Lite, Location: Dallas, Status: Not associated, Group: Default). At the bottom of the dialog are two buttons: 'Cancel' and a blue 'Associate' button.

- 3) The selected service will be associated with the project.

## IBM Data Virtualization

IBM Cloud Pak for Data

Projects / Multicloud Data Integration

Overview Assets Jobs Manage

Services & integrations

IBM services (1) Third-party integrations

Associate IBM Cloud services with this project to add tools, compute environments, or other capabilities. [Learn more.](#)

Find services

Name Service type

IBM Cognos Dashboard Embedded-am Cognos Dashboard Embedded

Associate service +

4) You should be able to add a **Dashboard Editor** as a **New Asset** to the project now.

IBM Cloud Pak for Data

Projects / Multicloud Data Integration

Overview Assets Jobs Manage

Find assets

13 assets

All assets

Asset types

- Data access (4)
- Data (6)** (selected)
- Flows (2)
- Source Code (1)

Data

Name	Last modified
DV_IBMD_6650024CT3.APPLICANTS_APPLICATIONS_CREDIT_SCORE_JOINED	15 minutes ago SANDEEP VED (You)
Customers.csv	4 weeks ago SANDEEP VED (You)
Sales representatives.csv	4 weeks ago SANDEEP VED (You)
MORTGAGE_APPLICANTS_INTEREST_RATES_CSV_shaped	4 weeks ago SANDEEP VED (You)
MORTGAGE_APPLICANTS_INTEREST_RATES.CSV	4 weeks ago Service
MORTGAGE_INTEREST_RATES.CSV	4 weeks ago Service

Import asset New asset +

Data in this project

Drop data files here or browse for files to upload

IBM Cloud Pak for Data

Projects / Multicloud Data Integration

New asset

Select a tool based on what type of asset you want and how you want to work.

Tool type

- All types** (selected)
- Data access tools
- Automated builders
- Graphical builders
- Code editors
- Component editors

Connected data moved To add connected data, close this dialog and click Import asset.

Find tools by name or description

<b>Dashboard editor</b> Create a set of visualizations of analytical results on a graphic canvas without coding. <a href="#">Learn more</a>	<b>Data Refinery</b> Create a flow of ordered operations to cleanse and shape data. Visualize data to identify problems and discover insights. <a href="#">Learn more</a>	<b>DataStage</b> Create a flow with a set of connectors and stages to transform and integrate data. Provide enriched and tailored information for your enterprise. <a href="#">Learn more</a>	<b>Decision Optimization</b> Create and manage scenarios to find the best solution to your optimization problem by comparing different combinations of your model, data, and solutions. <a href="#">Learn more</a>
<b>Masking flow</b> Create and run masking flows to prepare copies of data assets that are masked by advanced data protection rules. <a href="#">Learn more</a>	<b>Pipelines</b> Automate the model lifecycle, including preparing data, training models, and creating deployments. <a href="#">Learn more</a>	<b>SPSS Modeler</b> 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. <a href="#">Learn more</a>	
<b>Code editors</b>			
<b>Federated Learning</b>	<b>Jupyter notebook editor</b>		

Show descriptions

## IBM Data Virtualization

Create asset  
Create a dashboard

New Local file

Define details

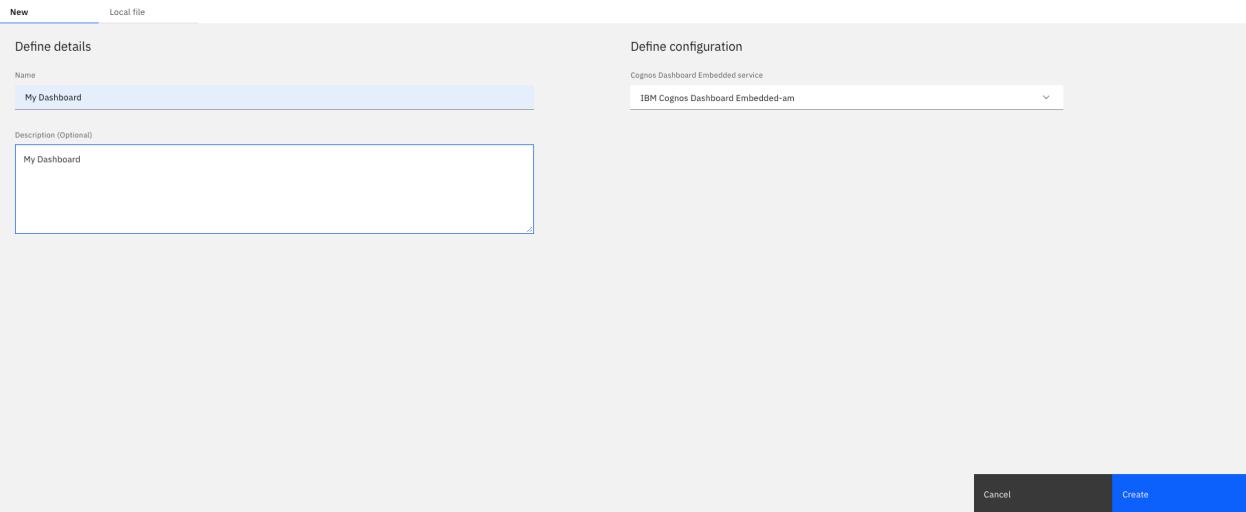
Name: My Dashboard

Description (Optional): My Dashboard

Define configuration

Cognos Dashboard Embedded service: IBM Cognos Dashboard Embedded-am

Cancel Create



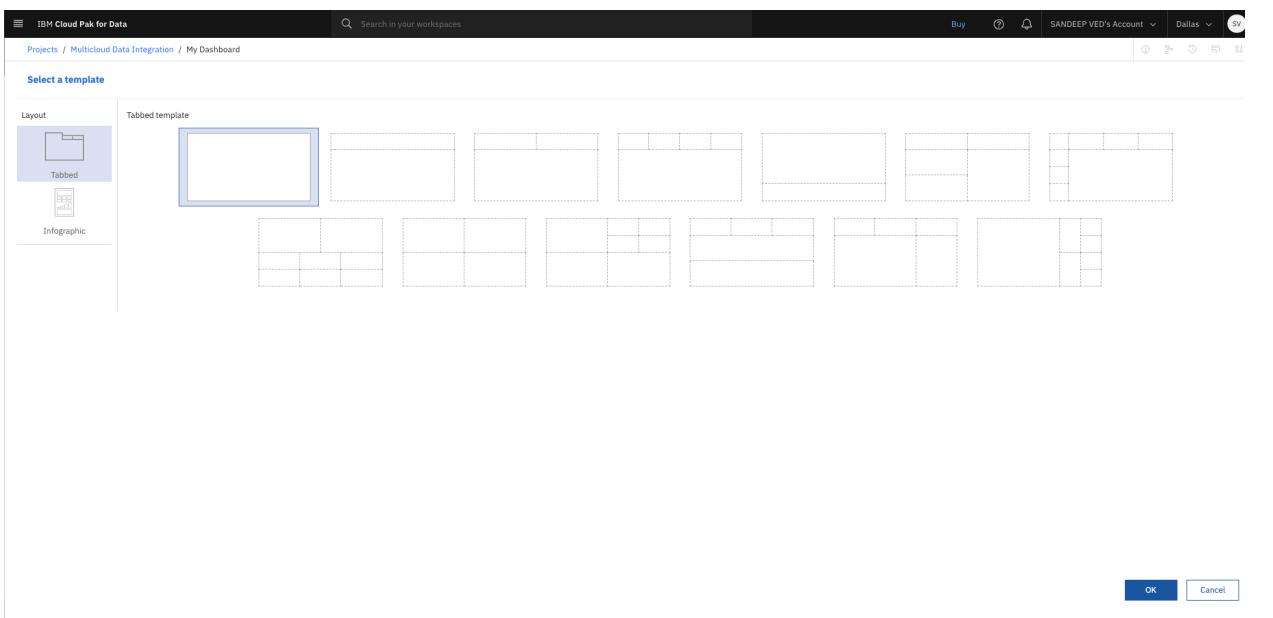
IBM Cloud Pak for Data Search in your workspaces SANDEEP VED's Account Dallas SV

Projects / Multicloud Data Integration / My Dashboard

Select a template

Layout: Tabbed

Tabbed template:



OK Cancel



IBM Cloud Pak for Data Search in your workspaces SANDEEP VED's Account Dallas SV

Projects / Multicloud Data Integration / My Dashboard

All tabs This tab

Drag and drop data here to filter all tabs.

