

# Lab Guide

## EAD25T340BH: Hands-on-Lab Watson Knowledge Catalog powered by Cloud Pak for Data

Rick Buglio

Digital Technical Engagement Manager – DataOps

[rbuglio@us.ibm.com](mailto:rbuglio@us.ibm.com)



# Fuel Data Innovation and AI with IBM® Watson Knowledge Catalog

Explore the self-service capabilities of IBM® Watson Knowledge Catalog powered by Cloud Pak for Data. Spend less time searching and preparing data and more time putting it to use doing analytics to extract business value

## Tutorial

IBM Watson Knowledge Catalog powers intelligent, self-service discovery of data, models and more, activating them for artificial intelligence, machine learning and deep learning. Access, curate, categorize and share data, knowledge assets and their relationships wherever they reside.

In this tutorial, you will explore the following key capabilities:

- Creating a Governed Knowledge Catalog
- Discovering and Cataloging Data Assets
- Understanding and Socializing Data Assets
- Shopping for Data
- Preparing Data for Analytics and AI
- Protecting Sensitive Information

## Introduction

In the insurance industry, claims processing is an area with many inefficiencies and risks to the insurance provider. A significant amount of the risk involved lies in the amount of time it takes to process a claim. The more time required to make the required adjustments, the higher the risk of lawsuits, which are a costly outcome.

Processing insurance claims is an expensive, time consuming and risk-intensive process. Challenges around claims processing become especially intense during natural calamities, when insurers need to process a sudden spike in claims, even to the point of transporting adjusters to the impacted location.

With a data-driven approach, the information gathering process can be expedited tremendously with immediate access to relevant information at the first notice of loss. The use of data analytics and AI can help identify potential claim fraud using deep learning and detailed data analysis.

Using information that's available in the insurance company's enterprise Knowledge Catalog, the business can easily develop a data-driven claims process that:

- Reduces the median time for a claim to be processed.
- Minimizes the risk of fraud.
- Automates as much of the claims and adjustment process as possible, while triaging more complex claims for adjusters to process.

In this use case, the insurance company's goal is to create a dashboard for a claims agent to interact with the information pushed up to the insurance company from the customer's mobile app. To help mitigate the fraud potential of remotely adjusting auto insurance claims, the customer is prompted for an image of their vehicle to validate that they are indeed a legitimate customer. Using IBM Cloud Pak for Data, the business can easily prepare a visual recognition model to assess the make, model, and year of the customer's vehicle in the photo. The app also needs the customer's account, logistics, policy and claims information to validate that the vehicle in the image matches the vehicle under the policies coverage.

This tutorial introduces you to the intelligent and collaborative capabilities of the IBM Watson Knowledge Catalog, and the integrated, common fabric of IBM Cloud Pak for Data. These offerings empower the insurance company's business analysts, data scientists and data professionals to quickly and easily discover, curate, catalog, shape and share data assets in preparation for the analytics and AI processes that will help them achieve their business goals.

