# Lab Guide

# Hands-on-Lab
# Virtualizing Db2 Warehouse data with data virtualization

Shivam R Solanki
Data Scientist
Shivam.raj.solanki@ibm.com

Cloud Pak for Data Workshop 2020

For decades, companies have tried to break down silos by copying data from different operational systems into central data stores for analysis, such as data marts, data warehouses, and data lakes. This is often costly and prone to error. Most struggle to manage an average of 33 unique data sources, which are diverse in structure and type, and are often trapped in data silos that are hard to find and access.

With *data virtualization,* you can query data across many systems without having to copy and replicate data, which reduces costs. It also can simplify your analytics and make them more up to date and accurate because you're querying the latest data at its source.

In this tutorial, we're going to learn how to virtualize Db2® Warehouse data with data virtualization on IBM Cloud Pak for Data to make queries across multiple data sources.

### Learning objectives

In this tutorial, you will learn how to:

- Setup the project on IBM Cloud Pak for Data
- Add datasets to IBM Cloud Pak® for Data.
- Add a data source for data virtualization.
- Virtualize the data and create a joined view.
- Assign virtualized data to a project.
- Add roles to users and perform admin tasks.

## Steps

### Step 1. About the dataset

The dataset used for this tutorial contains information about fraud auto insurance claims for an insurance company. The data was collected in three CSV files.

1. ***claims.csv***

Some of the important attributes in this file are

- Claim Id

- Capital Gains *($)*
- Capital Loss *($)*
- Incident Type *(Single Vehicle collision, Vehicle Theft etc.)*
- Collison Type *(Read Collision, Side Collision etc.)*
- Incident Severity *(Minor Damage, Total Loss etc.)*
- Authorities Contacted *(Ambulance, Police etc.)*
- Incident Hour of the day
- Number of vehicles involved
- Witnesses
- Total claim amount *($)*
- Fraud reported *(Yes, No)*

2. ***customer.csv***

Some of the important attributes in this file are

- Customer Id
- Insured Sex *(Male, Female)*
- Insured Occupation *(Craft repair, sales etc.)*
- Insured Hobbies *(Chess, Cross fit etc.)*

3. ***policies.csv***

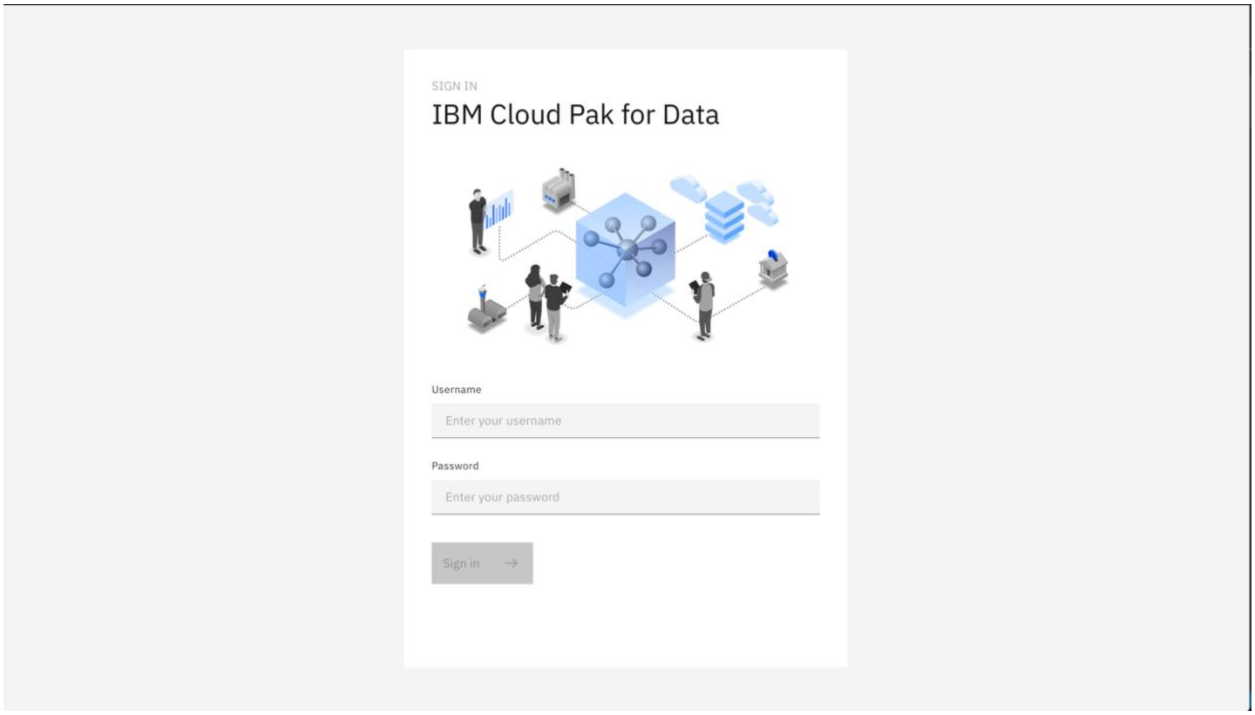
Some of the important attributes in this file are

- Claim Id
- Policy Id
- Coverage *(Basic, Premium etc.)*
- Policy Annual Premium *($)*
- Auto Make *(Dodge, Chevrolet etc.)*

## Step 2. Set up the project on IBM Cloud Pak for Data

### Log in to IBM Cloud Pak for Data

1. Launch a browser and navigate to the IBM Cloud Pak for Data url.



SIGN IN

IBM Cloud Pak for Data

Username

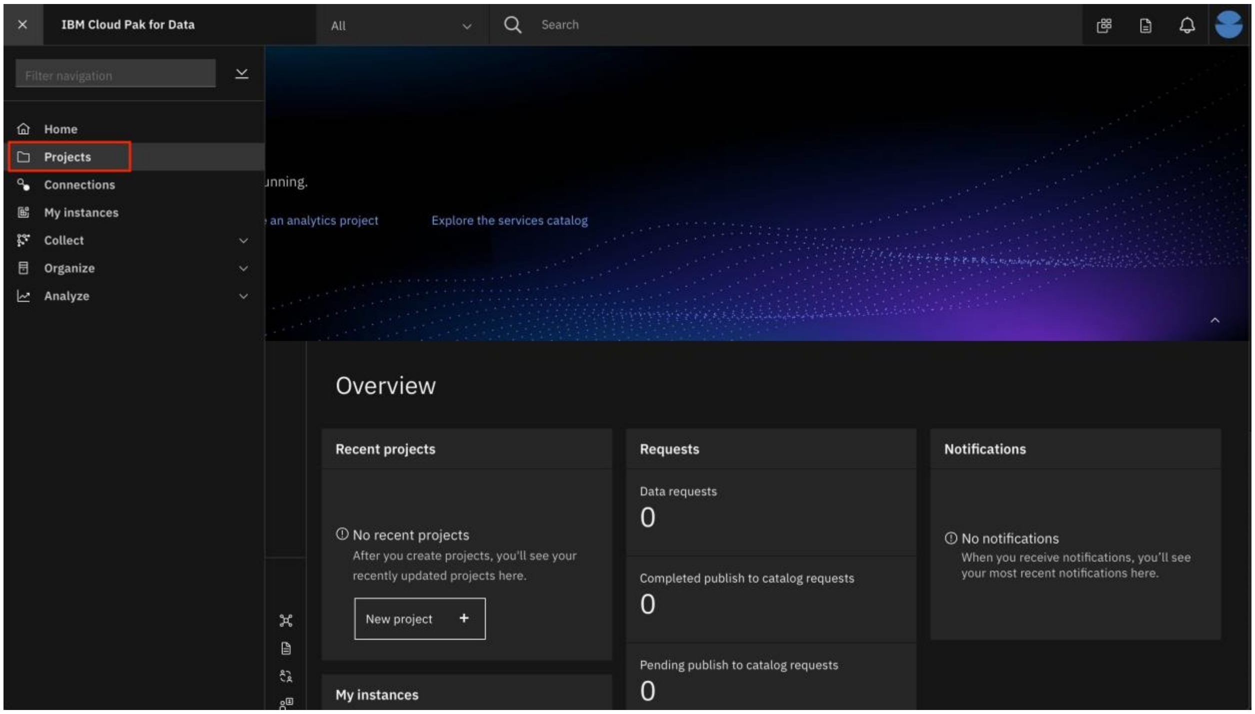
Enter your username

Password

Enter your password

### Create a new IBM Cloud Pak for Data project

1. Go the hamburger (☰) menu and click **Projects**.



No recent projects
After you create projects, you'll see your recently updated projects here.

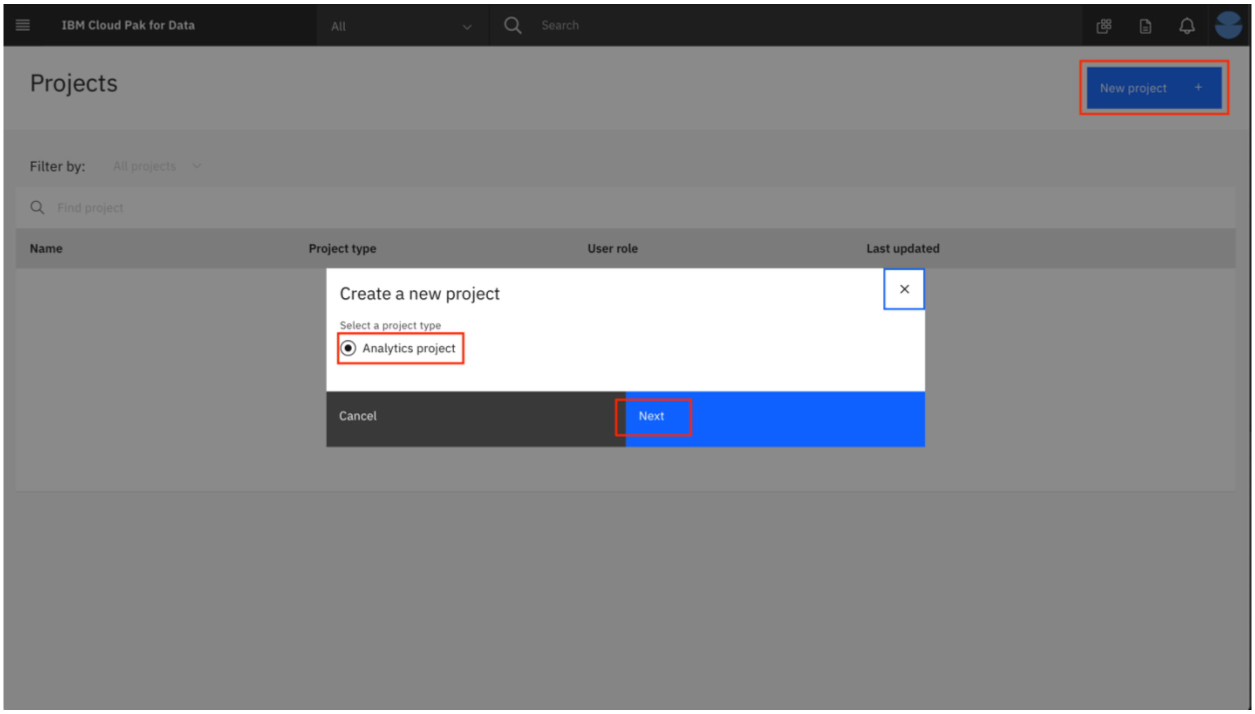
Data requests
0

Completed publish to catalog requests
0

Pending publish to catalog requests
0

No notifications
When you receive notifications, you'll see your most recent notifications here.

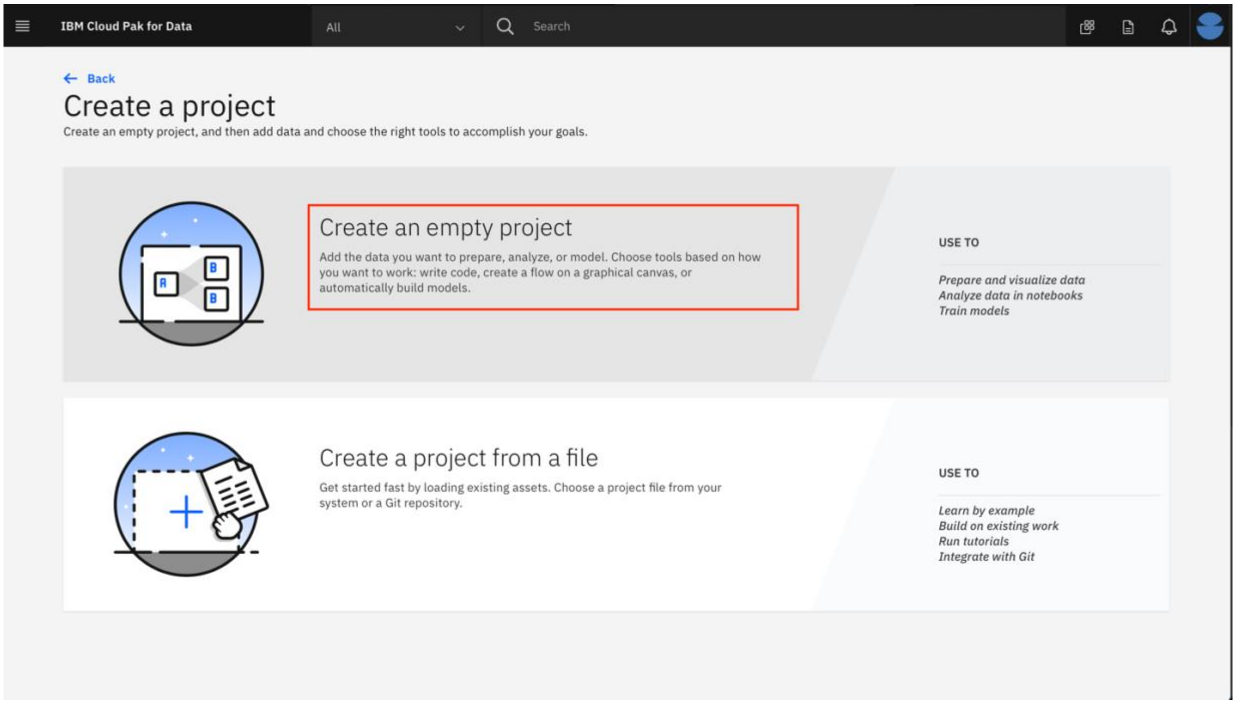
2. Click on **New project**.



Create a new project

Select a project type
● Analytics project

3. Select **Create an empty project**.



← Back

Create a project

Create an empty project, and then add data and choose the right tools to accomplish your goals.