Drag and drop data here to filter this tab.

Selected sources

Tab 1

...

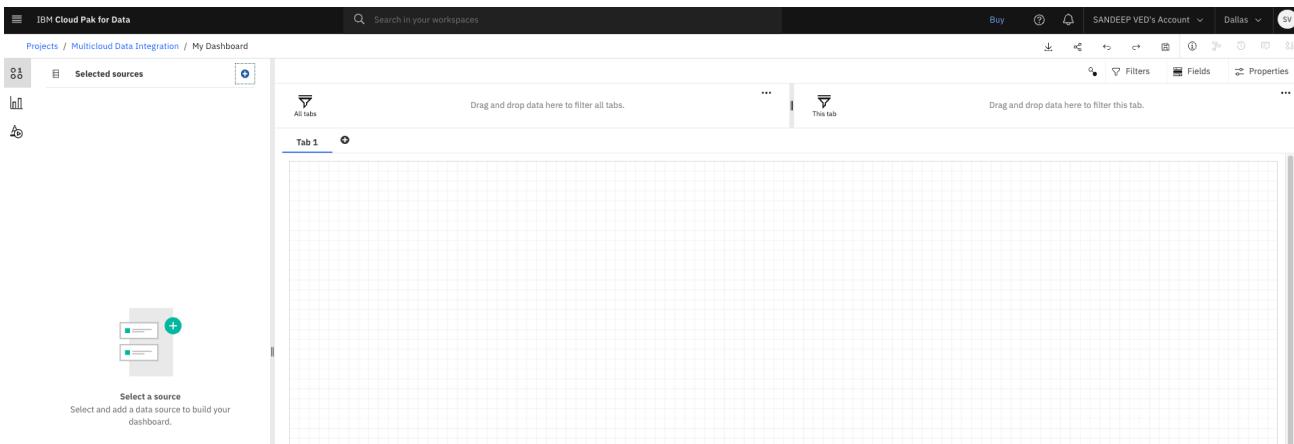
Properties

Filters Fields

OK Cancel

Select a source

Select and add a data source to build your dashboard.



## Task 9: Use Jupyter Notebook Dashboard (Optional)

You can add a **Jupyter Notebook** in the project as a **New Asset** for your Data Science Tasks.

## IBM Data Virtualization

The screenshot shows the 'Assets' tab in the IBM Cloud Pak for Data interface. On the left, there's a sidebar with '15 assets' and a tree view of asset types: Data access (Connections: 4), Data (6), Flows (2), Visualizations (1), and Source Code (2). The main area displays a table of 'Connections' with columns for Name, Created by, and Last modified. There are four entries: DS1612349864180111 (Connection), Data Fabric Trial - Databases for PostgreSQL (Connection), Data Fabric Trial - Db2 Warehouse (Connection), and Data Fabric Trial - MongoDB (Connection). A sidebar on the right says 'Data in this project' with a note 'Drop data files here or browse for files to upload'.

### Find and Select Jupyter notebook editor.

#### New asset

Select a tool based on what type of asset you want and how you want to work.

The screenshot shows a 'New asset' dialog. In the search bar, 'jup' is typed. The results show a single item: 'Jupyter notebook editor'. A tooltip above it says 'Connected data moved To add connected data, close this dialog and click Import asset.' Below the result, a description says 'Create a notebook in which you run Python, R, or Scala code to prepare, visualize, and analyze data, or build a model.'

### Enter the Notebook Name, Description and select the Runtime and click **Create**.

The screenshot shows the 'New notebook' dialog. It has tabs for 'Blank', 'From file', and 'From URL'. The 'Name' field contains 'My Notebook'. The 'Description (optional)' field also contains 'My Notebook'. Under 'Select runtime', 'Runtime 22.1 on Python 3.9 XS (2 vCPU 8 GB RAM)' is selected. A note says 'The selected runtime has 2 vCPU and 8 GB RAM. It consumes 1 capacity unit per hour.' Below that, 'Language' is set to 'Python 3.9'. At the bottom, there are 'Cancel' and 'Create' buttons.

# Data Virtualization Lab – Using New Project & Connections

Use Watson Query to combine data from different data sources and with different types in a **new project and defining New Connections**. Use SQL syntax and access & combined data without data movement.

Your basic workflow includes these tasks:

1. Provision the required additional service (DB2).
2. Create databases in multiple data sources and collect database details and credentials.
3. Create a New Project.
4. Add connections to your data sources.
5. Create virtual objects by combining data from all your data sources.
6. Add virtualized data to your catalogues and projects.
7. Manage access to your virtual objects (optional).

## Task 1: Provision the required services

This lab requires the Watson Query service as well as a Db2 service.

1. From the navigation menu, click **Services > Service instances**.

Name	Group	Location	Product	Plan	Status
CloudObjectStorage	Default	Global	Cloud Object Storage	Lite	Active
DataStage-ia	Default	Dallas	DataStage	Lite	Active
Db2-tq	Default	Dallas	Db2	Lite	Active
WatsonMachineLearning	Default	Dallas	Machine Learning	Lite	Active
KnowledgeCatalog	Default	Dallas	Watson Knowledge Catalog	Lite	Active
WatsonOpenScale	Default	Dallas	Watson OpenScale	Lite	Active
Watson Query-eg	Default	Dallas	Watson Query	Lite	Active
WatsonStudio	Default	Dallas	Watson Studio	Lite	Active

2. If you have a *Watson Query* service listed, then there is no need to provision another instance. Otherwise, follow the steps in Lab Prerequisites Section to [provision Watson Studio Service](#).
3. If you have a *Db2 on Cloud* or *Db2 Warehouse on Cloud* service listed, then there is no need to provision another instance. Otherwise, follow these steps:
  1. Click **Add service**.
  2. Select **Db2**.

## IBM Data Virtualization

The screenshot shows the IBM Cloud Pak for Data Services catalog interface. A search bar at the top contains the query 'db2'. Below the search bar, a sidebar lists categories: AI / Machine Learning, Databases, Analytics, Integration, and Storage. The main area displays two service cards: 'Db2 Warehouse' under 'Databases' and 'Db2' under 'Databases'. Both cards include a brief description and a 'Lite • Free' status indicator.

**3. Select the **Lite** plan for Db2 on Cloud. Only Available in **Dallas** Region. Click **Create**.**

The screenshot shows the 'Db2' service creation page. The 'Create' tab is selected. It displays two plan options: 'Lite' (selected) and 'Standard'. The 'Lite' plan includes 200 MB of data storage, 5 simultaneous connections, and is described as a free service. The 'Standard' plan includes flexible scaling of compute and storage. On the right side, a summary panel shows the service details: Region: Dallas, Plan: Lite, Service name: Db2-ac, and Resource group: Default. A large blue 'Create' button is at the bottom.

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

**5. Access the *Db2 on Cloud* (or *Db2 Warehouse on Cloud*) instance menu, and select **Manage in IBM Cloud**. This will launch the service instance page in IBM Cloud.**

The screenshot shows the 'Service instances' page. At the top, there's a message about upgrading the service plan. Below it, a search bar and filter options (Resource Groups, Locations, Product, Service plan) are available. The main table lists various service instances, including 'CloudObjectStorage', 'DataStage-ia', 'Db2-tq' (highlighted with a red box), 'WatsonMachineLearning', 'KnowledgeCatalog', 'WatsonOpenScale', 'Watson Query-eg', and 'WatsonStudio'. For each instance, columns show Name, Group, Location, Product, Plan, Status, and actions (⋮, Upgrade service, Manage, Manage in IBM Cl..., Delete). The 'Db2-tq' row has its 'Manage in IBM Cl...' button highlighted with a red box.

## 6. Click **Service credentials**.

Resource list / Db2-tq Active cpdaas Actions...