## Prerequisites

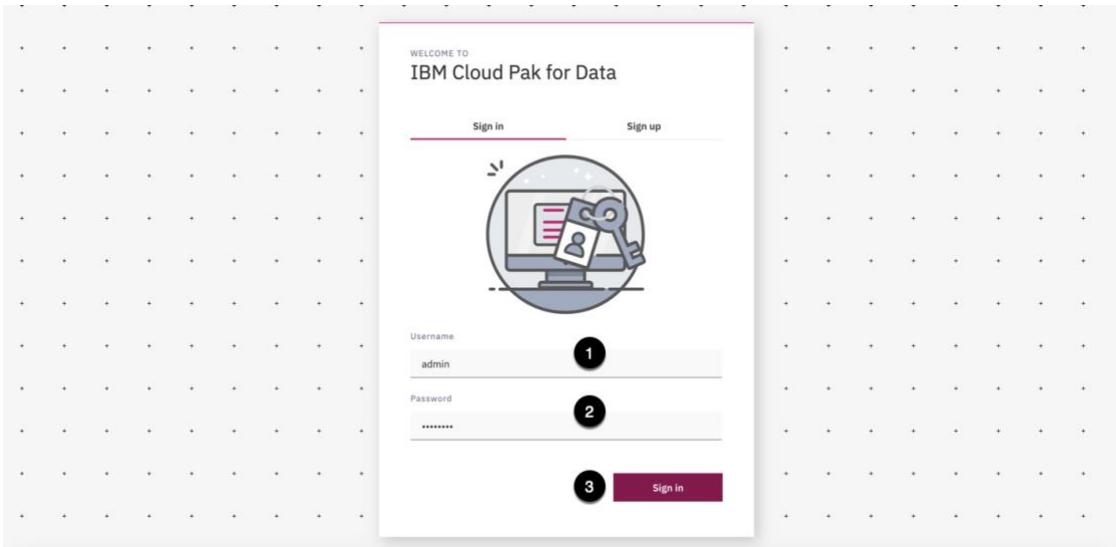
### Download Unstructured Files

In the **Discover and Catalog Data Assets** task, you are instructed to add two files to a new Knowledge Catalog using the **Local files** method. You need to download the files, to your desktop or local file system, from this [Tutorial Files](#) Box folder and remember where you placed them. Do this **now** before you proceed.

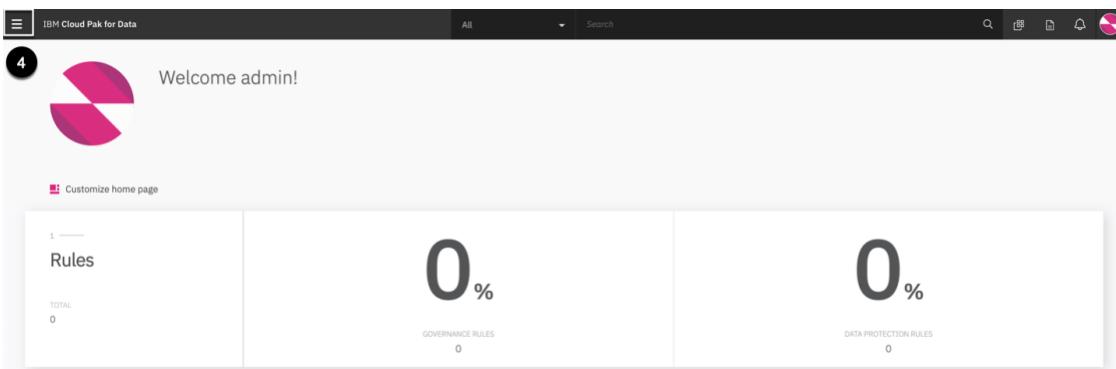
### Create an IBM Cloud Pak for Data User

In this section you will log in to Cloud Pak for Data using the **admin** credentials and create a **new** user and assign the user to **all** the available roles. This will provide you with the authority you need to complete the lab and an isolated account to only view what you create. This will shield you from work being done by other users doing the lab in the same environment.