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*

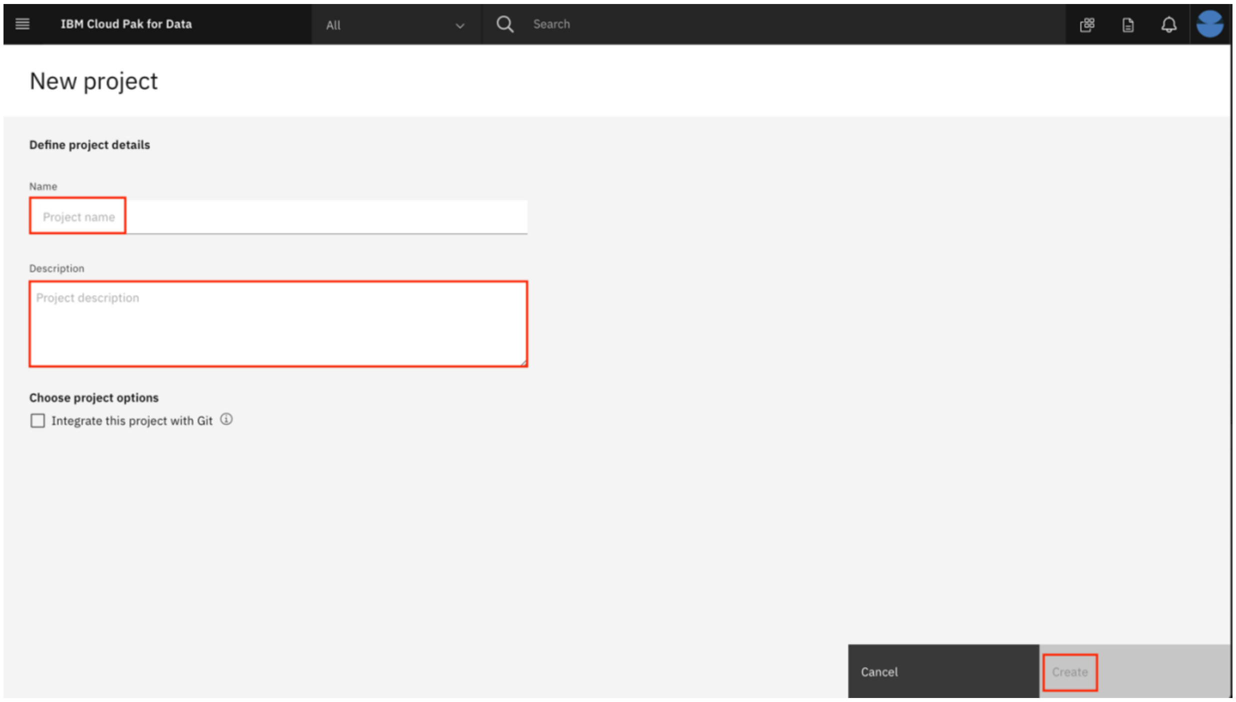
Create a project from a file

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

USE TO

*Learn by example*
*Build on existing work*
*Run tutorials*
*Integrate with Git*

4. Provide a name and optional description for the project and click **Create**. *(Note – For the purpose of this lab, please add your initials at the end of the project name to avoid conflict since all the participants are working in the same environment. For ex: Fraud-Claim-srs)*



New project

Define project details

Name

Project name

Description

Project description

Choose project options

□ Integrate this project with Git ⓘ

## Step 3. Set up the Db2 Warehouse on IBM Cloud

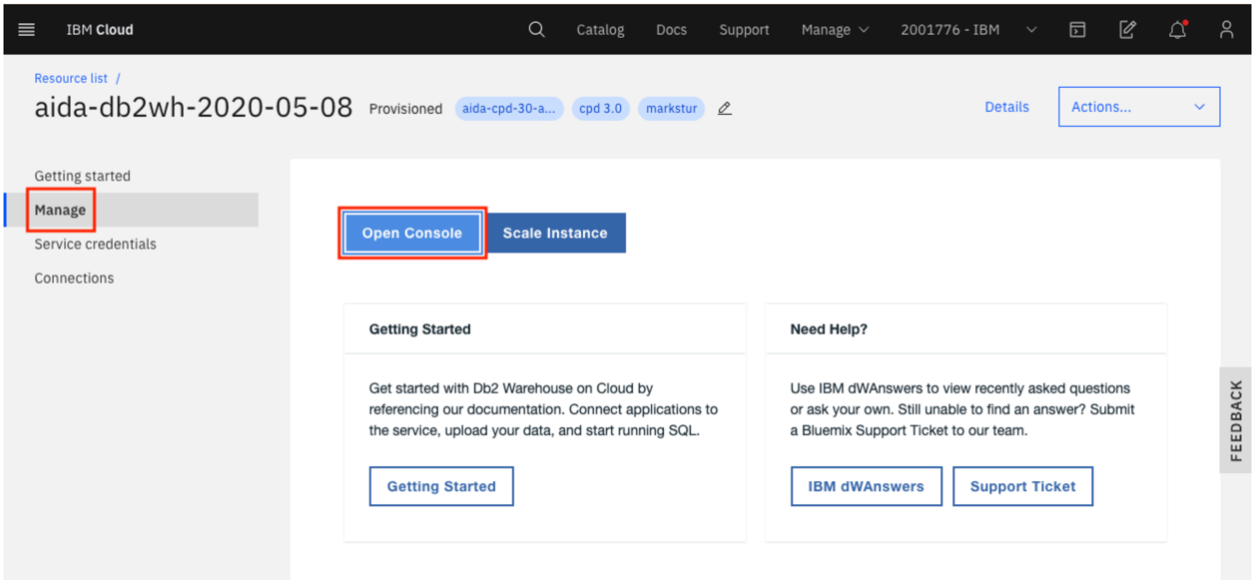
We’ll need a place to store our data. It is suggested to use Db2 Warehouse on IBM Cloud in order to conserve resources on the CPD cluster.

Note that IBM Cloud Pak for Data can work with any database with a JDBC connector, so Db2 warehouse is only one of many choices.

Steps below are shown for fetching connection details for Db2 Warehouse on IBM Cloud. *(Note - This step has already been done for you so you can use the credentials provided by the instructor. The screenshots shown here are for reference. You can skip to the Step 4 now)*

1. Go to Service Credentials and click **New credential +**. Click the **Copy to clipboard** icon and save the credentials for later.

2. Now go to Manage and click **Open Console**.



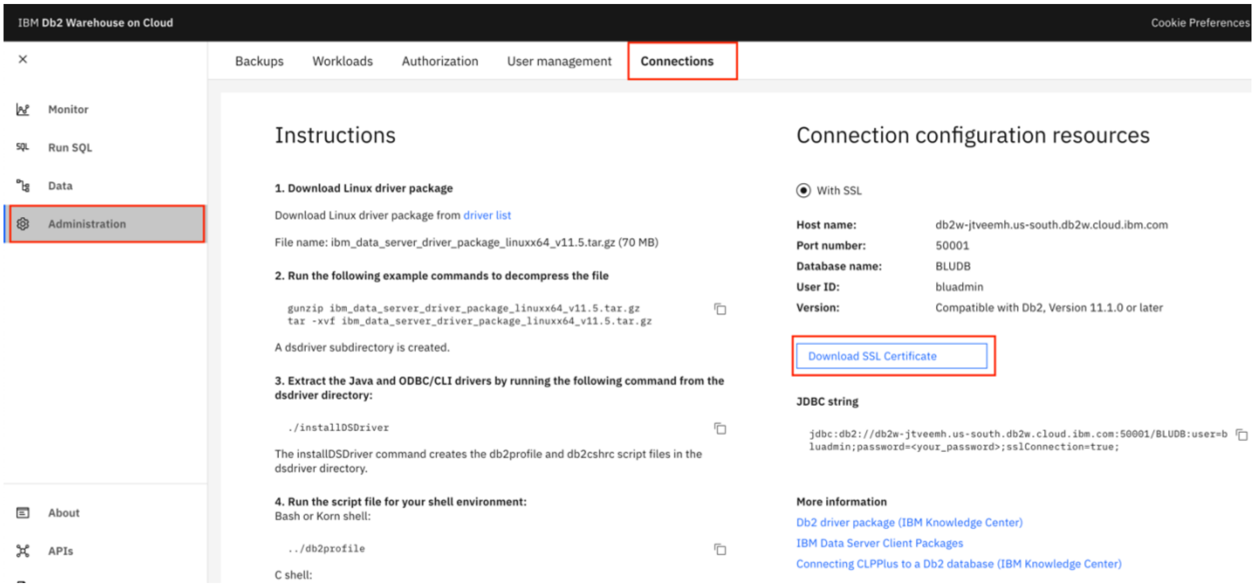
Get started with Db2 Warehouse on Cloud by referencing our documentation. Connect applications to the service, upload your data, and start running SQL.

Use IBM dWAnswers to view recently asked questions or ask your own. Still unable to find an answer? Submit a Bluemix Support Ticket to our team.

### Get SSL certificate for Db2 Warehouse on IBM Cloud

You will need an SSL cert for IBM Cloud Pak for Data to use the IBM Cloud Db2 Warehouse instance.

1. In the Db2 Warehouse console, from the upper-left hamburger (☰) menu, click **Administration → Connections**, then **Download SSL Certificate**.


IBM Db2 Warehouse on Cloud
Cookie Preferences
×
Monitor
Run SQL
Data
Administration
About
APIs
Backups
Workloads
Authorization
User management
Connections
Download SSL Certificate
driver list
Db2 driver package (IBM Knowledge Center)
IBM Data Server Client Packages
Connecting CLPPlus to a Db2 database (IBM Knowledge Center)

Instructions

**1. Download Linux driver package**

Download Linux driver package from driver list

File name: ibm_data_server_driver_package_linuxx64_v11.5.tar.gz (70 MB)

**2. Run the following example commands to decompress the file**

```
gunzip ibm_data_server_driver_package_linuxx64_v11.5.tar.gz
tar -xvf ibm_data_server_driver_package_linuxx64_v11.5.tar.gz
```

A dsdriver subdirectory is created.

**3. Extract the Java and ODBC/CLI drivers by running the following command from the dsdriver directory:**

```
./installDSDriver
```

The installDSDriver command creates the db2profile and db2cshrc script files in the dsdriver directory.

**4. Run the script file for your shell environment:**
Bash or Korn shell:

```
../db2profile
```

C shell:

Connection configuration resources

● With SSL

| | |
|---|---|
| **Host name:** | db2w-jtveemh.us-south.db2w.cloud.ibm.com |
| **Port number:** | 50001 |
| **Database name:** | BLUDB |
| **User ID:** | bluadmin |
| **Version:** | Compatible with Db2, Version 11.1.0 or later |

**JDBC string**

```
jdbc:db2://db2w-jtveemh.us-south.db2w.cloud.ibm.com:50001/BLUDB:user=b
luadmin;password=<your_password>;sslConnection=true;
```

**More information**

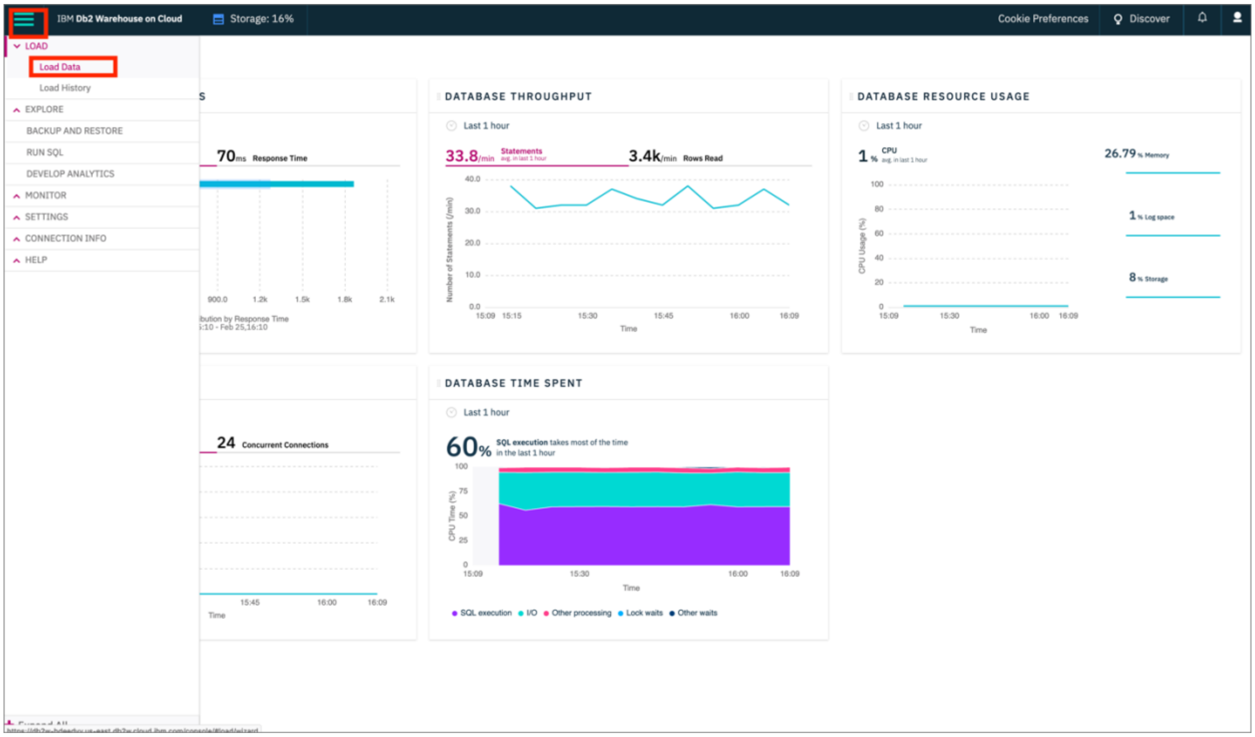
2. You will need to convert the SSL certificate from .crt to a .pem file using OpenSSL. Run the following command:

```
bash
openssl x509 -in DigiCertGlobalRootCA.crt -out DigiCertGlobalRootCA.pem -outform PEM -inform DER
```