Manage Getting started **Service credentials** Connections

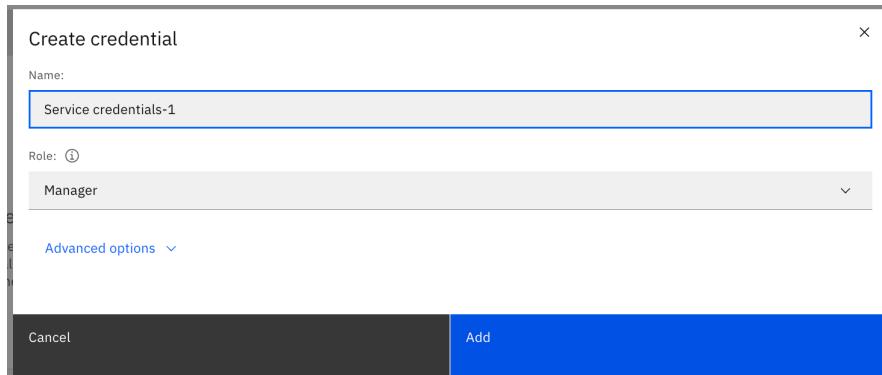
**Service credentials**  
You can generate a new set of credentials for cases where you want to manually connect an app or external consumer to an IBM Cloud service. [Learn more](#)

Search credentials... New credential +

Key name	Date created
No service credentials Credentials are provided in JSON format. The JSON snippet lists credentials, such as the API key and secret, as well as connection information for the service.	

## 7. If there are no credentials listed, follow these steps:

1. Click **New credential**.
2. Click **Add**.



## 8. Expand the service credentials, and copy them to the clipboard. You will need these credentials later to configure a connection, so you may want to save the credentials in a text file.

Resource list / Db2-tq Active cpdaas Actions...

Manage Getting started **Service credentials** Connections

**Service credentials**  
You can generate a new set of credentials for cases where you want to manually connect an app or external consumer to an IBM Cloud service. [Learn more](#)

Search credentials... New credential +

Key name	Date created
Service credentials-1	2022-07-28 5:57 PM

IBM Data Virtualization

For more information, see [Db2 on Cloud Pak for Data as a Service](#) and [Watson Query on Cloud Pak for Data as a Service](#).

## Task 2: Load data into your Db2 data source

The data sets you will use in this Lab are available in the **GitHub** and **IBM Gallery** (link below). You need to load those data sets into Db2 on Cloud.

1. Access these two data sets in the Project. Click **Add to Project** and select a project.
    - [Customers data set](#)
    - [Sales data set](#)

IBM Cloud Pak for Data

Search in your workspaces

Buy SANDEEP VED's Account Dallas SV

Gallery / Customers /

Back

## Customers

**Description**

This data set includes a list of fictitious general customers data along with the contact and payment information and their annual sales. Columns include:

- CUST\_ID
- CUSTNAME
- CITY
- STATE
- COUNTRY
- POSTAL\_CODE
- EMAIL\_ADDRESS
- PHONE\_NUMBER
- YTD\_SALES
- SALESREP\_ID
- NATIONALITY
- NATIONAL\_ID
- CREDITCARD\_NUMBER
- CREDITCARD\_TYPE
- CREDITCARD\_EXP
- CREDITCARD\_CVV

**Tags**

Economy & Business

**Modified**

Jul 24, 2021

Add to project

2. Download the Asset to your local computer to be added into DB2 DB later.

The screenshot shows the 'IBM Data Virtualization' service instance in the IBM Cloud Pak for Data. The 'Assets' tab is active. On the left, there's a sidebar with '1 assets' and 'All assets'. The main area shows a table with 'Name' and 'Last modified' columns. A file named 'Customers.csv' is listed, showing it was modified 'Now' by 'SANDEEP VED (You)'. To the right of the table is a context menu with options like 'Publish to catalog' (which is highlighted with a blue border), 'Promote to space', 'Refine', 'Download' (which is also highlighted with a blue border), and 'Delete'.

- From your *Db2 on Cloud* (or *Db2 Warehouse on Cloud*) service instance page in IBM Cloud, select the **Manage** panel.

The screenshot shows the 'Service instances' page in the IBM Cloud Pak for Data. It lists various service instances such as CloudObjectStorage, DataStage-ia, Db2-tq, WatsonMachineLearning, KnowledgeCatalog, WatsonOpenScale, WatsonQuery-eg, and WatsonStudio. A context menu is open for the 'WatsonMachineLearning' instance, with the 'Upgrade service' option highlighted with a blue border. Other options in the menu include 'Manage' (highlighted with a red border), 'Manage in IBM Cl...', and 'Delete'.

- Click **Go to UI**. The *Db2 on Cloud* console displays.

The screenshot shows the 'Db2-tq' service instance page in the Db2 on Cloud console. It has sections for 'Getting started' and 'Need help?'. In the 'Getting started' section, the 'Go to UI' button is highlighted with a red border. In the 'Need help?' section, the 'Support case' button is highlighted with a blue border.

- Select the **Data** panel. Drag the **customers.csv** file from your local machine into the *Load Data* panel in the *Db2 on Cloud* console, and click **Next**.

## IBM Data Virtualization

IBM Db2 on Cloud

Load Data Load History Tables Views Indexes Aliases MQTs Sequences Application objects

Source Target Define Finalize

You are loading the file

File selection

My Computer A single delimited text file (CSV) without header row.

S3 Amazon S3

Cloud Object Storage

Drag a file here or browse files

Next

This screenshot shows the 'File selection' step in the IBM Data Virtualization interface. On the left, there's a sidebar with icons for Home, Load Data, Load History, Tables, Views, Indexes, Aliases, MQTs, Sequences, Application objects, and Help. The main area has tabs for Source, Target, Define, and Finalize. Under Source, it says 'You are loading the file'. Below that is a 'File selection' dialog with a 'Drag a file here or browse files' area. To the right, a 'Selected file' section shows 'Customers.csv'. At the bottom right is a 'Next' button.

IBM Db2 on Cloud

Load Data Load History Tables Views Indexes Aliases MQTs Sequences Application objects

Source Target Define Finalize

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

File selection

My Computer A single delimited text file (CSV) without header row.

S3 Amazon S3

Cloud Object Storage

Drag a file here or browse files

Selected file

Customers.csv

Next

This screenshot shows the same interface as the previous one, but with a file selected. The 'Selected file' section now shows 'Customers.csv'. The 'Next' button at the bottom right is highlighted in blue.

6. Select a schema. Click **New table +**, type `customers` for the table name, and click **Create**.

## IBM Data Virtualization

IBM Db2 on Cloud

Load Data Load History Tables Views Indexes Aliases MQTs Sequences Application objects

Select a load target

Schema

Table

No entries found.

Back Next

IBM Db2 on Cloud

Load Data Load History Tables Views Indexes Aliases MQTs Sequences Application objects

Select a load target

Schema

Table

New table +

Create a new table

Customers

Create

No entries found.

Back Next

7. Click **Next** to continue.

IBM Db2 on Cloud

Load Data Load History Tables Views Indexes Aliases MQTs Sequences Application objects

Select a load target

Schema

Table

CUSTOMERS

Back Next

## IBM Data Virtualization

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

	CUST_ID SMALLINT	CUSTNAME VARCHAR(20)	CITY VARCHAR(19)	STATE VARCHAR(3)	COUNTRY_CODE VARCHAR(2)	POSTAL_CODE VARCHAR(8)	EMAIL_ADDRESS VARCHAR(37)
1	10001	Michael Golden	Abbadia Alpina	TO	IT	10060	Michael.E.golden@spambob.com
2	10002	Renee Mullins	Columbus	OH	US	45101	Rene.K.mullins@dodgeit.com
3	10003	Allen Schmidt	Abeto	PG	IT	6040	Allen.M.schmidt@spambob.com
4	10004	Robert May	Houston	TX	US	79601	Robert.C.may@spambob.com
5	10005	Joe Cruz	Harrisonburg	VA	US	24210	Joe.D.cruz@spambob.com
6	10006	Rebecca White	Acate	RG	IT	97011	Rebecca.white@spambob.com
7	10352	Gary Neal	Chicago	IL	US	61701	Gary.N.neal@dodgeit.com
8	10422	Steve Huynh	Detroit	MI	US	48614	Steve.A.huynh@dodgeit.com
9	10007	Anthony Johnson	Achstetten		DE	88480	Anthony.H.johnson@pookmail.com
10	10008	Alberto Fabian	Acquabona	BL	IT	32043	Alberto.T.fabian@pookmail.com