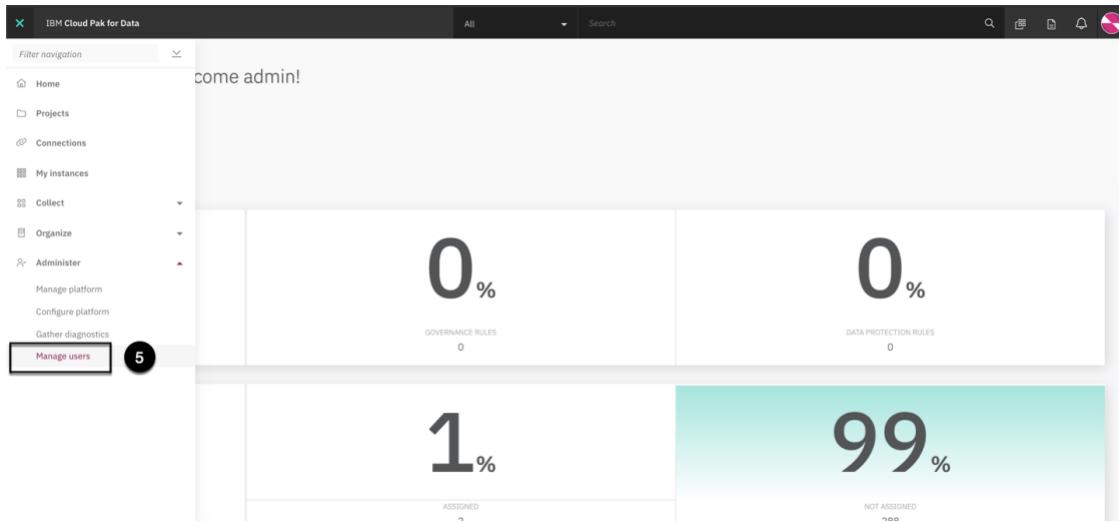
Go to the [Fast Start 2020 Las Vegas Skytap Labs](#) website. Wait for the instructor to provide you with a Cloud Pak for Data cluster access code.



1. Enter **admin** as the **Username**.
2. Enter **password** as the **Password**.
3. Click the **Sign in** button.



4. Click the **IBM Cloud Pak for Data** navigation menu in the top left corner.



## 5. Click the Administer > Manage users menu.

The screenshot shows the 'Manage users' page under the 'Administer' section. It has tabs for 'Users' and 'Roles'. There's a search bar and a 'New user' button with a circled number 6. A table lists three existing users: admin, ctp, and data citizen, each with their status, username, date added, user ID, and roles assigned.

## 6. Click the + New user button.

The screenshot shows the 'New user' creation form. It includes fields for User (7), Username (8), Password (9) and Re-enter new password (10), Email (11), and Roles (12). A list of available roles is shown on the right: Administrator, Business Analyst, Data Engineer, Data Quality Analyst, Data Scientist, Data Steward, and Developer. A circled number 13 points to the 'Save' button at the bottom right.

## 7. Enter your Full Name as the User.

## 8. Enter your Last Name as the Username.

9. Enter your First Name as the **Password**.
10. Enter your First Name again in the **Re-enter new password** field.
11. Enter your Email address as the **Email**.
12. Select all of the **Roles** listed.
13. Click the **Save** button.

The screenshot shows the 'Manage users' page. At the top, there are tabs for 'Users' and 'Roles'. Below is a search bar with a placeholder 'Find users' and a magnifying glass icon. To the right of the search bar is a '+ New user' button. The main area displays a table with columns: Name, Status, Username, Date added, User ID, and Roles. Two users are listed: 'Rick Buglio' (Status: ✓, Username: buglio, Date added: 12 Feb, 2020 3:24 PM, User ID: 1000331004, Roles: Administrator, Business Analyst + 5 more) and 'admin' (Status: ✓, Username: admin, Date added: --, User ID: 1000330999, Roles: Administrator, Data Scientist + 5 more). In the top right corner, there is a circular profile icon with a pink and white design, and next to it is a 'Configure LDAP' link. A callout bubble with the number '14' is positioned above the profile icon.