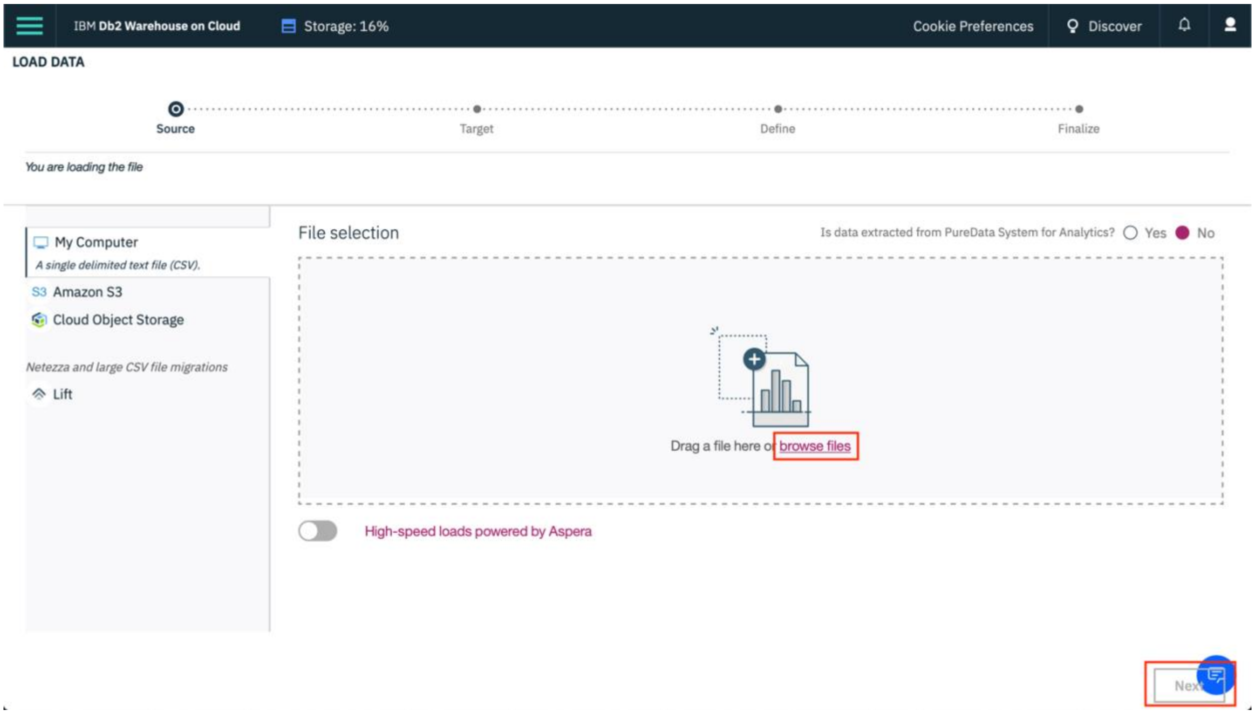
### Seed the Db2 Warehouse on IBM Cloud

Steps below are shown for uploading data to Db2 Warehouse on IBM Cloud. *(Note - The 3 data files claims.csv, customer.csv and policies.csv have been uploaded to Db2 Warehouse on IBM Cloud for you so you can skip to the Step 4 now)*

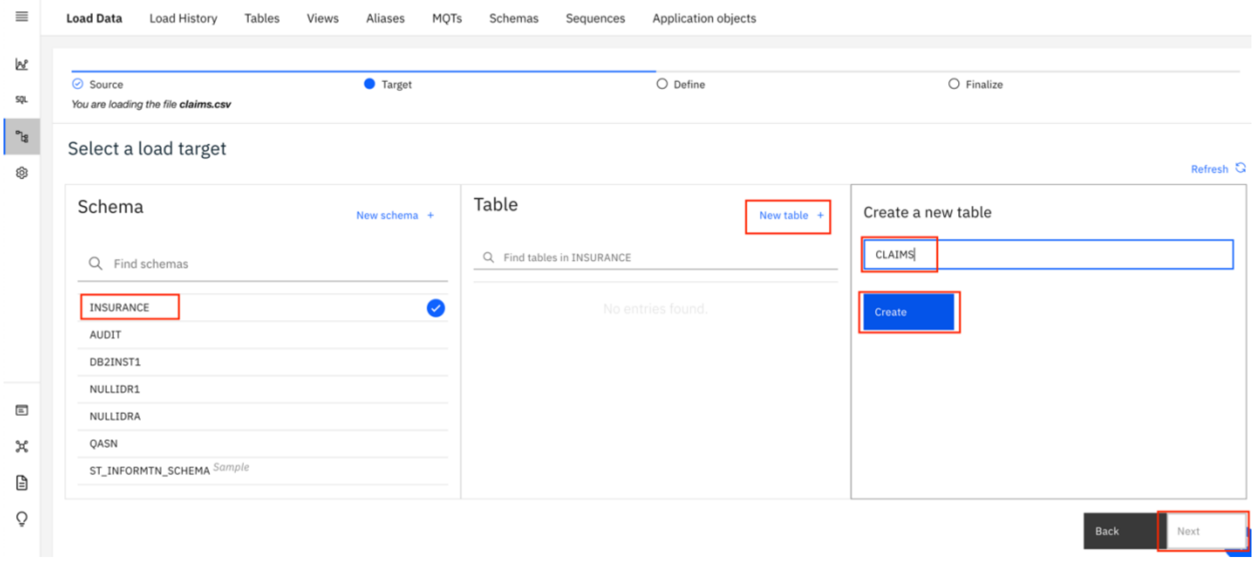
1. From the upper-left hamburger (☰) menu, click **LOAD** → **Load data**.



2. Click **browse files** and select the [claims.csv ](claims.csv) file after downloading it from the link, then click **Next**.


IBM Db2 Warehouse on Cloud
Storage: 16%
Cookie Preferences
Discover
LOAD DATA
Source
Target
Define
Finalize
You are loading the file
My Computer
A single delimited text file (CSV).
Amazon S3
Cloud Object Storage
Netezza and large CSV file migrations
Lift
File selection
Is data extracted from PureData System for Analytics? ○ Yes ● No
Drag a file here or browse files
High-speed loads powered by Aspera
Next

3. Choose Schema **INSURANCE** and click **+ New table**. Under Create a new Table, provide CLAIMS as the name of the table and click **Create**, then **Next**.



You are loading the file *claims.csv*

Select a load target

Schema

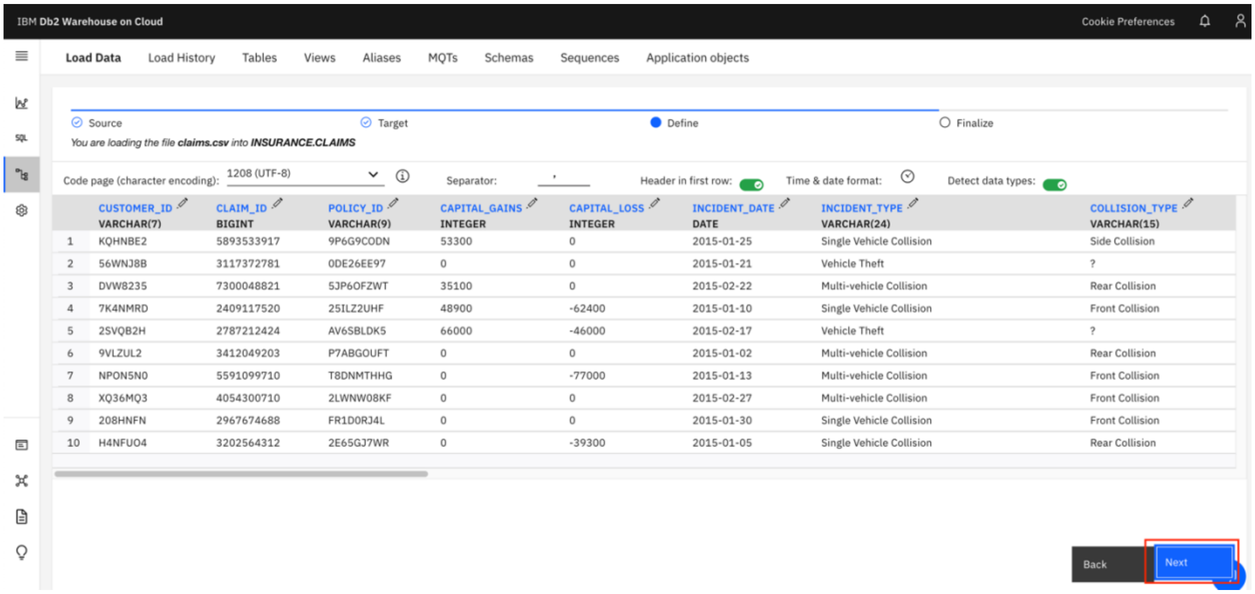
- INSURANCE ✓
- AUDIT
- DB2INST1
- NULLIDR1
- NULLIDRA
- QASN
- ST_INFORMTN_SCHEMA *Sample*

Table

Create a new table

CLAIMS

4. Accept the defaults and click **Next**. On the next screen, click **Begin Load**.



You are loading the file *claims.csv* into *INSURANCE.CLAIMS*

Code page (character encoding): 1208 (UTF-8)

Separator: ,

Header in first row: ●

Time & date format:

Detect data types: ●

| | CUSTOMER_ID | CLAIM_ID | POLICY_ID | CAPITAL_GAINS | CAPITAL_LOSS | INCIDENT_DATE | INCIDENT_TYPE | COLLISION_TYPE |
|---|---|---|---|---|---|---|---|---|
| | VARCHAR(7) | BIGINT | VARCHAR(9) | INTEGER | INTEGER | DATE | VARCHAR(24) | VARCHAR(15) |
| 1 | KQHNBE2 | 5893533917 | 9P6G9CODN | 53300 | 0 | 2015-01-25 | Single Vehicle Collision | Side Collision |
| 2 | 56WNJ8B | 3117372781 | 0DE26EE97 | 0 | 0 | 2015-01-21 | Vehicle Theft | ? |
| 3 | DVW8235 | 7300048821 | 5JP6OFZWT | 35100 | 0 | 2015-02-22 | Multi-vehicle Collision | Rear Collision |
| 4 | 7K4NMRD | 2409117520 | 25ILZ2UHF | 48900 | -62400 | 2015-01-10 | Single Vehicle Collision | Front Collision |
| 5 | 2SVQB2H | 2787212424 | AV6SBLDK5 | 66000 | -46000 | 2015-02-17 | Vehicle Theft | ? |
| 6 | 9VLZUL2 | 3412049203 | P7ABGOUFT | 0 | 0 | 2015-01-02 | Multi-vehicle Collision | Rear Collision |
| 7 | NPON5N0 | 5591099710 | T8DNMTHHG | 0 | -77000 | 2015-01-13 | Multi-vehicle Collision | Front Collision |
| 8 | XQ36MQ3 | 4054300710 | 2LWNW08KF | 0 | 0 | 2015-02-27 | Multi-vehicle Collision | Front Collision |
| 9 | 208HNFN | 2967674688 | FR1D0RJ4L | 0 | 0 | 2015-01-30 | Single Vehicle Collision | Front Collision |
| 10 | H4NFUO4 | 3202564312 | 2E65GJ7WR | 0 | -39300 | 2015-01-05 | Single Vehicle Collision | Rear Collision |

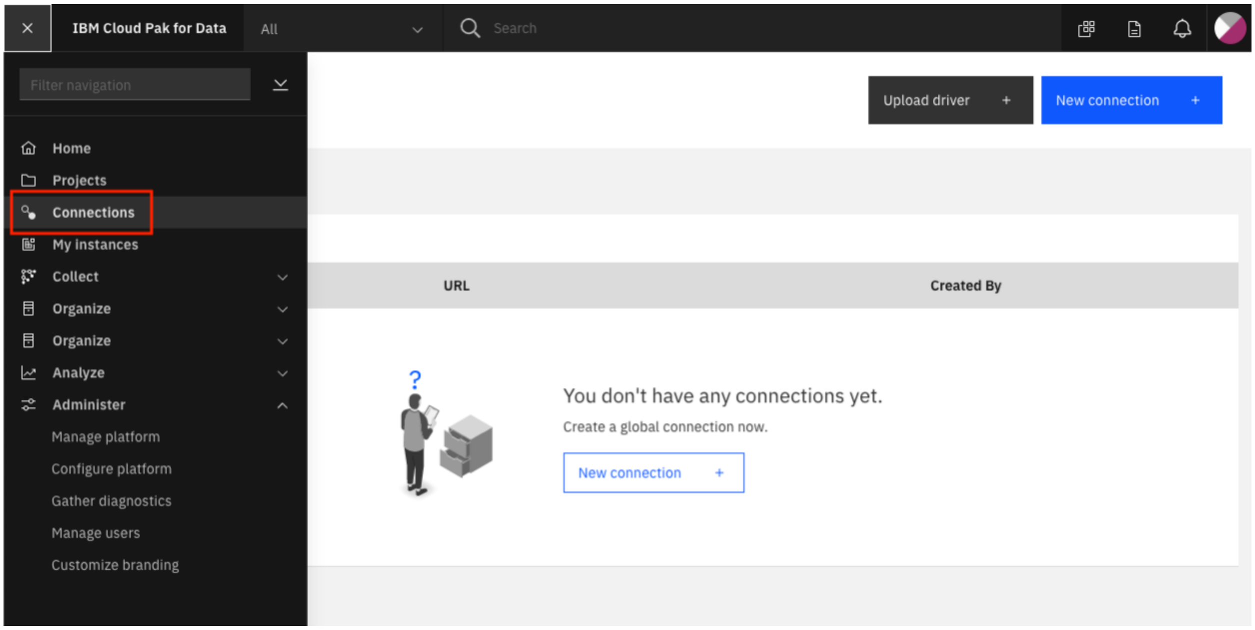
3. Repeat for the policies.csv file, naming the table `POLICIES` and the customers.csv table `CUSTOMERS`.

Now the Db2 warehouse has been set up on IBM Cloud.

## Step 4. Add a new data source connection

### Add the new data source

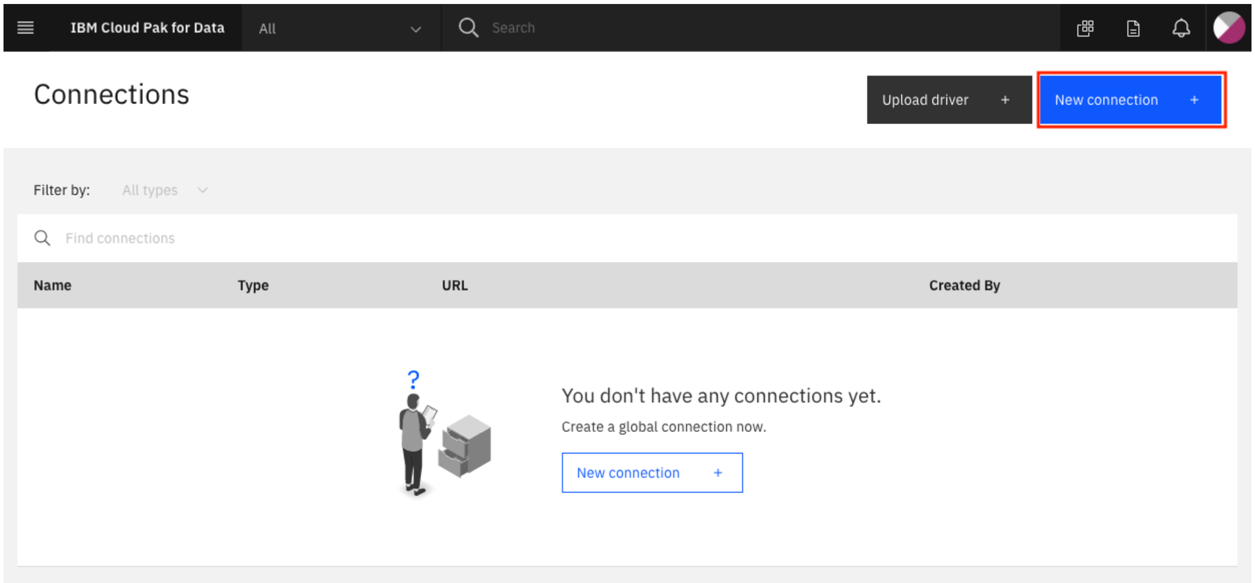
1. To add a new data source, go to the hamburger (☰) menu and click on the **Connections** option.