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

Summary		Option
Code page:	1208 (Default)	Maximum number of warnings
Separator:	,	1000
Time format:	HH:MM:SS (Default)	
Date format:	YYYY-MM-DD (Default)	
Timestamp format:	YYYY-MM-DD HH:MM:SS (Default)	
String delimiter:	(Default)	

10. When the load is successful, click **Load More Data**.

## IBM Data Virtualization

The screenshot shows the 'Load Data' interface in the IBM Data Virtualization tool. At the top, there's a navigation bar with links like 'Load Data', 'Load History', 'Tables', 'Views', etc. On the left, there are icons for 'SQL', 'PQL', and 'Tables'. The main area is titled 'Load details' and shows a summary: 'My computer' (Source) has loaded 'Customers.csv' into 'MLX86164.CUSTOMERS' (Target). A large blue circle indicates the status is 'COMPLETE'. Below this, a summary table shows: 100 Rows read, 100 Rows loaded, and 0 Rows rejected. It also lists the start time (08/22/2022 1:54:51 PM) and end time (08/22/2022 1:54:56 PM). To the right, a 'Notifications' panel shows a green checkmark indicating 'The data load job succeeded.' with the message 'Load Customers.csv from My Computer to MLX86164.CUSTOMERS' and the timestamp '2022/08/22, 01:54 PM'. Below the notifications, it says 'No errors'.

**11.** Repeat steps 5-10 to load the **sales.csv** data set into the same schema with the table name **sales**.

This screenshot shows the 'Select a load target' step in the data loading wizard. At the top, there are tabs for 'Source' (radio button selected), 'Target' (radio button selected), 'Define', and 'Finalize'. The 'Source' tab shows the message 'You are loading the file Sales representatives.csv into MLX86164.SALES'. The 'Target' tab shows a table list with 'SALES' and 'CUSTOMERS' selected. On the left, there's a 'Schema' section with 'MLX86164' selected. At the bottom right, there are 'Back' and 'Next' buttons.

## IBM Data Virtualization

The screenshot shows the IBM Data Virtualization interface. At the top, there's a navigation bar with links like Load Data, Load History, Tables, Views, Indexes, Aliases, MQTs, Sequences, and Application objects. Below this is a section titled "Load details" showing a summary of a completed job:

- My computer: Sales representatives.csv
- Target: MLX86164.SALES
- Status: COMPLETE
- Rows read: 103
- Rows loaded: 103
- Rows rejected: 0
- Start time: 08/22/2022 3:29:48 PM
- End time: 08/22/2022 3:29:53 PM

In the center, a large blue circle indicates the job status. To the right, a message says "The data load job succeeded. You can now work with your data." Below this, it says "No errors". A notifications panel on the right shows a green checkmark icon and the message "The data load job succeeded. Load Sales representatives.csv from My Computer to MLX86164.SALES" with a timestamp of "2022/08/22, 03:29 PM" and a "View details" link.

## Task 3: Create a project

You need a new project to store the virtualized data.

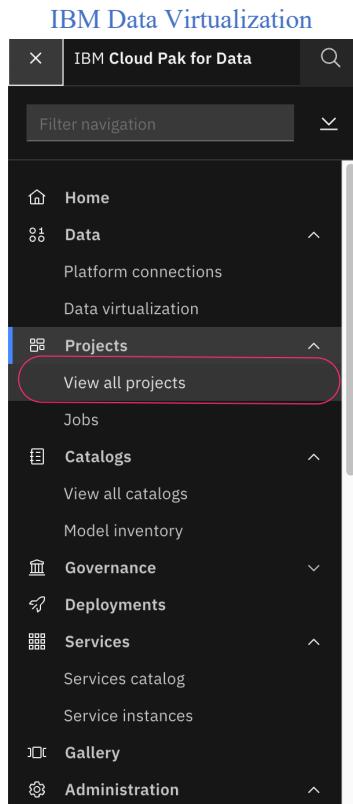
1. Login to CPD Platform.
2. From the Cloud Pak for Data navigation menu choose **Projects** > Select **View all Projects**.

The screenshot shows the IBM Cloud Pak for Data interface. At the top, there's a navigation bar with a menu icon (circled in red), a search bar, and user account information. The main dashboard features a "Welcome, Sandeep!" message and three main sections:

- Take a tutorial**: Step through implementing a Data fabric use case in a sample project. →
- Work with data**: Create a project for your team to prepare data, find insights, or build models. →
- Learn what's new**: Stay current with new features, enhancements, and other changes. →

On the right, there's a 3D graphic illustrating data flow and analysis. Below these sections are four main cards:

- Quick start**: A list of quick tasks including Build dashboards, Create data pipelines, Build customer profiles, Catalog and govern data, Build and manage ML models, and Query data anywhere.
- Projects**: Shows a list item: Multicloud Data Integration (Jul 25, 2022 06:31 PM) with a small error icon.
- Catalogs**: Shows a list item: Platform assets catalog (Jul 26, 2022 04:10 PM).
- Deployments**: Shows a list item: Mortgage Approval Catalog (Jul 25, 2022 07:12 PM).
- Notifications**: Shows a notification: Project import complete (Multicloud Data Integration was imported successfully). View import summary.
- New in gallery**: Shows a card for Scoring for Beijing PM 2.5.



3. You should have an existing project named Multicloud Data Integration, with default data source connections. Let's create another project to keep new data source connections. Click **New project** on your **Projects** page.

A screenshot of the 'Projects' page in IBM Cloud Pak for Data. The top navigation bar includes 'IBM Cloud Pak for Data', a search bar, and user account information. Below is a table titled 'Projects' with columns: Name, Date created, Your role, and Collaborators. A single row is visible for 'Multicloud Data Integration', created 3 days ago by 'Admin'. In the top right corner of the table header, there is a blue button labeled 'New project' with a '+' sign. A red arrow points to this button.

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

[← Back](#)

## Create a project

Choose whether to create an empty project or to preload your project with data and analytical assets. Add collaborators and data, and then choose the right tools to accomplish your goals. Add services as necessary.



### 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

Prepare and visualize data  
Analyze data in notebooks  
Train models



### Create a project from a sample or file

Get started fast by loading existing assets. Choose a project file from your system, or choose a curated sample project.

#### 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.
6. Choose an existing [object storage service instance](#) or create a new one.
7. Click **Create**.

## New project

### Define details

#### Name

Data Virtualisation Project

### Storage

CloudObjectStorage

#### Description

To Virtualise External Data and publish the virtualised view to catalog.

### Choose project options

Restrict who can be a collaborator ①

Project includes integration with [Cloud Object Storage](#) for storing project assets.

Cancel

Create

For more information, see [Creating a project](#).

The screenshot shows the IBM Data Virtualization interface. At the top, there's a navigation bar with 'IBM Cloud Pak for Data' and a search bar. Below the navigation bar, the page title is 'IBM Data Virtualization'. The main content area has tabs for 'Overview', 'Assets', 'Jobs', and 'Manage'. The 'Overview' tab is selected. It contains three main sections: 'Assets' (with a note about creating assets), 'Resource usage' (showing 0 CUH for the month), and 'Project history' (with a note about notifications). There are also 'Readme' and 'Type project notes, reminders, or instructions' sections.

Now Continuing on the below tasks, perform Task 4a and Optionally Task 4b.

## Task 4a: Add a connection to your Db2 data source

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

1. Switch back to Cloud Pak for Data as a Service.
2. From the main navigation menu select **Data > Data virtualization**. The list of configured *Data sources* displays.
3. Click **Add connection > New connection**.

The screenshot shows the 'Data sources' page in Cloud Pak for Data. At the top, there's a navigation bar with 'IBM Cloud Pak for Data' and a search bar. Below the navigation bar, the page title is 'Data virtualization / Data sources'. The main content area shows a table with columns 'Name', 'Endpoint', 'Type', and 'Status'. A tooltip for governance is visible, stating: 'Set up a primary catalog to enforce governance. If you want to enforce governance for your published assets, you need to set up a governed catalog to publish your assets to. Go to Service settings > Governance to enforce governance and select a primary catalog.' There are also buttons for 'Add connection', 'New connection', 'Existing platform connection', and 'Go to Governance'.