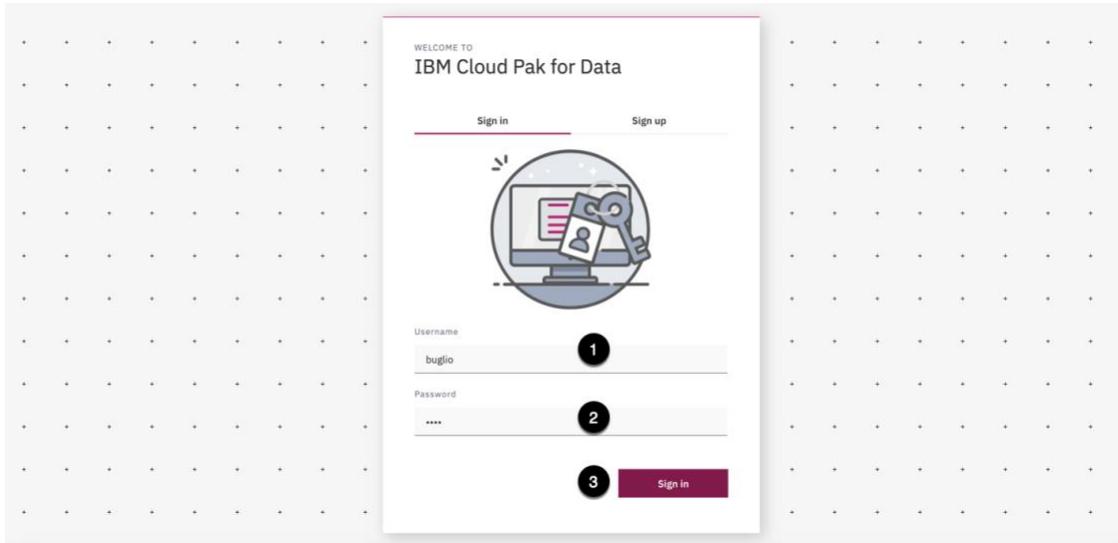
14. Click on the **Profile and settings** icon in the top right corner.

The screenshot shows the same 'Manage users' page after step 14. A vertical user dropdown menu has appeared on the right side of the screen, showing the name 'admin' and the 'Profile and settings' option. Below this, there are links for 'Getting Started', 'About', 'Community', and 'Support'. At the bottom of the dropdown is a 'Log out' button with a right-pointing arrow. A callout bubble with the number '15' is positioned above the 'Log out' button.

15. Click **Log out**.

## Log in to Cloud Pak for Data

In this section you will log into Cloud Pak for Data using the credentials of the **new** user you just created in the previous step to do the lab.



1. Enter your Last Name as the **Username**.
2. Enter your First Name as the **Password**.
3. Click the **Sign in** button.

A screenshot of the IBM Cloud Pak for Data welcome page. The top navigation bar includes a menu icon, the title "IBM Cloud Pak for Data", a search bar, and a user profile icon. The main area features a large "WELCOME, ctp!" message with a pie chart icon. Below it is a "Let's get started!" callout. A sidebar on the left lists resources: "IBM Cloud Pak for Data: Collect and organize", "IBM Cloud Pak for Data: Analyze", and "IBM Cloud Pak for Data: Accelerate". The "Collect and organize" section includes a video thumbnail titled "IBM Cloud Pak for Data: Collect and organize" and a list of tasks: "Set up data connections", "Create catalogs", "Explore business terms", "Explore policies", "Discover assets", "Explore catalogs", "Explore information assets", "View data quality assets", "Transform and integrate data", and "Create a database".

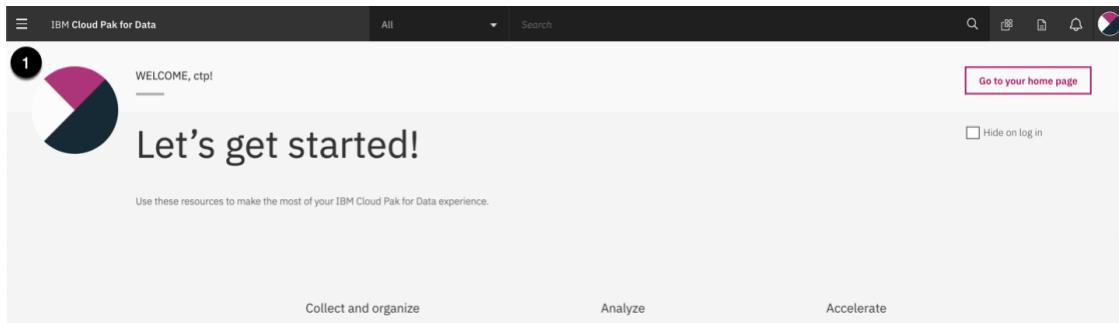
You will be brought into the IBM Cloud Pak for Data welcome page.