You don't have any connections yet.

Create a global connection now.

2. At the overview, click **New connection +**.



You don't have any connections yet.

Create a global connection now.

### Enter connection details for Db2 Warehouse on IBM Cloud

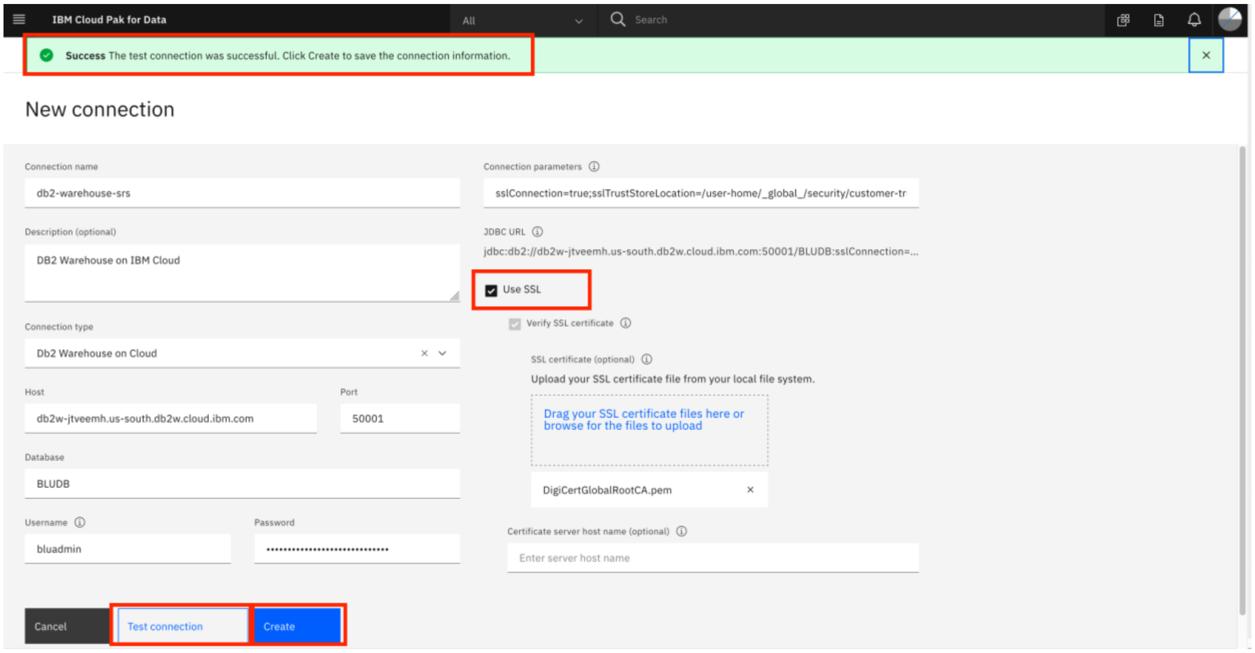
1. Start by giving your new connection a name, and select **Db2 Warehouse on Cloud** as your connection type. More fields should appear. Fill in the new fields with the credentials for your Db2 Warehouse connection. Click the checkbox for **Use SSL**.

   *(Note – Please add your initials at the end of the connection name with a dash. For example, db2-warehouse-srs in order to avoid conflict)*

2. Enter the following service credentials to create a connection to the Db2 warehouse on IBM Cloud.

   "**hostname**": "db2w-jtveemh.us-south.db2w.cloud.ibm.com",
   "**password**": "GxdualWj_f4aZq82Ah@4@KqmHAkEF",
   "**port**": 50001,
   "**db**": "BLUDB",
   "**username**": "bluadmin",

3. Download the SSL certificate and upload it to create connection.



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

New connection

Connection name: db2-warehouse-srs

Description (optional): DB2 Warehouse on IBM Cloud

Connection type: Db2 Warehouse on Cloud

Host: db2w-jtveemh.us-south.db2w.cloud.ibm.com

Port: 50001

Database: BLUDB

Username: bluadmin

Password: ••••••••••••••••••••••••••••

Connection parameters: sslConnection=true;sslTrustStoreLocation=/user-home/_global_/security/customer-tr

JDBC URL: jdbc:db2://db2w-jtveemh.us-south.db2w.cloud.ibm.com:50001/BLUDB:sslConnection=...

☑ Use SSL

☑ Verify SSL certificate

SSL certificate (optional)

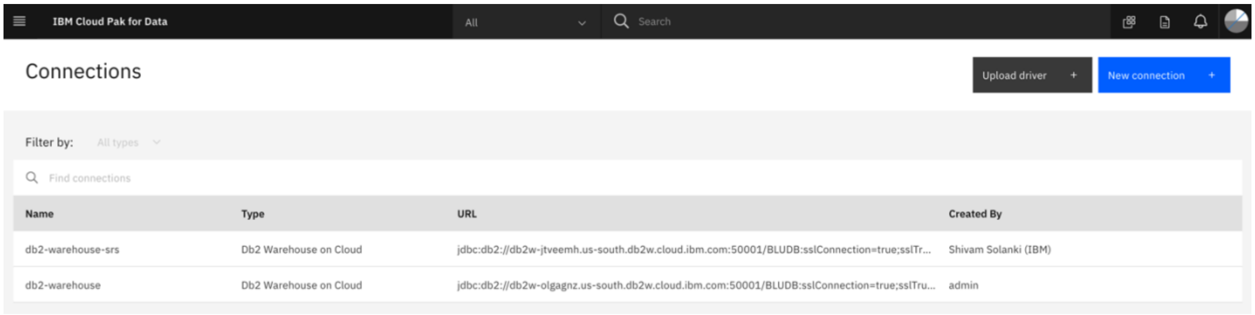
Upload your SSL certificate file from your local file system.

Drag your SSL certificate files here or browse for the files to upload

DigiCertGlobalRootCA.pem

Certificate server host name (optional): Enter server host name

The new connection will be listed in the overview.



| Name | Type | URL | Created By |
|---|---|---|---|
| db2-warehouse-srs | Db2 Warehouse on Cloud | jdbc:db2://db2w-jtveemh.us-south.db2w.cloud.ibm.com:50001/BLUDB:sslConnection=true;sslTr... | Shivam Solanki (IBM) |
| db2-warehouse | Db2 Warehouse on Cloud | jdbc:db2://db2w-olgagnz.us-south.db2w.cloud.ibm.com:50001/BLUDB:sslConnection=true;sslTru... | admin |

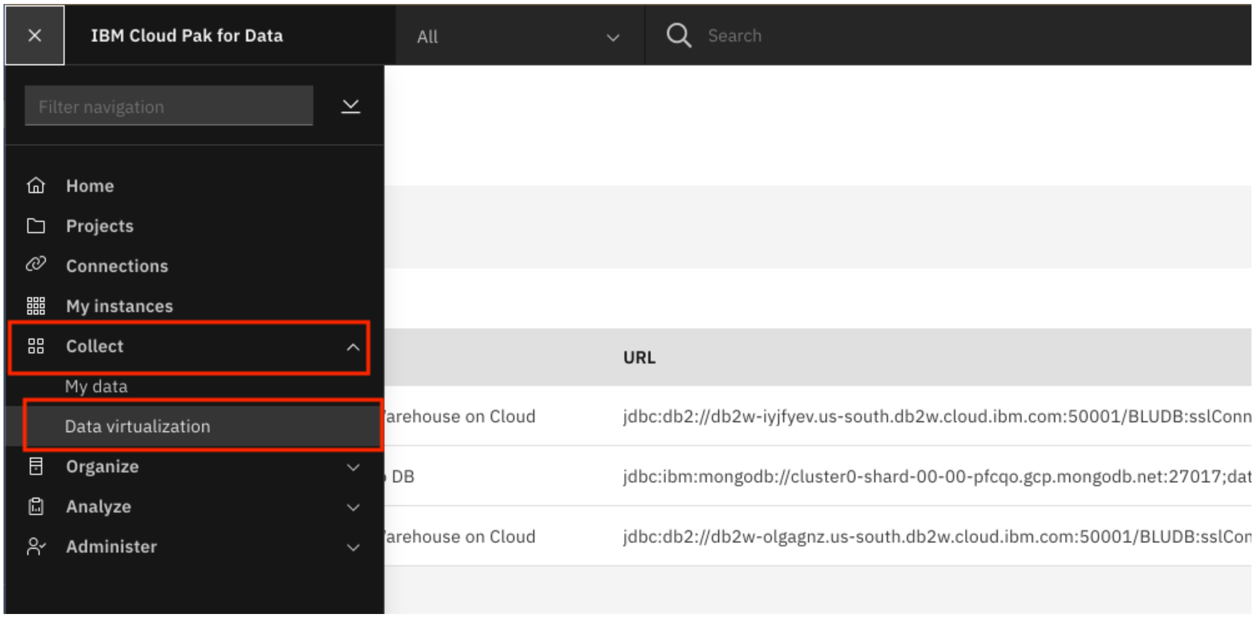
## Step 5. Virtualize Db2 data with data virtualization

**NOTE**: This section requires Admin user access to the IBM Cloud Pak for Data cluster.

For this section, we'll use the data virtualization tool to import the data from Db2 Warehouse, which is now exposed as a connection in IBM Cloud Pak for Data.

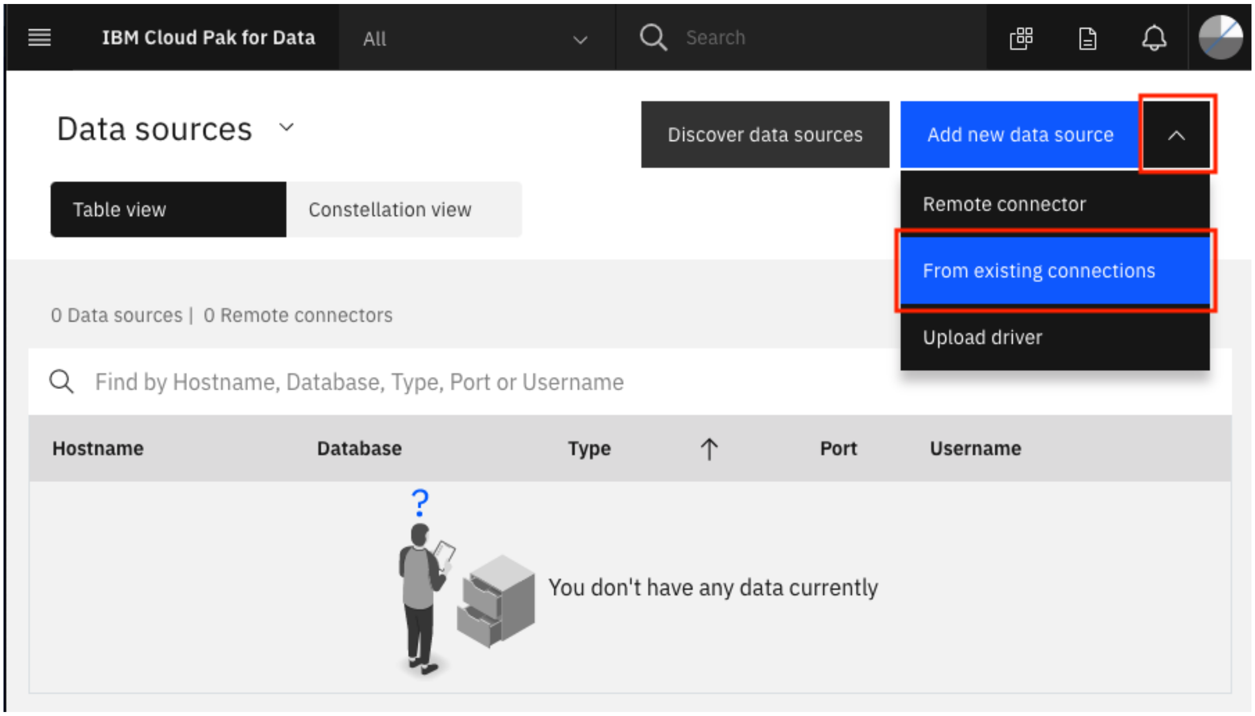
### Add a data source to Data Virtualization

1. To launch the data virtualization tool, go the hamburger (☰) menu and click **Collect** and then **Data Virtualization**.



| | URL |
|---|---|
| arehouse on Cloud | jdbc:db2://db2w-iyjfyev.us-south.db2w.cloud.ibm.com:50001/BLUDB:sslConn |
| DB | jdbc:ibm:mongodb://cluster0-shard-00-00-pfcqo.gcp.mongodb.net:27017;dat |
| arehouse on Cloud | jdbc:db2://db2w-olgagnz.us-south.db2w.cloud.ibm.com:50001/BLUDB:sslCon |