4. Select **Db2 on Cloud** (or **Db2 Warehouse on Cloud**). Click **Select**.

## IBM Data Virtualization

The screenshot shows the 'Data virtualization' workspace in IBM Cloud Pak for Data. A search bar at the top has 'db2' typed into it. Below the search bar, there's a sidebar with filters for 'Provider' (IBM, Third-party) and 'Private communication protocol' (Satellite, Secure Gateway). The main area displays a list of connection types under 'All connection types': IBM Db2, IBM Db2 Event Store, IBM Db2 on Cloud (selected), IBM Db2 Big SQL, IBM Db2 Hosted, and IBM Db2 Warehouse. To the right of the list is a panel titled 'Selected connection type' which provides details about 'IBM Db2 on Cloud' and lists compatible services like Catalogs, DataStage, and Data Virtualization.

5. Complete the connection details based on the credentials you copied for your instance in IBM Cloud.

The screenshot shows the 'Create connection: IBM Db2 on Cloud' page. On the left, a sidebar lists tabs: 'Connection overview' (selected), 'Connection details', 'Credentials', 'Certificates', 'Private connectivity', and 'Location and sovereignty'. The main area is titled 'Connection overview' and contains fields for 'Name' (set to 'Db2 on cloud bluedb') and 'Description' (set to 'Db2 on cloud bluedb'). At the bottom right of this section is a 'Test connection' button. Below this is another section titled 'Connection details' with a 'Cancel' button. At the very bottom right of the page are 'Back' and 'Create' buttons.

## IBM Data Virtualization

Create connection: IBM Db2 on Cloud

Enter the connection information.

**Connection overview**

**Connection details**

Database\* ⓘ bludb

Hostname or IP address\* ⓘ 6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud

Port ⓘ 30376

Credentials

Username\* ⓘ mlx86164

Password\* ⓘ ..... ⓘ

Certificates

Port is SSL-enabled ⓘ

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

6. Click **Test Connection**. The connection Should be successful. Click **Create** to create the connection.

The test was successful.  
Click Create to save the connection information.

Create connection: IBM Db2 on Cloud

Enter the connection information.

**Connection overview**

**Connection details**

Database\* ⓘ bludb

Hostname or IP address\* ⓘ 6667d8e9-9d4d-4ccb-ba32-21da3bb5aafc.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud

Port ⓘ 30376

Credentials

Username\* ⓘ mlx86164

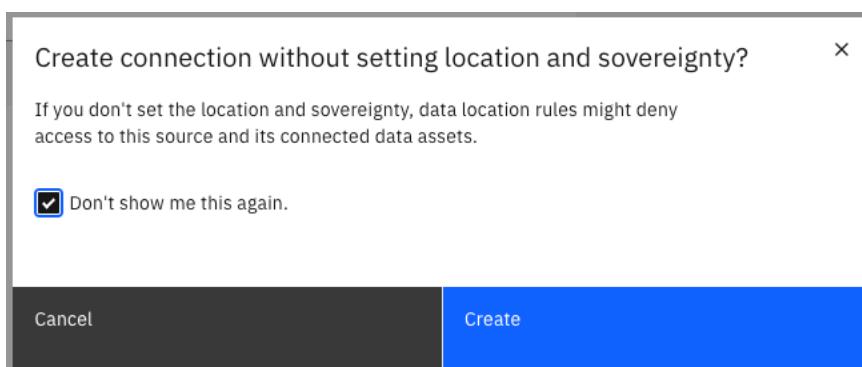
Password\* ⓘ ..... ⓘ

Certificates

Port is SSL-enabled ⓘ

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

7. Click **Create** to confirm.



## IBM Data Virtualization

### 8. The connection should be added Successfully.

The screenshot shows the 'Data sources' page in the IBM Cloud Pak for Data interface. At the top, there's a banner with a blue icon and the text 'Set up a primary catalog to enforce governance'. Below it, a note says 'If you want to enforce governance for your published assets, you need to set up a governed catalog to publish your assets to. Go to Service settings > Governance to enforce governance and select a primary catalog.' On the right, there are buttons for 'Go to Governance' and a close 'X' button. Below the banner, there are two tabs: 'Table view' (which is selected) and 'Constellation view'. A search bar at the top allows filtering by 'Name', 'Endpoint', and 'Type'. The main table lists connections with columns for 'Name', 'Endpoint', 'Type', and 'Status'. One connection, 'Db2 on cloud bluedb', is listed with its endpoint and type details. The status is 'Active' with a green checkmark.

For more information, see [Adding and connecting to data sources in Watson Query](#).

### Task 4b: 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 'Platform connections' page in the IBM Cloud Pak for Data interface. At the top, there's a search bar and a 'New connection' button. Below it, a section titled 'Connected data sources' with the sub-instruction 'Manage existing data source connections or create new connections that can be used across the platform.' A 'Filter by:' dropdown is set to 'All types'. The main table lists connections with columns for 'Name', 'Type', 'Created by', 'Modified by', and 'Last updated'. The connections listed are: 'Data Fabric Trial - Databases for PostgreSQL' (Type: IBM Cloud Databases for PostgreSQL), 'Db2 on cloud bluedb' (Type: IBM Db2 on Cloud), 'Data Fabric Trial - Db2 Warehouse' (Type: IBM Db2 Warehouse), and 'Data Fabric Trial - MongoDB' (Type: MongoDB). Each row has a small edit icon on the far right.

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

IBM Data Virtualization

#### 4. Enter the Connection Credentials as below and Click on **Test Connection**.

- Database Name: vlyawtap
- Host Name: tiny.db.elephantsql.com
- Port: 5432
- User Name: vlyawtap
- Password: fXt4TyCB\_W0d0LCaCaPF7MbLKWLpti60

Create connection: PostgreSQL

Enter the connection information.

## IBM Data Virtualization

Create connection: PostgreSQL

Enter the connection information.

**Credentials**

Username\* ⓘ  
vlyawtap

Password\* ⓘ  
.....

Port is SSL-enabled ⓘ

**Satellite Link** ⓘ  
Use a Satellite Link endpoint to allow any client that runs in your Satellite Location to connect to any service with access to this connection.

**Secure Gateway** ⓘ  
IBM Secure Gateway for IBM Cloud enables users to integrate cloud services with enterprise systems on premises.

**Location and sovereignty**  
Select the location and sovereignty of this source. If you don't set these properties, data location rules might deny access to this source and its connected data assets. [Learn more](#)

Location ⓘ

<https://dataplatform.cloud.ibm.com/home2?context=cpdaas>

5. The connection should be successful. Click **Create**.

Create connection: PostgreSQL

Enter the connection information.

**Credentials**

The test was successful.  
Click Create to save the connection information.

Username\* ⓘ  
vlyawtap

Password\* ⓘ  
.....

6. The new connection should be added. Now this connection can be added in Data Virtualization Data Sources.

Platform connections

**Connections** [Access control](#) [Supported connection types](#)

**Connected data sources**  
Manage existing data source connections or create new connections that can be used across the platform.

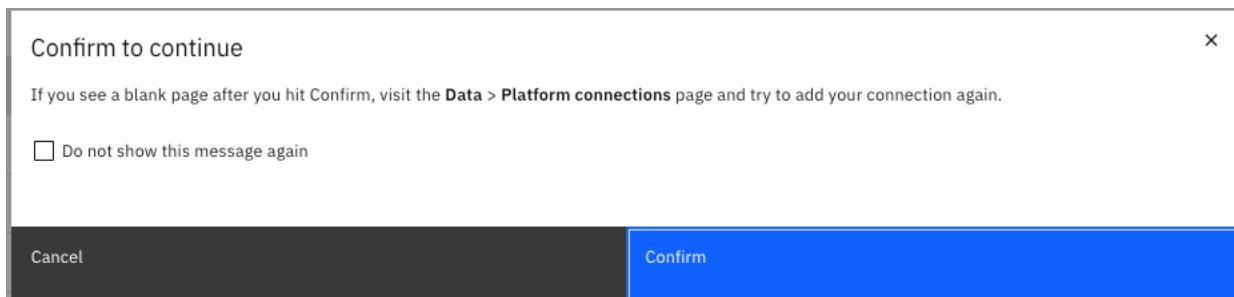
Filter by: All types ▾

Name	Type	Created by	Modified by	Last updated	Action
Db2 on cloud bluedb	IBM Db2 on Cloud	SANDEEP VED	SANDEEP VED	Jul 26, 2022	
PostgreSQL on ElephantsQL	PostgreSQL	SANDEEP VED	SANDEEP VED	Aug 23, 2022	

7. From the main navigation menu , select **Data > Data virtualization**. The list of configured *Data sources* displays.  
 8. Click **Add connection > Existing Platform connection**.