## Create a Governed Catalog

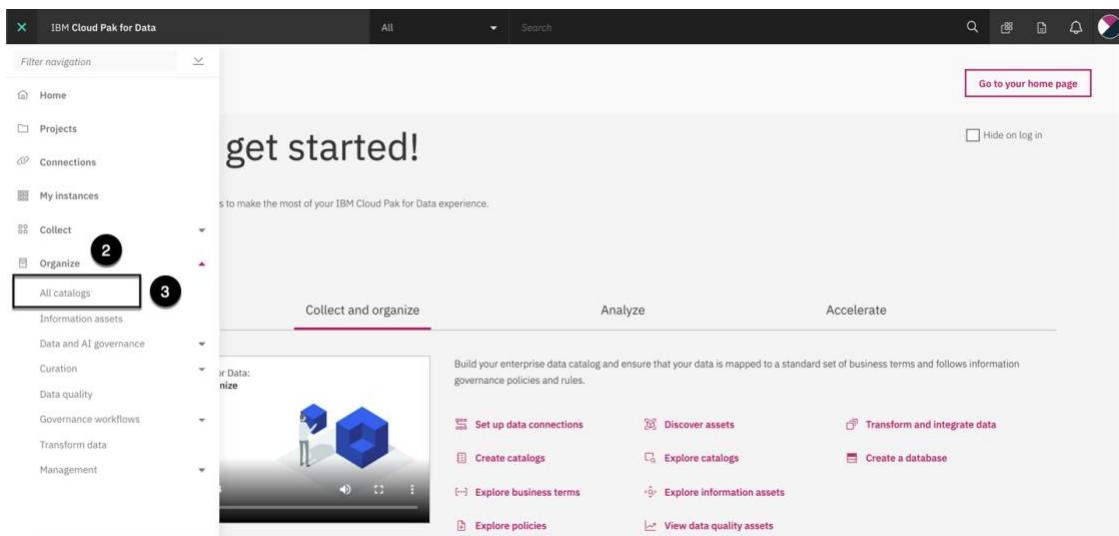
In this task you create a governed **Knowledge Catalog**. Watson Knowledge Catalog is a secure and collaborative catalog of metadata used to organize and govern information assets. It is tightly integrated with the global business glossary of data governance artifacts that describe and govern the information managed by the catalog, providing self-service capabilities for data professionals to quickly and easily search, find, understand and use data.

A **Default Catalog** is provided out of the box. However, organizations can create as many catalogs as they need. In this lab, you will create an additional catalog to house the **Auto Insurance** claims fraud analysis information assets that will be used by the analytic project team.

You will learn how to discover, curate and catalog data assets using an additional catalog other than the **Default Catalog** and by using some alternative methods. It will still have integration to the global business glossary and the business policies and rules to govern and protect it.