2. At the empty overview, click the drop-down next to Add new data source and select **From existing connections**

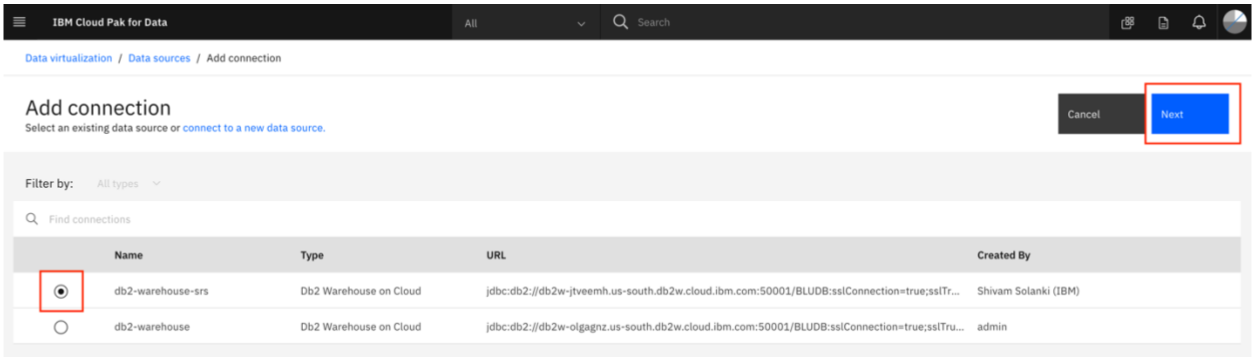


0 Data sources | 0 Remote connectors

| Hostname | Database | Type ↑ | Port | Username |
|---|---|---|---|---|

You don't have any data currently

3. Select the data source we made in the previous step and click **Next**.



Add connection

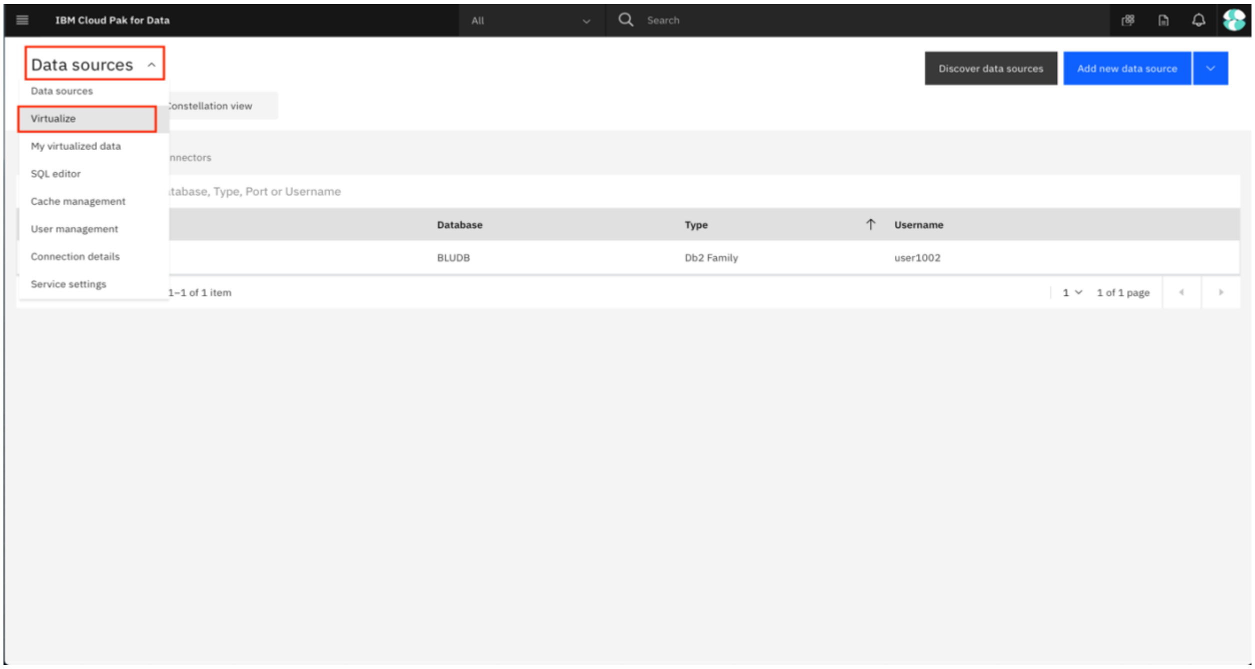
Select an existing data source or connect to a new data source.

| | Name | Type | URL | Created By |
|---|---|---|---|---|
| ● | db2-warehouse-srs | Db2 Warehouse on Cloud | jdbc:db2://db2w-jtveemh.us-south.db2w.cloud.ibm.com:50001/BLUDB:sslConnection=true;sslTr... | Shivam Solanki (IBM) |
| ○ | db2-warehouse | Db2 Warehouse on Cloud | jdbc:db2://db2w-olgagnz.us-south.db2w.cloud.ibm.com:50001/BLUDB:sslConnection=true;sslTru... | admin |

### Start virtualizing data

Because we now have access to the Db2 Warehouse data, we can virtualize the data to our IBM Cloud Pak for Data project.

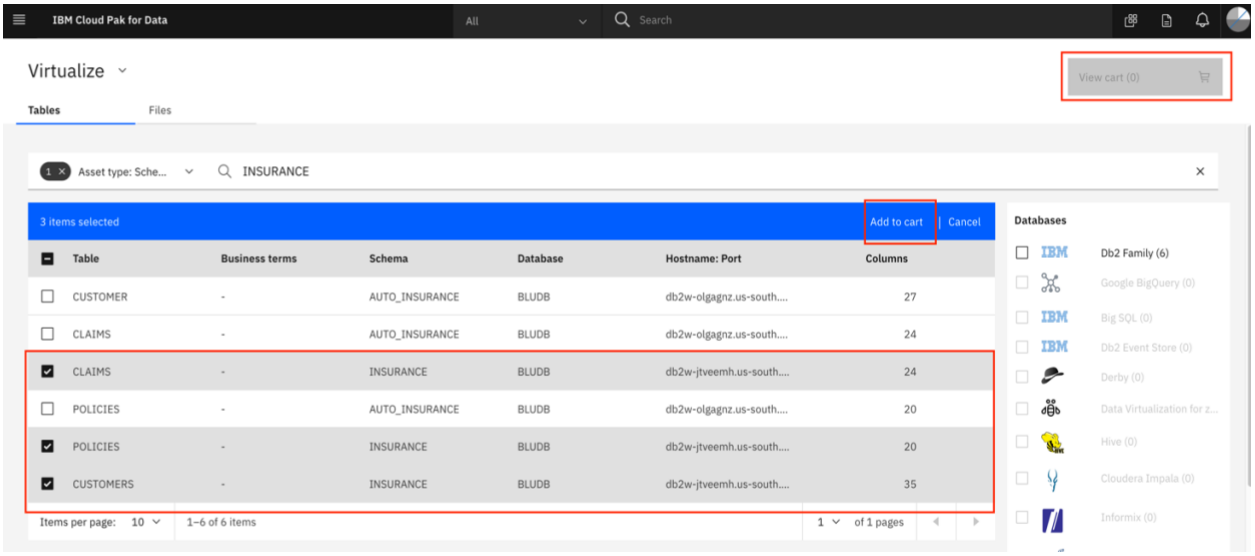
1. Click on the **Data Sources** drop-down and choose **Virtualize**.



| Database | Type | Username |
|---|---|---|
| BLUDB | Db2 Family | user1002 |

1–1 of 1 item  1  1 of 1 page

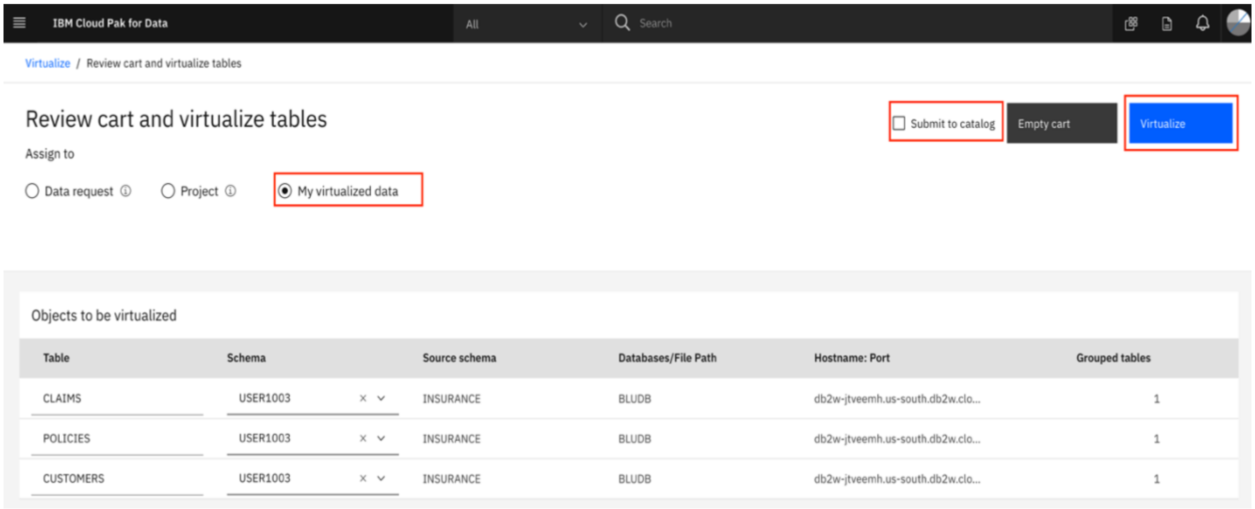
2. Several tables will appear (many are created as sample data when a Db2 Warehouse instance is provisioned) in the table. Find the tables that was created earlier, the previous instructions suggested naming them CUSTOMERS, POLICIES, and CLAIMS. Or you can search them by filtering on Schema = INSURANCE. Once selected, click **Add to cart** and then **View Cart**.



| | Table | Business terms | Schema | Database | Hostname: Port | Columns |
|---|---|---|---|---|---|---|
| □ | CUSTOMER | - | AUTO_INSURANCE | BLUDB | db2w-olgagnz.us-south.... | 27 |
| □ | CLAIMS | - | AUTO_INSURANCE | BLUDB | db2w-olgagnz.us-south.... | 24 |
| ☑ | CLAIMS | - | INSURANCE | BLUDB | db2w-jtveemh.us-south.... | 24 |
| □ | POLICIES | - | AUTO_INSURANCE | BLUDB | db2w-olgagnz.us-south.... | 20 |
| ☑ | POLICIES | - | INSURANCE | BLUDB | db2w-jtveemh.us-south.... | 20 |
| ☑ | CUSTOMERS | - | INSURANCE | BLUDB | db2w-jtveemh.us-south.... | 35 |

Items per page: 10  1–6 of 6 items  1  of 1 pages

3. The next panel prompts the user to choose which project to assign the data to. Choose **My virtualized data** and *uncheck* the box that says Submit to catalog. Click **Virtualize** to start the process.



Review cart and virtualize tables

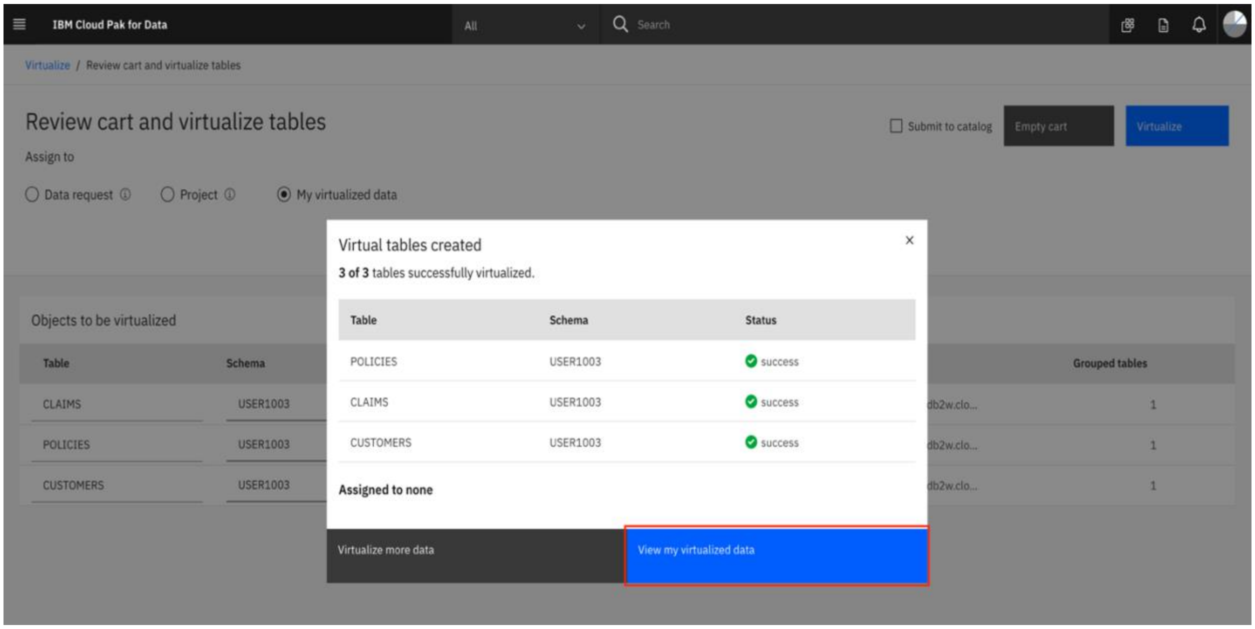
Assign to

○ Data request ○ Project ● My virtualized data

Objects to be virtualized

| Table | Schema | Source schema | Databases/File Path | Hostname: Port | Grouped tables |
|---|---|---|---|---|---|
| CLAIMS | USER1003 | INSURANCE | BLUDB | db2w-jtveemh.us-south.db2w.clo... | 1 |
| POLICIES | USER1003 | INSURANCE | BLUDB | db2w-jtveemh.us-south.db2w.clo... | 1 |
| CUSTOMERS | USER1003 | INSURANCE | BLUDB | db2w-jtveemh.us-south.db2w.clo... | 1 |

4. You’ll be notified that the virtual tables have been created. Let’s see the new virtualized data from the data virtualization tool by clicking **View my virtualized data**.


IBM Cloud Pak for Data
All
Search
Virtualize / Review cart and virtualize tables
Review cart and virtualize tables
Assign to
○ Data request ○ Project ● My virtualized data
□ Submit to catalog
Empty cart
Virtualize
Objects to be virtualized
Virtualize more data
View my virtualized data

Virtual tables created

**3 of 3** tables successfully virtualized.

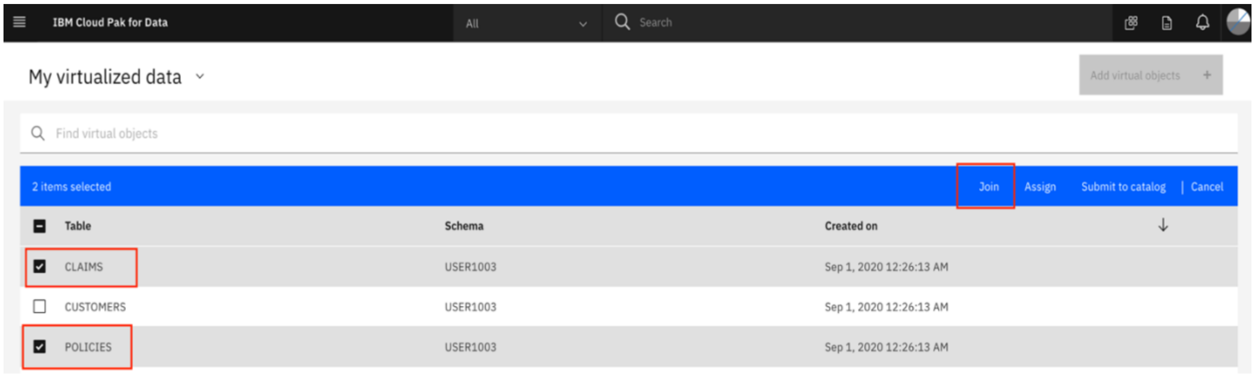
| Table | Schema | Status |
|---|---|---|
| POLICIES | USER1003 | ✓ success |
| CLAIMS | USER1003 | ✓ success |
| CUSTOMERS | USER1003 | ✓ success |

**Assigned to none**

### Join the virtualized data

Now we are going to *join* the tables we created so we have a merged set of data. It will be easier to do it here rather than in a notebook where we would have to write code to handle three different datasets.