The screenshot shows the IBM Data Virtualization interface. At the top, there's a navigation bar with 'IBM Cloud Pak for Data', a search bar, and account information for 'SANDEEP VED's Account'. Below the navigation is a 'Data sources' dropdown menu. A prominent message box appears, stating: 'Set up a primary catalog to enforce governance. If you want to enforce governance for your published assets, you need to set up a governed catalog to publish your assets to. Go to Service settings > Governance to enforce governance and select a primary catalog.' There are two tabs at the bottom: 'Table view' (selected) and 'Constellation view'. On the right side of the interface, there's a sidebar with options like 'Add connection', 'New connection', 'Existing platform connection', and 'Go to Governance'.

## 9. Click **Confirm**.



## 10. Select **PostgreSQL** Connection and click **Add**. The connection will be added as a Data Source for Data Virtualization.

The screenshot shows the 'Add existing connection' screen. The left sidebar has 'Data sources' selected. The main area displays a table of existing connections. One connection, 'PostgreSQL on ElephantSQL', is selected (indicated by a blue circle). The table columns are: Name, Type, Created by, Modified by, and Last updated. The data rows are:

Name	Type	Created by	Modified by	Last updated
Data Fabric Trial - Databases for PostgreSQL	IBM Cloud Databases for PostgreSQL	SANDEEP VED	SANDEEP VED	Aug 23, 2022
Db2 on cloud bluedb	IBM Db2 on Cloud	SANDEEP VED	SANDEEP VED	Jul 26, 2022
Data Fabric Trial - Db2 Warehouse	IBM Db2 Warehouse	SANDEEP VED	SANDEEP VED	Aug 23, 2022
Data Fabric Trial - MongoDB	MongoDB	SANDEEP VED	SANDEEP VED	Aug 23, 2022
<b>PostgreSQL on ElephantSQL</b>	PostgreSQL	SANDEEP VED	SANDEEP VED	Aug 23, 2022

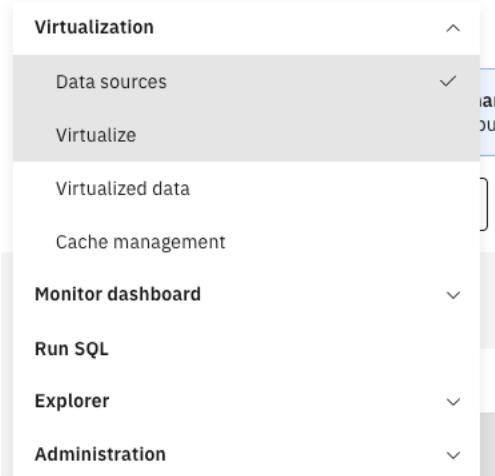
At the bottom are 'Cancel' and 'Add' buttons. The 'Add' button is highlighted in blue.

## Task 5: 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**, and wait for the available tables to load.

## Data sources ^



- Locate and select the **customers** and **sales** tables from the list, and click **Add to cart**.  
*You can select **Sales** Table from DB2 Connection and/or **Sales Rep** table from PostgreSQL Connection.*

This screenshot shows the 'Virtualize' interface under the 'Tables' tab. The search bar at the top has 'Find tables by name, schema, column, or business term' entered. The results table shows two items selected:

Table	Business terms	Schema	Connection	Hostname: Port	Database	Columns
<input checked="" type="checkbox"/> SALES	-	MLX86164	Db2 on cloud bluedb	6667d8e9-9d4d-4ccb...	bludb	11
<input checked="" type="checkbox"/> CUSTOMERS	-	MLX86164	Db2 on cloud bluedb	6667d8e9-9d4d-4ccb...	bludb	16

This screenshot shows the 'Virtualize' interface under the 'Tables' tab. The search bar at the top has 'Find tables by name, schema, column, or business term' entered. The results table shows one item selected:

Table	Business terms	Schema	Connection	Hostname: Port	Database	Columns
<input checked="" type="checkbox"/> SALES_REP	-	"sample"	Sales Tables in Postgre...	tiny.db.elephantsql.co...	vlyawtap	11

- Click **View cart**.

## IBM Data Virtualization

Showing matched tables: 2/62

Loaded: Aug 22, 2022 3:34 PM

Table	Business terms	Schema	Connection	Hostname: Port	Database	Columns
SALES	-	MLX86164	Db2 on cloud bluedb	6667d8e9-9d4d-4ccb...	bludb	11
CUSTOMERS	-	MLX86164	Db2 on cloud bluedb	6667d8e9-9d4d-4ccb...	bludb	16

Showing matched tables: 1/62

Loaded: not available

Table	Business terms	Schema	Connection	Hostname: Port	Database	Columns
SALES_REP	-	"sample"	Sales Tables in Postgre...	tiny.db.elephantsql.co...	vlyawtap	11

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

Review cart and virtualize tables

Empty cart Virtualize

Table	Schema	Source schema	Connections	Databases/File Path	Hostname: Port	Grouped tables
SALES	DV_IBMID_6650024CT3	MLX86164	Db2 on cloud bluedb	bludb	6667d8e9-9d4d-4ccb...	1
CUSTOMERS	DV_IBMID_6650024CT3	MLX86164	Db2 on cloud bluedb	bludb	6667d8e9-9d4d-4ccb...	1

Review cart and virtualize tables

Empty cart Virtualize

Table	Schema	Source schema	Connections	Databases/File Path	Hostname: Port	Grouped tables
SALES_REP	DV_IBMID_6650024CT3	"sample"	Sales Tables in Postgre...	vlyawtap	tiny.db.elephantsql.co...	1

- 5. 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.

[Cancel](#)

[Confirm](#)

- After the selected Customers and Sales/Sales\_Rep are virtualized, Click on **Go to Virtualized Data**.

### Virtualize objects

Table	Schema	Virtualization status
SALES	DV_IBMID_6650024CT3	Success
CUSTOMERS	DV_IBMID_6650024CT3	Success

[Virtualize more data](#)

[Go to virtualized data](#)

### Virtualize objects

Table	Schema	Virtualization status
SALES_REP	DV_IBMID_6650024CT3	Success

[Virtualize more data](#)

[Go to virtualized data](#)

- You can see the tables virtualized.

Table	Schema name	Created on
SALES_REP	DV_IBMID_6650024CT3	Aug 23, 2022 11:36 AM
joined_customers_sales_table	DV_IBMID_6650024CT3	Aug 22, 2022 3:42 PM
CUSTOMERS	DV_IBMID_6650024CT3	Aug 22, 2022 3:39 PM
SALES	DV_IBMID_6650024CT3	Aug 22, 2022 3:39 PM

Now Continuing on the below tasks, perform Either Task 6a or 6b.

## Task 6a: Publish virtualized data to a catalog and project

Join two tables from same IBM DB2 Data source, to create a virtualized asset and publish that to a catalog and project.

## IBM Data Virtualization

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

2. For each table, search for `salesrep`. Connect the **SALESREP\_ID** columns in the two tables.

3. Click **Preview** and then close popup and click **Next**.

CUST_ID	CUSTNAME	CITY	STATE	COUNTRY_CODE	POSTAL_CODE	EMAIL_ADDRESS	PHONE_NUMBER	YTD_SALES	SALESREP_ID
10102	Evelyn Conner	Andria	BA	IT	70031	Evelyn.R.Conner@mailinator.com	845-679-3051	904.86	SE208
10001	Michael Golden	Abbadia Alpina	TO	IT	10060	Michael.E.golden@spambob.com	724-454-8453	90.30	SE133
10002	Renee Mullins	Columbus	OH	US	45101	Rene.K.mullins@dodgeit.com	229-990-2162	0.00	NC169
10003	Allen Schmidt	Abeto	PG	IT	6040	Allen.M.schmidt@spambob.com	288-202-8653	304.00	RP385
10004	Robert May	Houston	TX	US	79601	Robert.C.may@spambob.com	630-492-6535	304.00	SE337
10005	Joe Cruz	Harrisonburg	VA	US	24210	Joe.D.cruz@spambob.com	448-788-1089	180.30	WE355

## IBM Data Virtualization

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

CUST_ID	CUSTNAME	CITY	STATE	COUNTRY_CODE	POSTAL_CODE	EMAIL_ADDRESS	PHONE_NUM
10102	Evelyn Conner	Andria	BA	IT	70031	Evelyn.R.Conner@mailinator.com	845-679-305
10001	Michael Golden	Abbadia Alpina	TO	IT	10060	Michael.E.golden@spambob.com	724-454-845
10002	Renee Mullins	Columbus	OH	US	45101	Rene.K.mullins@dodgeit.com	229-990-216
10003	Allen Schmidt	Abeto	PG	IT	6040	Allen.M.schmidt@spambob.com	288-202-865
10004	Robert May	Houston	TX	US	79601	Robert.C.may@spambob.com	630-492-653
10005	Joe Cruz	Harrisonburg	VA	US	24210	Joe.D.cruz@spambob.com	448-788-108

### 5. For the view name, type `joined_customers_sales_table`. Select a project from the list. Optionally Select the **Publish to catalog** option, and select a catalog. Click **Create view**.

Assign and review

View name: joined\_customers\_sales\_table

Schema name: DV\_IBMID\_6650024CT3

Publish to:

- Project [View](#)
- Virtualized data

Catalog [View](#)

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

Join view created

Assigned to: Data Virtualisation Project [View](#)

Published to: Mortgage Approval Catalog's catalog [View](#)

View	Schema	View creation status	Assignment status	Publish status
joined_customers_sales_table	DV_IBMID_6650024CT3	<span style="color: green;">✓ Success</span>	<span style="color: green;">✓ Success</span>	<span style="color: green;">✓ Success</span>

Virtualize more data      Go to virtualized data

### 7. You can either view the project or the catalog to preview the virtualized data.

## IBM Data Virtualization

The screenshot shows the IBM Data Virtualization interface. At the top, there's a navigation bar with 'IBM Cloud Pak for Data', a search bar, and account information for 'SANDEEP VED's Account' and 'Dallas'. Below the navigation is a breadcrumb trail: 'Data virtualization /'. The main content area is titled 'Virtualized data' and contains a search bar 'Find virtual objects'. A filter dropdown is set to 'All types'. A table lists three virtual objects:

Table	Schema name	Created on
joined_customers_sales_table	DV_IBMID_6650024CT3	Aug 22, 2022 3:42 PM
CUSTOMERS	DV_IBMID_6650024CT3	Aug 22, 2022 3:39 PM
SALES	DV_IBMID_6650024CT3	Aug 22, 2022 3:39 PM

At the top right of the table is a blue button labeled 'Add virtual objects +'

8. You will need an IBM Cloud API key to view the data in the project or catalog. See [Creating an IBM Cloud API key](#). Alternatively you can save the IBM Cloud API Key in the connection settings in project level.

The screenshot shows the IBM Cloud Pak for Data interface. At the top, there's a navigation bar with 'IBM Cloud Pak for Data', a search bar, and account information for 'SANDEEP VED's Account' and 'Dallas'. Below the navigation is a breadcrumb trail: 'Projects / Data Virtualisation Project / DV\_IBMID\_6650024CT3.joined\_\_...'. The main content area has tabs: 'Preview asset' (which is selected), 'Profile', and 'Visualization'. The 'Preview asset' tab shows a table schema with 26 columns. A 'Prepare data' button is visible. Below the schema, a message says 'Enter your personal credentials to unlock this connection and access its associated assets.' A 'Check connection with personal credentials' button is shown. On the right side, there's an 'Information' panel with details about the data asset:

- Data asset: DV\_IBMID\_6650024CT3.join\_ed\_customers\_sales\_table
- Description: No description available for this asset
- Tags: connected-data
- Creator: SANDEEP VED
- Usage: Created on Aug 22, 2022, 03:42 PM

Below the main preview area, there's a separate 'Authentication method' section with options: 'API key' (selected) and 'Username and password'.

## IBM Data Virtualization

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 the navigation bar, it says 'Projects / Data Virtualisation Project / DV\_IBMID\_6650024CT3.joined\_...'. On the left, there are tabs for 'Preview asset', 'Profile', and 'Visualization'. The main area displays a table with 26 columns. A sample row is shown:

CUST_ID	CUSTNAME	CITY	STATE	COUNTRY_C...	POSTAL_C...	DV_IBMID_6650024CT3_CUSTOMERS_EMAIL_ADD...	DV_IBMID_6650024CT3_CUSTOMERS_PHONE_NU...
10102	Evelyn Conner	Andria	BA	IT	70031	Evelyn.R.Conner@mailinator.com	845-679-3051
10001	Michael Golden	Abbadia Alpina	TO	IT	10060	Michael.E.golden@spambob.com	724-454-8453
10002	Renee Mullins	Columbus	OH	US	45101	Rene.K.mullins@dodgeit.com	229-990-2162
10003	Allen Schmidt	Abeto	PG	IT	6040	Allen.M.schmidt@spambob.com	288-202-8653
10004	Robert May	Houston	TX	US	79601	Robert.C.may@spambob.com	630-492-6535
10005	Joe Cruz	Harrisonburg	VA	US	24210	Joe.D.cruz@spambob.com	448-788-1089
10006	Rebecca White	Acate	RG	IT	97011	Rebecca.white@spambob.com	828-268-4303
10352	Gary Neal	Chicago	IL	US	61701	Gary.N.neal@dodgeit.com	267-929-9454
10422	Steve Huynh	Detroit	MI	US	48614	Steve.A.huynh@dodgeit.com	644-101-7549
10007	Anthony Johnson	Achstetten		DE	88480	Anthony.H.johnson@pookmail.com	907-111-5490
10008	Alberto Fabian	Acquabona	BL	IT	32043	Alberto.T.fabian@pookmail.com	113-785-7672
10009	Ronald Gordon	Acquaicia Del Ca	LE	IT	73040	Ronald.J.gordon@dodgeit.com	994-960-8006
10010	Christopher Marc	Acquarola	SA	IT	84080	Christopher.V.marcello@trashymail.com	717-217-7230
10012	Chris Green	ACTON		UK	SY9 6CX	Chris.M.green@pookmail.com	933-753-6160
10013	Norbert Crawford	Lansing	MI	US	49301	Norbert.T.crawford@spambob.com	983-926-6221
10014	Mike Dunmire	Louisville	KY	US	47202	Mike.I.dunmire@snamhoh.com	778-473-5757

On the right, there's an 'Information' panel with details about the asset:

- Data asset: DV\_IBMID\_6650024CT3.joined\_customers\_sales\_table
- Description: No description available for this asset
- Tags: connected-data
- Creator: SANDEEP VED
- Usage: Created on Aug 22, 2022, 03:42 PM

For more information, see [Governing virtual data in Watson Query](#).

## Task 6b: Publish virtualized data to a catalog and project

Join two tables from IBM DB2 Data source and PostgreSQL Data Source, to create a virtualized asset and publish that to a catalog and project.

1. You can Select **Customers** and **Sales\_Rep** tables (Sales and Sales\_Rep contains same data, just difference data source connection). Click on **Join**.

The screenshot shows the 'Virtualized data' section in the IBM Cloud Pak for Data interface. It lists selected items:

Table	Schema name	Created on
<input checked="" type="checkbox"/> SALESREP	DV_IBMID_6650024CT3	Aug 23, 2022 11:36 AM
<input type="checkbox"/> joined_customers_sales_table	DV_IBMID_6650024CT3	Aug 22, 2022 3:42 PM
<input checked="" type="checkbox"/> CUSTOMERS	DV_IBMID_6650024CT3	Aug 22, 2022 3:39 PM
<input type="checkbox"/> SALES	DV_IBMID_6650024CT3	Aug 22, 2022 3:39 PM

Buttons at the bottom include: Manage access, Join, Assign, Publish to catalog, and Cancel.