1. Click on any two tables (CLAIMS and POLICIES, for instance), then click the **Join** button.



| Table | Schema | Created on |
|---|---|---|
| ☑ CLAIMS | USER1003 | Sep 1, 2020 12:26:13 AM |
| □ CUSTOMERS | USER1003 | Sep 1, 2020 12:26:13 AM |
| ☑ POLICIES | USER1003 | Sep 1, 2020 12:26:13 AM |

2. To join the tables, we need to pick a key that is common to both datasets. Here we choose to map ***claim_id*** and ***policy_id*** from the first table to ***claim_id*** and ***policy_id*** on the second table. Do this by clicking one and dragging it to the other. When the line is drawn, click **Next**.

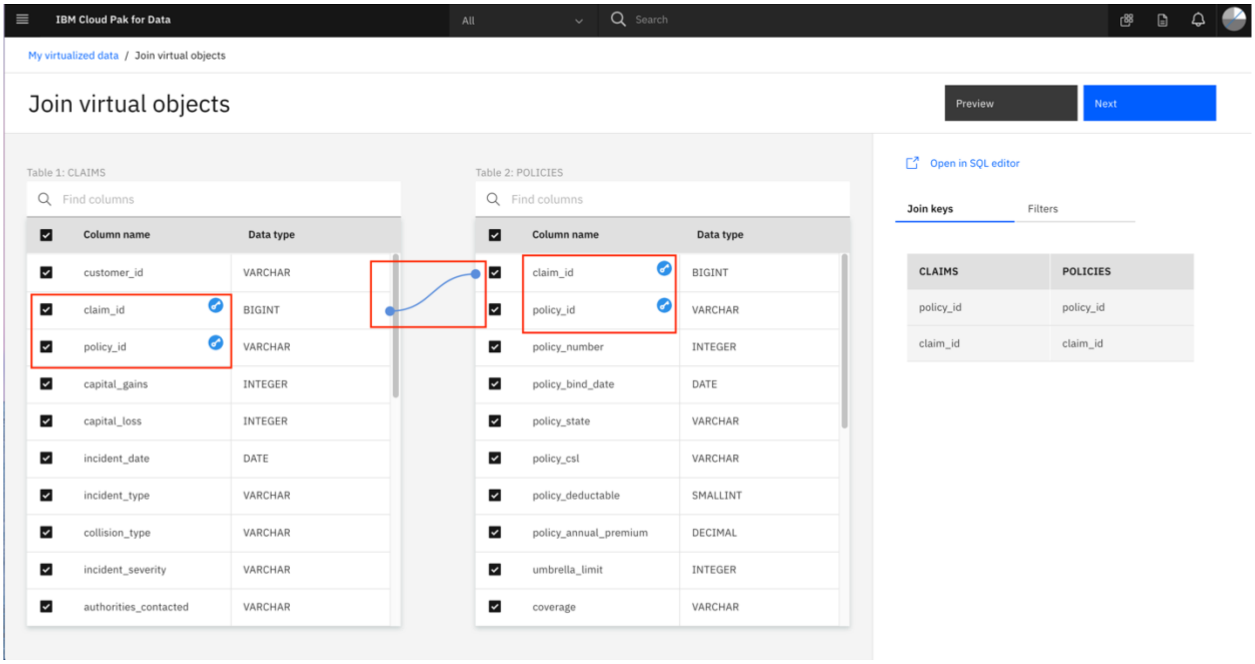


Table 1: CLAIMS

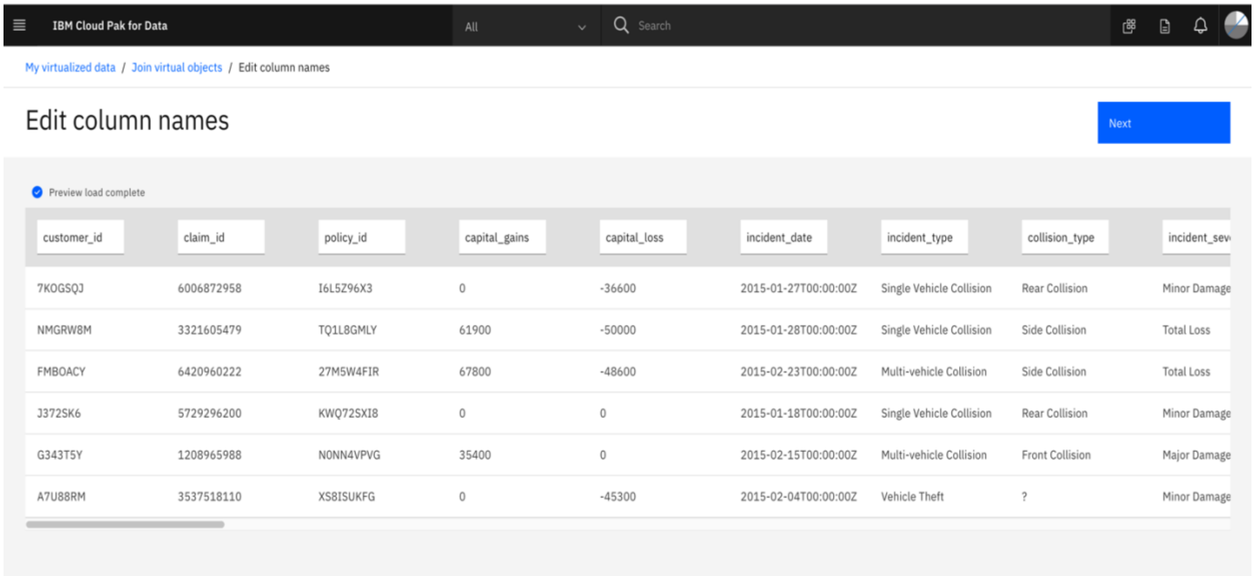
| Column name | Data type |
|---|---|
| ☑ customer_id | VARCHAR |
| ☑ claim_id | BIGINT |
| ☑ policy_id | VARCHAR |
| ☑ capital_gains | INTEGER |
| ☑ capital_loss | INTEGER |
| ☑ incident_date | DATE |
| ☑ incident_type | VARCHAR |
| ☑ collision_type | VARCHAR |
| ☑ incident_severity | VARCHAR |
| ☑ authorities_contacted | VARCHAR |

Table 2: POLICIES

| Column name | Data type |
|---|---|
| ☑ claim_id | BIGINT |
| ☑ policy_id | VARCHAR |
| ☑ policy_number | INTEGER |
| ☑ policy_bind_date | DATE |
| ☑ policy_state | VARCHAR |
| ☑ policy_csl | VARCHAR |
| ☑ policy_deductable | SMALLINT |
| ☑ policy_annual_premium | DECIMAL |
| ☑ umbrella_limit | INTEGER |
| ☑ coverage | VARCHAR |

| CLAIMS | POLICIES |
|---|---|
| policy_id | policy_id |
| claim_id | claim_id |

3. Next, you have a chance to edit column names, but we will keep them as-is. Click **Next**.



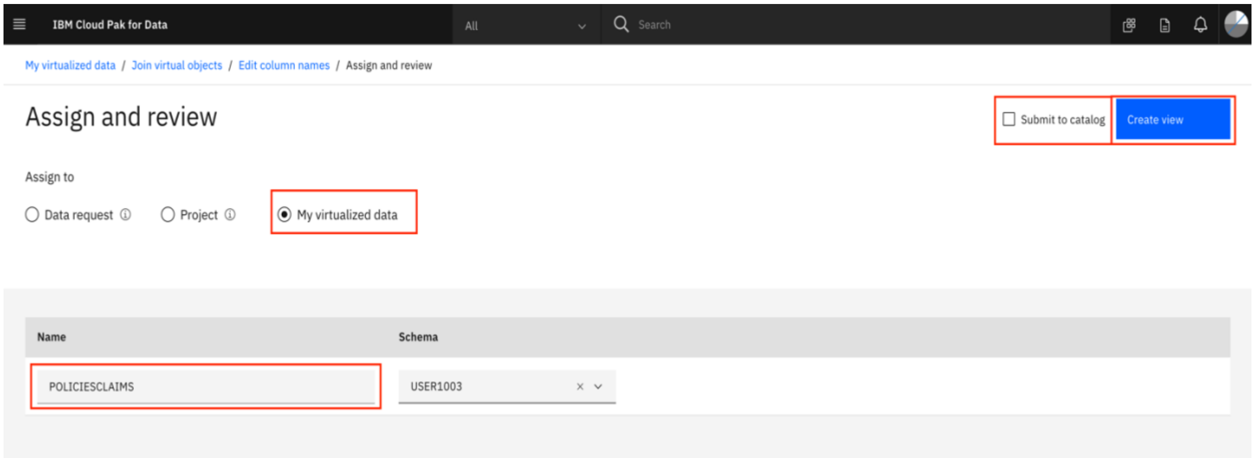
My virtualized data / Join virtual objects / Edit column names

Edit column names

Preview load complete

| customer_id | claim_id | policy_id | capital_gains | capital_loss | incident_date | incident_type | collision_type | incident_sev |
|---|---|---|---|---|---|---|---|---|
| 7KOGSQJ | 6006872958 | I6L5Z96X3 | 0 | -36600 | 2015-01-27T00:00:00Z | Single Vehicle Collision | Rear Collision | Minor Damage |
| NMGRW8M | 3321605479 | TQ1L8GMLY | 61900 | -50000 | 2015-01-28T00:00:00Z | Single Vehicle Collision | Side Collision | Total Loss |
| FMBOACY | 6420960222 | 27M5W4FIR | 67800 | -48600 | 2015-02-23T00:00:00Z | Multi-vehicle Collision | Side Collision | Total Loss |
| J372SK6 | 5729296200 | KWQ72SXI8 | 0 | 0 | 2015-01-18T00:00:00Z | Single Vehicle Collision | Rear Collision | Minor Damage |
| G343T5Y | 1208965988 | NONN4VPVG | 35400 | 0 | 2015-02-15T00:00:00Z | Multi-vehicle Collision | Front Collision | Major Damage |
| A7U88RM | 3537518110 | XS8ISUKFG | 0 | -45300 | 2015-02-04T00:00:00Z | Vehicle Theft | ? | Minor Damage |

4. In the next panel, we'll give our joined data a unique name such as POLICIESCLAIMS (to be consistent with SQL standards, pick an uppercase name). Under **Assign to**, choose **My virtualized data** and uncheck the box that says Submit to catalog. Click **Create view** to start the process.



My virtualized data / Join virtual objects / Edit column names / Assign and review

Assign and review

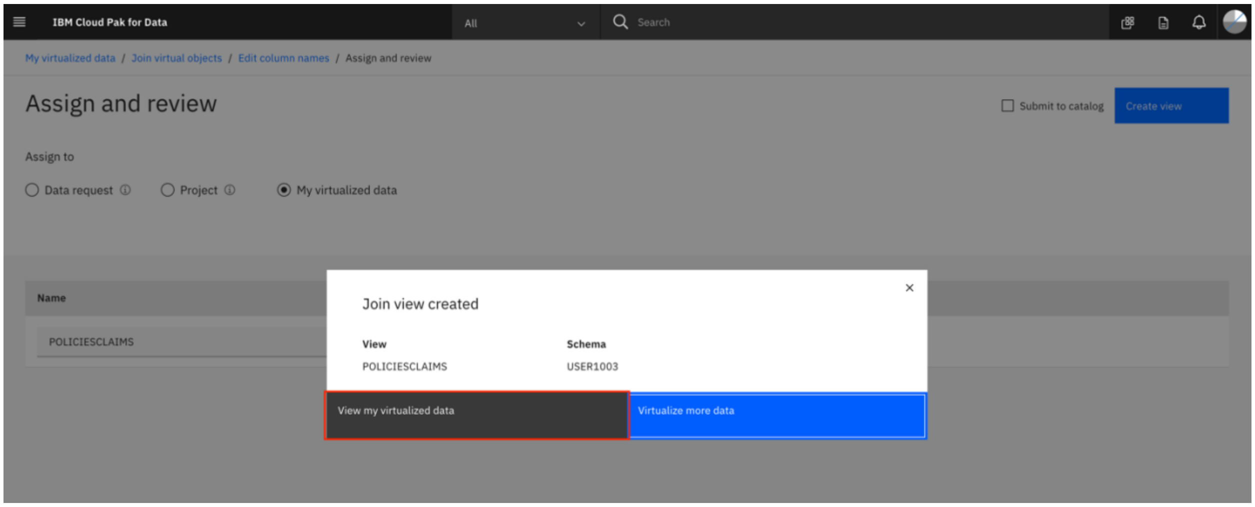
□ Submit to catalog

Assign to

○ Data request ○ Project ● My virtualized data

| Name | Schema |
|---|---|
| POLICIESCLAIMS | USER1003 |

5. You'll be notified that the join has succeeded. Click on **View my virtualized data** to go back and see all your virtualized data.



My virtualized data / Join virtual objects / Edit column names / Assign and review

Assign and review

□ Submit to catalog  Create view

Assign to

○ Data request ○ Project ● My virtualized data

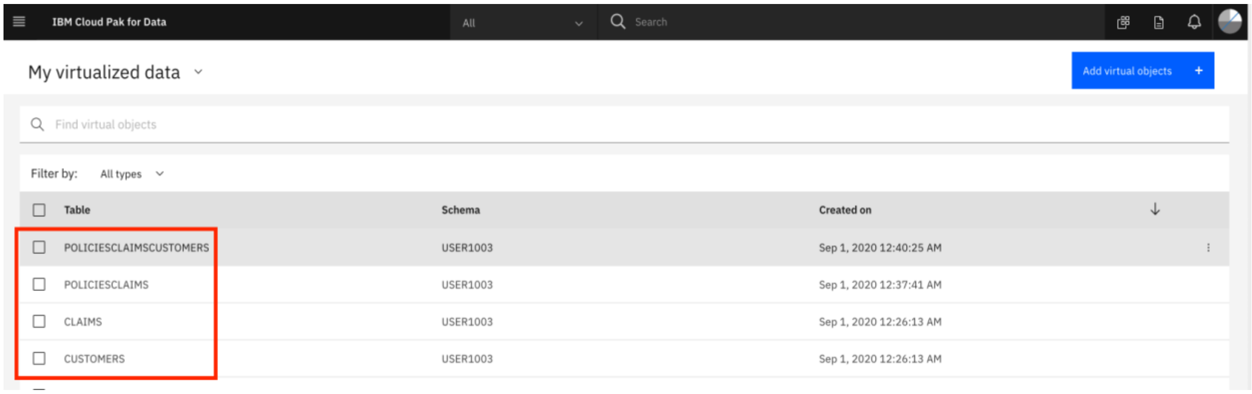
Name

POLICIESCLAIMS

Join view created

| View | Schema |
|---|---|
| POLICIESCLAIMS | USER1003 |

6. **IMPORTANT** Now join the new joined view (POLICIESCLAIMS) and the last virtualized table (CUSTOMERS) to create a new joined view that has all three tables; let's call it POLICIESCLAIMSCUSTOMERS. Switching back to the **My virtualized data** screen should show all three virtualized tables and two joined tables. Do not go to the next section until this step is performed.



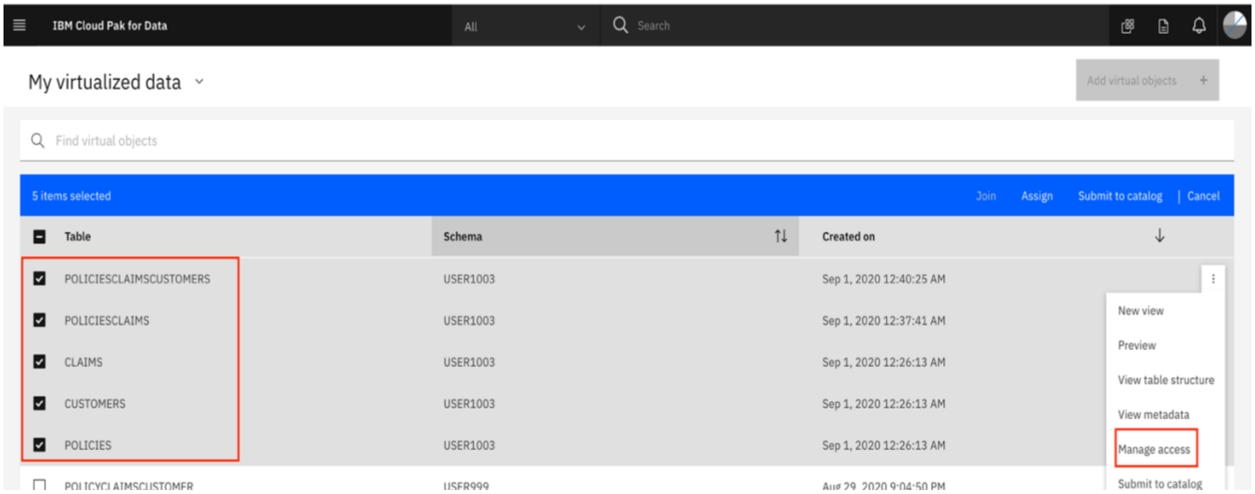
Filter by: All types

| □ Table | Schema | Created on |
|---|---|---|
| □ POLICIESCLAIMSCUSTOMERS | USER1003 | Sep 1, 2020 12:40:25 AM |
| □ POLICIESCLAIMS | USER1003 | Sep 1, 2020 12:37:41 AM |
| □ CLAIMS | USER1003 | Sep 1, 2020 12:26:13 AM |
| □ CUSTOMERS | USER1003 | Sep 1, 2020 12:26:13 AM |

## Grant access to the virtualized data

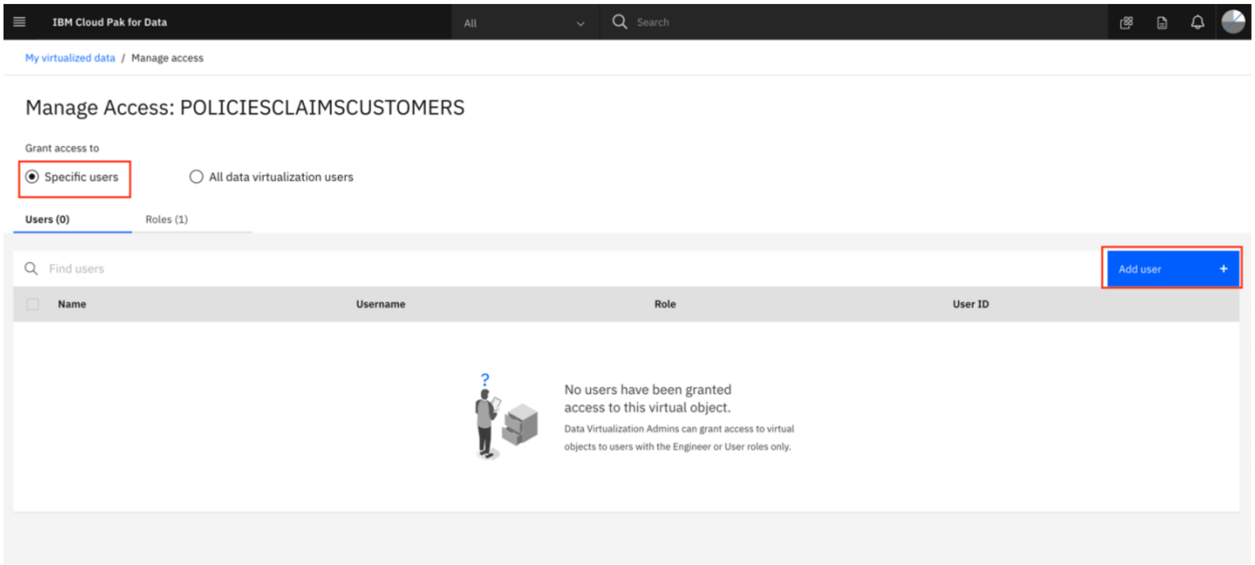
For other users to have access to the data you just virtualized, you need to grant it. Follow these steps to make your virtualized data visible to them:

1. Go to **Data Virtualization** from the hamburger (☰) menu. Click on **Menu** → **My virtualized data**.
2. Click on the virtualized data you've created, then click the three vertical dots to the right, and choose **Manage access.**



| | Table | Schema | Created on |
|---|---|---|---|
| ☑ | POLICIESCLAIMSCUSTOMERS | USER1003 | Sep 1, 2020 12:40:25 AM |
| ☑ | POLICIESCLAIMS | USER1003 | Sep 1, 2020 12:37:41 AM |
| ☑ | CLAIMS | USER1003 | Sep 1, 2020 12:26:13 AM |
| ☑ | CUSTOMERS | USER1003 | Sep 1, 2020 12:26:13 AM |
| ☑ | POLICIES | USER1003 | Sep 1, 2020 12:26:13 AM |
| □ | POLICYCLAIMSCUSTOMER | USER999 | Aug 29, 2020 9:04:50 PM |

3. Click the **Specific users** radio button, then **Add user +**.



Manage Access: POLICIESCLAIMSCUSTOMERS

Grant access to

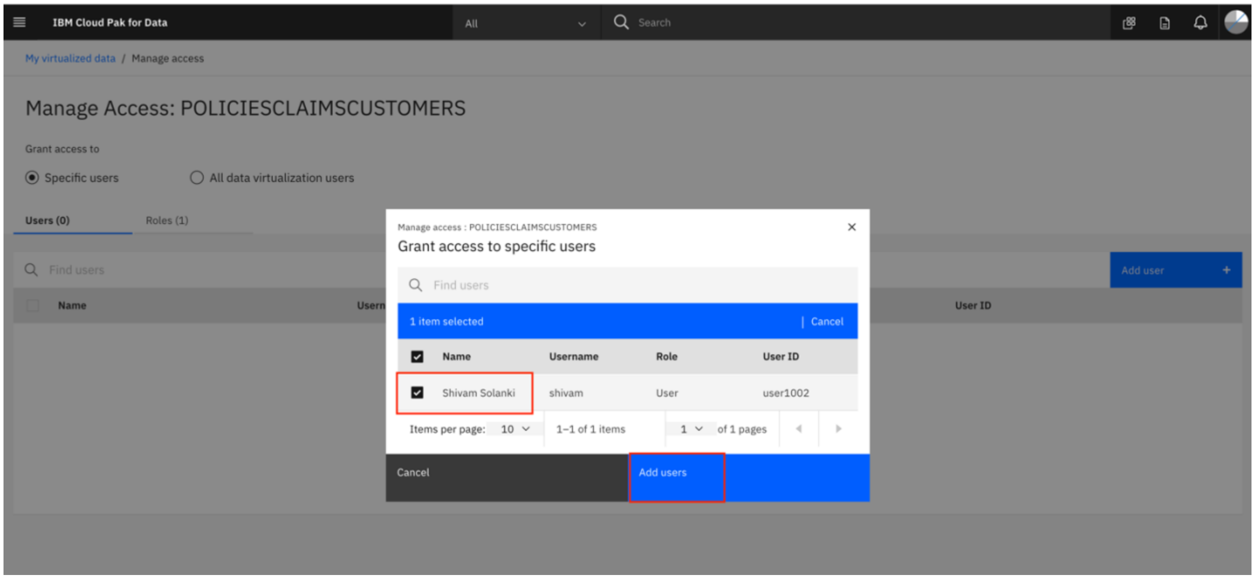
● Specific users ○ All data virtualization users

| □ | Name | Username | Role | User ID |
|---|---|---|---|---|

No users have been granted access to this virtual object.

Data Virtualization Admins can grant access to virtual objects to users with the Engineer or User roles only.

4. Select the users you wish to grant access to and click **Add users**.



Manage Access: POLICIESCLAIMSCUSTOMERS

● Specific users ○ All data virtualization users

Grant access to specific users

| ☑ | Name | Username | Role | User ID |
|---|---|---|---|---|
| ☑ | Shivam Solanki | shivam | User | user1002 |

Items per page: 10 1–1 of 1 items 1 of 1 pages

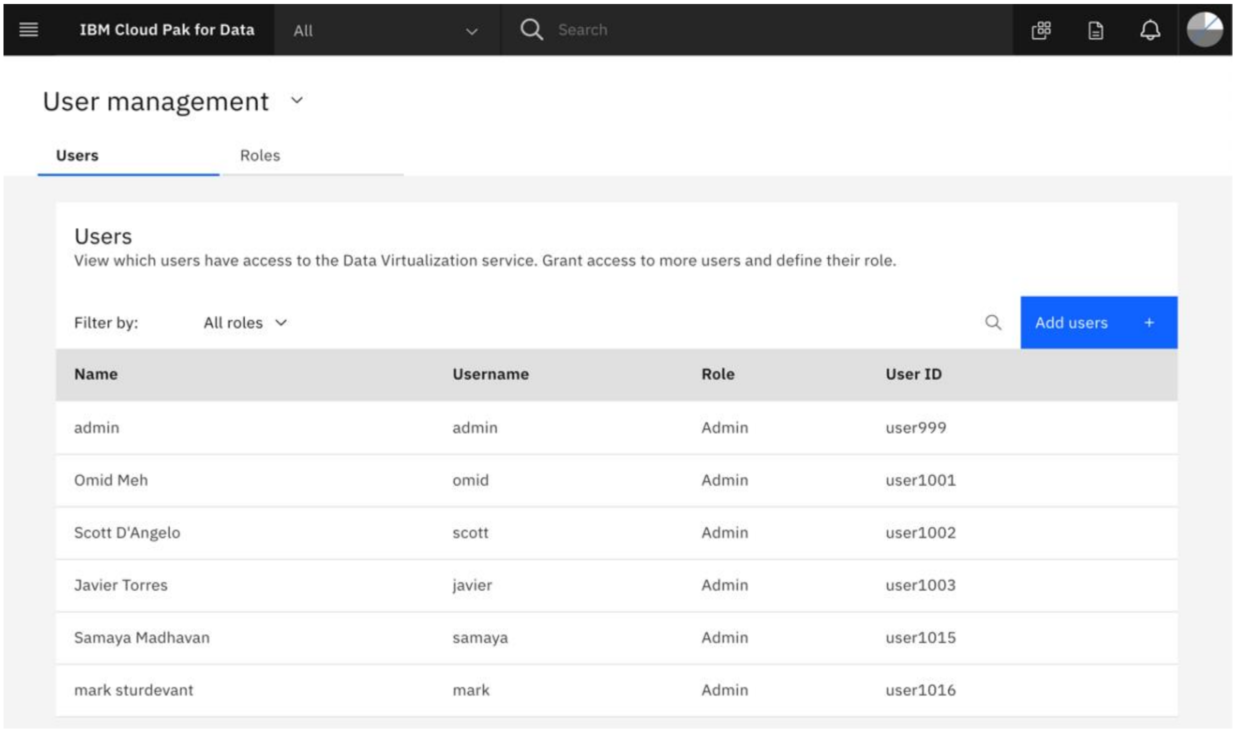
Repeat the above steps for the remaining tables and views.

**Assign the Engineer role to the users**

IBM Cloud Pak for Data users that need to use data virtualization functions must be assigned specific roles based on their job descriptions. These roles are Admin, Engineer, User, and Steward. You can learn more about these roles on the IBM Cloud Pak for Data product hub.

Let’s assign the Engineer role to some users:

1. From the hamburger (☰) menu, choose the **Data Virtualization** option, then click **My virtualized data** → **User management**.