2. For each table, search for salesrep. Connect the **SALESREP\_ID** columns in the two tables. Click on **Preview** Button.

## IBM Data Virtualization

Join virtual objects

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

Table: SALES\_REP | Schema: DV\_IBMID\_6650024CT3

<input type="checkbox"/> 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
<input checked="" type="checkbox"/> EMAIL_ADDRESS	CLOB
<input checked="" type="checkbox"/> MANAGER_ID	CLOB

Table: CUSTOMERS | Schema: DV\_IBMID\_6650024CT3

<input type="checkbox"/> Column name	Data type
<input checked="" type="checkbox"/> CUST_ID	SMALLINT
<input checked="" type="checkbox"/> CUSTNAME	VARCHAR
<input checked="" type="checkbox"/> CITY	VARCHAR
<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	VARCHAR
<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 then next on **Next**.

Join virtual objects

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

Table: SALES\_REP | Schema: DV\_IBMID\_6650024CT3

<input type="checkbox"/> Column name	New join preview										
<input checked="" type="checkbox"/> SALESREP_ID	SALESREP_ID	FIRST_NAME	LAST_NAME	NATIONALITY	NATIONAL_ID	PHONE_NUMBER	AGE	SEX	TERRITORY	EMAIL_ADDRESS	MANAGER_ID
SE133	SE133	Amy	Thompson	IT	845-184-238	243-717-8900	42	F	SouthEast	Amy.Thompson@abxyzertainment.com	MA000
NC169	NC169	Fredericka	King	ES	766-887-613	448-788-1089	88	F	NorthCentral	Fredericka_King@abxyzertainment.com	MA008
RP385	RP385	Margereta	Vance	US	439-589-844	553-715-8233	30	F	Massachusetts	Margereta_Vance@abxyzertainment.com	MA006
SE337	SE337	Nathaniel	Hawk	IT	111-128-943	433-971-3885	67	M	SouthEast	Nathaniel_Hawk@abxyzertainment.com	MA000
WE355	WE355	Elane	Patterson	US	851-613-458	406-630-1559	19	F	West	Elane_Patterson@abxyzertainment.com	MA030
SC325	SC325	Rosa	Newman	US	113-743-369	738-126-4234	28	F	SouthCentral	Rosa_Newman@abxyzertainment.com	MA002
					<input checked="" type="checkbox"/> YTD_SALES	DECIMAL					
					<input checked="" type="checkbox"/> SALESREP_ID	VARCHAR					
					<input checked="" type="checkbox"/> NATIONALITY	VARCHAR					
					<input checked="" type="checkbox"/> NATIONAL_ID	VARCHAR					

Table: CUSTOMERS | Schema: DV\_IBMID\_6650024CT3

<input type="checkbox"/> Column name									
<input checked="" type="checkbox"/> CUST_ID	SMALLINT								
<input checked="" type="checkbox"/> CUSTNAME	VARCHAR								
<input checked="" type="checkbox"/> CITY	VARCHAR								
<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	VARCHAR								
<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

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

Edit column names

Cancel Back Next

SALESREP_ID	FIRST_NAME	LAST_NAME	NATIONALITY	NATIONAL_ID	PHONE_NUMBER	AGE	SEX	TERRITORY
SE133	Amy	Thompson	IT	845-184-238	243-717-8900	42	F	SouthEast
NC169	Fredericka	King	ES	766-887-613	448-788-1089	88	F	NorthCentral
RP385	Margereta	Vance	US	439-589-844	553-715-8233	30	F	Massachusetts
SE337	Nathaniel	Hawk	IT	111-128-943	433-971-3885	67	M	SouthEast
WE355	Elane	Patterson	US	851-613-458	406-630-1559	19	F	West
SC325	Rosa	Newman	US	113-743-369	738-126-4234	28	F	SouthCentral

## IBM Data Virtualization

- For the view name, type `joined_customers_salesrep_table`. Select the newly created project from the list. Optionally select the **Publish to catalog** option and select a catalog, and Click **Create View**.

The screenshot shows the 'Assign and review' step in the IBM Cloud Pak for Data interface. It includes fields for 'Assign to (all tables will be assigned to the same project)', 'Publish to', and 'View name'. The 'View name' field contains `joined_customers_salesrep_table`, and the 'Schema name' field contains `DV_IBMID_6650024CT3`. The 'Publish to' section has 'Catalog' selected and 'Mortgage Approval Catalog' chosen.

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

The screenshot shows a confirmation dialog titled 'Join view created'. It states 'Assigned to: Data Virtualisation Project [View](#)' and 'Published to: Mortgage Approval Catalog's catalog [View](#)'. A table summarizes the creation status: 'joined\_customers\_salesrep\_table' was successful in all categories (View creation status, Assignment status, Publish status). Below the table are buttons for 'Virtualize more data' and 'Go to virtualized data'.

- You can either view the project or the catalog to preview the virtualized data. You will need an IBM Cloud API key to view the data in the project or catalog. See [Creating an IBM Cloud API key](#).

The screenshot shows the 'Virtualized data' list view. It displays a table of virtual objects, including tables like `joined_customers_salesrep_table`, `SALES_REP`, `joined_customers_sales_table`, `CUSTOMERS`, and `SALES`, along with their schema names (`DV_IBMID_6650024CT3`) and creation dates (e.g., Aug 23, 2022 11:41 AM).

For more information, see [Creating virtual objects in Watson Query](#).

## Task 7: Update IBM API Keys in Project Connection

To avoid supplying IBM API Key again and again while viewing/accessing the Connected Asset, add API Key in the project connection settings. Click on one of the Connection Asset in the Project.

## IBM Data Virtualization

IBM Cloud Pak for Data

Search in your workspaces

Buy ? SANDEEP VED's Account Dallas SV

Projects / Data Virtualisation Project

Overview Assets Jobs Manage

Find assets Import asset New asset +

**8 assets**

All assets

**Asset types**

- > Data access (2)
- ↓ Data (4)
- Data assets (4)
- > Visualizations (1)
- > Source Code (1)

**Data access**

Name	Created by	Last modified
Data Virtualization_c887a514-a9a1-42a1-b348-26511cc631cb Connection	SANDEEP VED (You)	1 hour ago SANDEEP VED (You)
DS16611541722370214 Connection	SANDEEP VED (You)	2 hours ago SANDEEP VED (You)

Data in this project

Drop data files here or browse for files to upload

Click on **Credentials** in the left pane.

IBM Cloud Pak for Data

Search in your workspaces

Buy ? SANDEEP VED's Account Dallas SV

Edit connection: IBM Data Virtualization

Review the connection information.

Test connection

**Connection overview**

Name: Data Virtualization\_c887a514-a9a1-42a1-b348-26511cc631cb

Description: Connection description

Connection details

Database: bludb

Cancel Save

Under **Credentials**, Select **API Key** as the **authentication method**.

IBM Cloud Pak for Data

Search in your workspaces

Buy ? SANDEEP VED's Account Dallas SV

Edit connection: IBM Data Virtualization

Review the connection information.

Test connection

**Credentials**

Authentication method\* ⓘ

Choose which credentials you want to provide

API key

Username and password

Port is SSL-enabled ⓘ

SSL certificate ⓘ

Private connectivity

Connect privately to your data using Satellite Link or Secure Gateway.

Satellite Link       Secure Gateway

Use a Satellite Link endpoint      IBM Secure Gateway for IBM

Cancel Save

Supply the IBM API Key as downloaded/noted [before as per prerequisites section](#). Click **Save**.

## Task 7: Next Actions (Optional)

Now your virtual data is ready to be used. You can now optionally perform:

- [Manage user access to virtual objects](#)
- [Add the data asset to a catalog to share it with your organization](#)
- [Analyze the data in a notebook](#)
- [Build and train a model with the data asset](#)
- [Visualize the data with a dashboard](#)