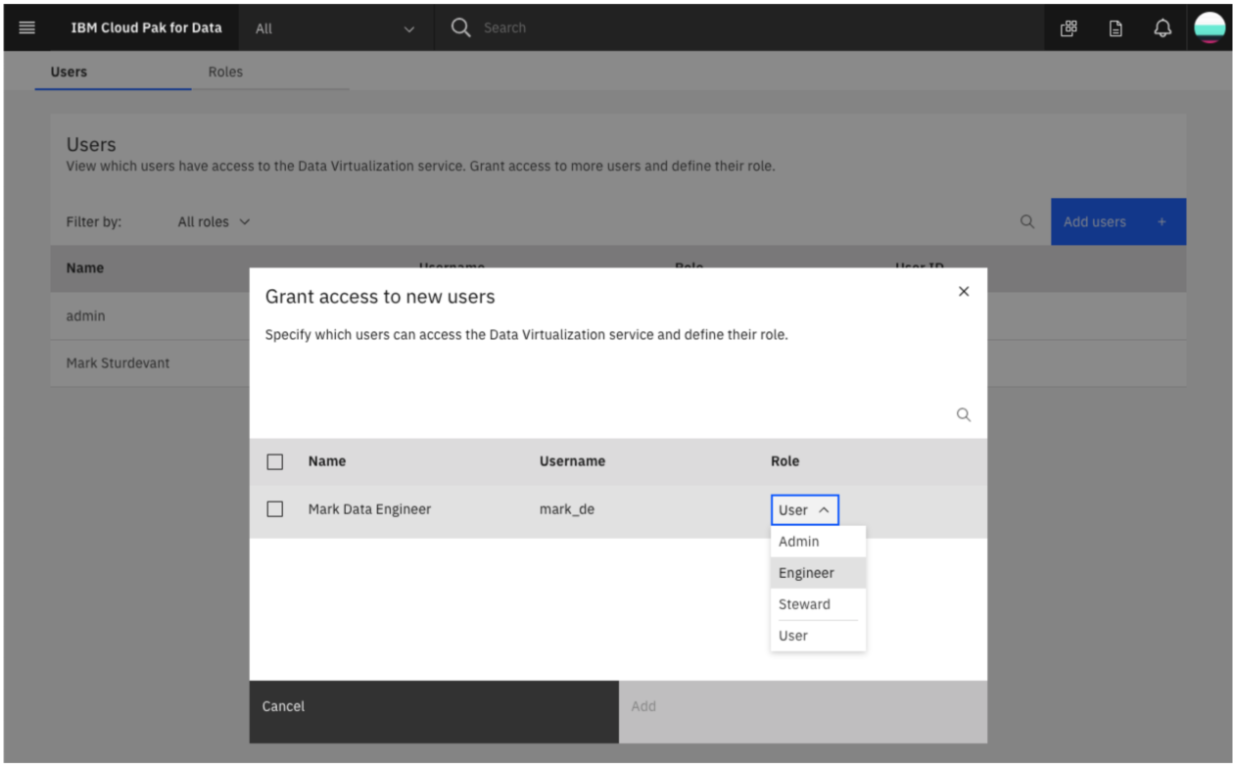


Users

View which users have access to the Data Virtualization service. Grant access to more users and define their role.

| Name | Username | Role | User ID |
|---|---|---|---|
| admin | admin | Admin | user999 |
| Omid Meh | omid | Admin | user1001 |
| Scott D'Angelo | scott | Admin | user1002 |
| Javier Torres | javier | Admin | user1003 |
| Samaya Madhavan | samaya | Admin | user1015 |
| mark sturdevant | mark | Admin | user1016 |

2. Click on **Add users +** and update the role of the user to Engineer.



Users

View which users have access to the Data Virtualization service. Grant access to more users and define their role.

Grant access to new users

Specify which users can access the Data Virtualization service and define their role.

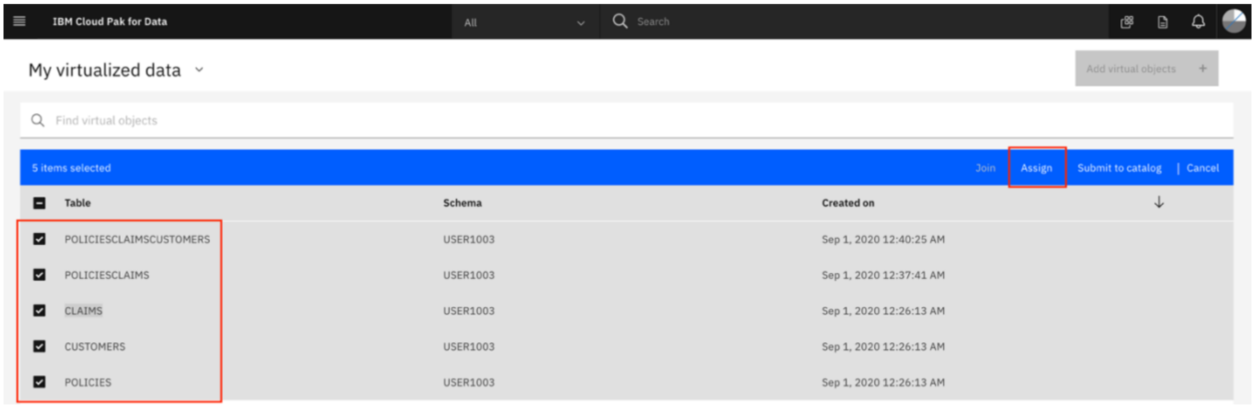
| □ | Name | Username | Role |
|---|---|---|---|
| □ | Mark Data Engineer | mark_de | User |

## Step 6. Users assign virtualized data

Now let's look at how a user who has access to virtualized data can assign the data to their project — how to add the virtualized data as asset to a project.

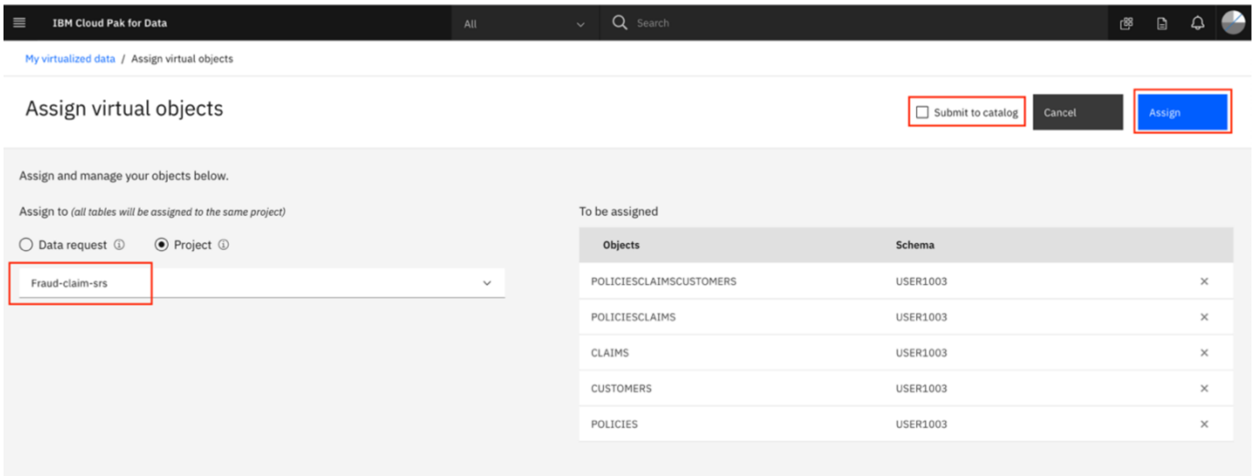
### Assign the data to your project

1. From the hamburger (☰) menu, click on **Collect** → **Data Virtualization**. You will be brought to the My virtualized data section. Here you should see the data you can access (or that the administrator has assigned to you). Select the checkbox next to our original tables (CLAIMS, POLICIES, CUSTOMERS) and the joined tables (POLICIESCLAIMS, POLICIESCLAIMS CUSTOMERS), and click the **Assign** button to import them into your project.



| Table | Schema | Created on |
|---|---|---|
| ☑ POLICIESCLAIMSCUSTOMERS | USER1003 | Sep 1, 2020 12:40:25 AM |
| ☑ POLICIESCLAIMS | USER1003 | Sep 1, 2020 12:37:41 AM |
| ☑ CLAIMS | USER1003 | Sep 1, 2020 12:26:13 AM |
| ☑ CUSTOMERS | USER1003 | Sep 1, 2020 12:26:13 AM |
| ☑ POLICIES | USER1003 | Sep 1, 2020 12:26:13 AM |

2. On the Assign virtual objects screen, choose the project to assign the data. If there is a **Submit to catalog** checkbox on the top right, uncheck it and click the **Assign** button to add the data to your project.



Assign and manage your objects below.

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

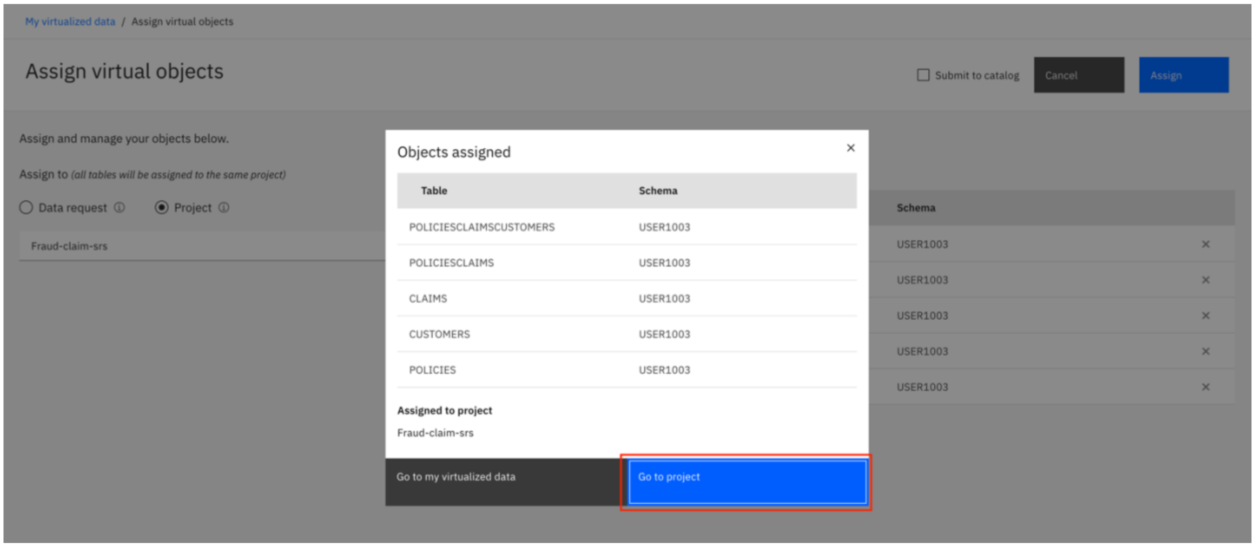
○ Data request ● Project

Fraud-claim-srs

To be assigned

| Objects | Schema |
|---|---|
| POLICIESCLAIMSCUSTOMERS | USER1003 |
| POLICIESCLAIMS | USER1003 |
| CLAIMS | USER1003 |
| CUSTOMERS | USER1003 |
| POLICIES | USER1003 |

3. In the pop-up panel, you will receive a confirmation that the objects have been assigned to your project. Click the **Go to project** button.


My virtualized data / Assign virtual objects
Assign virtual objects
□ Submit to catalog
Cancel
Assign
Assign and manage your objects below.
Assign to (all tables will be assigned to the same project)
○ Data request ● Project
Fraud-claim-srs
Schema
Objects assigned
×
Go to my virtualized data
Go to project

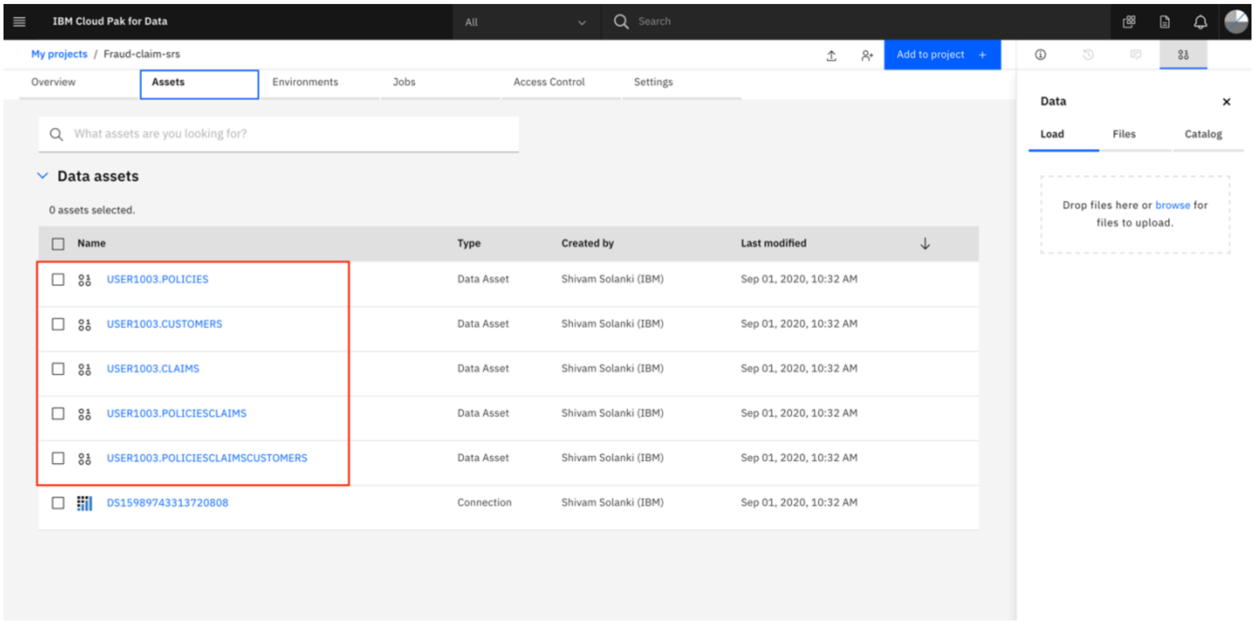
| Table | Schema |
|---|---|
| POLICIESCLAIMSCUSTOMERS | USER1003 |
| POLICIESCLAIMS | USER1003 |
| CLAIMS | USER1003 |
| CUSTOMERS | USER1003 |
| POLICIES | USER1003 |

**Assigned to project**

Fraud-claim-srs

Alternatively, close the modal and go to your projects by clicking on the hamburger (☰) menu, then choosing **Projects**.

4. On the project page, click on the **Assets** tab to display the virtualized tables and joined tables that are now in your project.



Data assets

0 assets selected.

| Name | Type | Created by | Last modified |
|---|---|---|---|
| USER1003.POLICIES | Data Asset | Shivam Solanki (IBM) | Sep 01, 2020, 10:32 AM |
| USER1003.CUSTOMERS | Data Asset | Shivam Solanki (IBM) | Sep 01, 2020, 10:32 AM |
| USER1003.CLAIMS | Data Asset | Shivam Solanki (IBM) | Sep 01, 2020, 10:32 AM |
| USER1003.POLICIESCLAIMS | Data Asset | Shivam Solanki (IBM) | Sep 01, 2020, 10:32 AM |
| USER1003.POLICIESCLAIMSCUSTOMERS | Data Asset | Shivam Solanki (IBM) | Sep 01, 2020, 10:32 AM |
| DS15989743313720808 | Connection | Shivam Solanki (IBM) | Sep 01, 2020, 10:32 AM |

## Summary

This lab tutorial explained how to virtualize Db2 Warehouse data with data virtualization on IBM Cloud Pak for Data to make queries across multiple data sources. To continue the series and learn more about IBM Cloud Pak for Data, take a look at the next tutorial, Data visualization with data refinery.