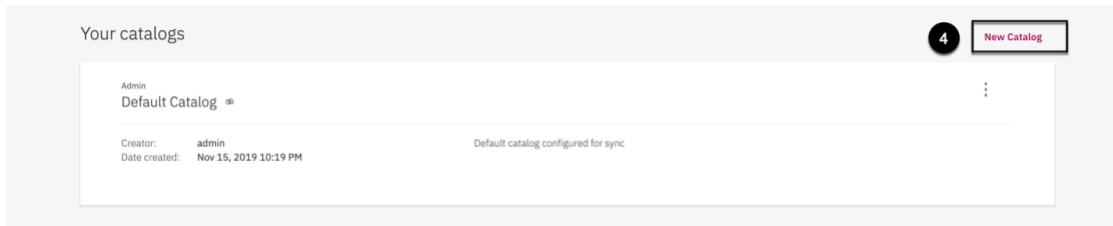


1. Click the **IBM Cloud Pak for Data** navigation menu in the top left corner.

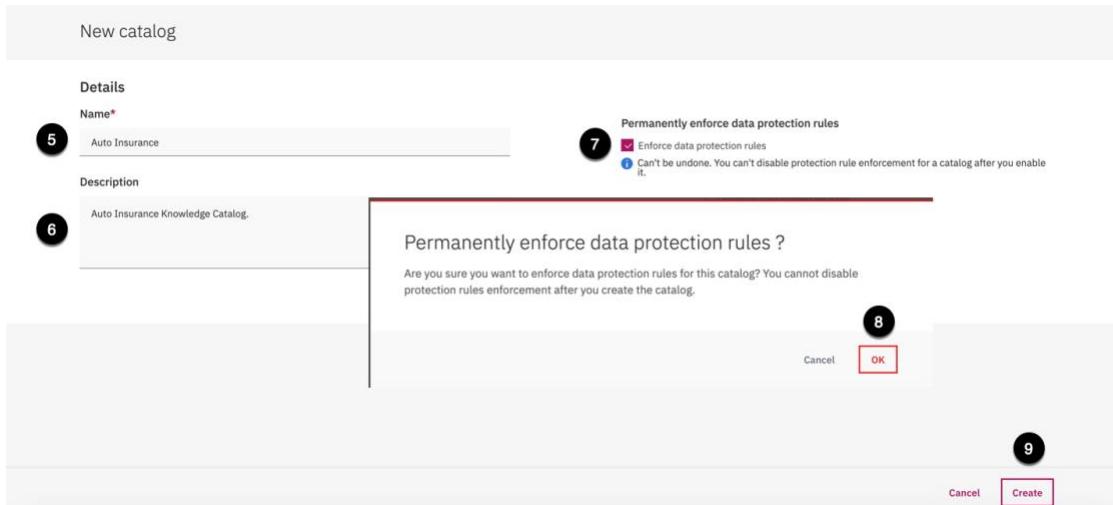


2. Click the **Organize** menu.

3. Click the **All Catalogs** menu item.



4. Click the **New Catalog** button in the top right corner.



5. Enter a Name of **Auto Insurance**.

6. Enter a Description of **Auto Insurance Knowledge Catalog**.

7. Select the **Enforce data policies** checkbox.

The **Permanently enforce data protection rules** warning dialog will be displayed, asking if you are sure you want to set this option and informing you that the setting is permanent.

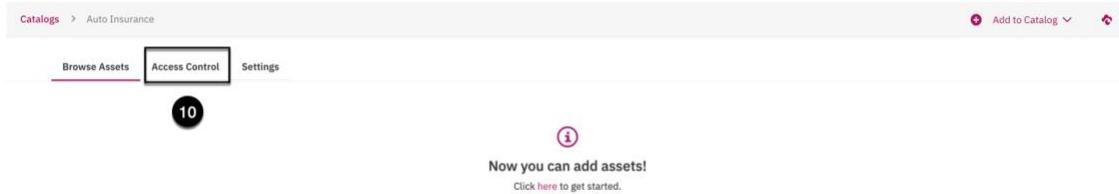
8. Click the **OK** button.

By default, access to data assets in a catalog is only restricted by the privacy settings of the data assets. Privacy settings and policy rules can limit which members of the catalog can view and use the assets. You can implement data protection rules to restrict access to data based on the contents of the data. These rules help you control data access and ensure that the right people can access the right data. Selecting the option to **Enforce data protection rules** enables the enforcement of data protection rules to allow or deny access to a data asset or mask, substitute and redact data at the data asset field level.

Setting this option for a catalog is a good best practice. Once it is enabled, it cannot be undone, but it does not restrict or impede any functionality, it provides additional security measures to protect data assets.

9. Click the **Create** button.

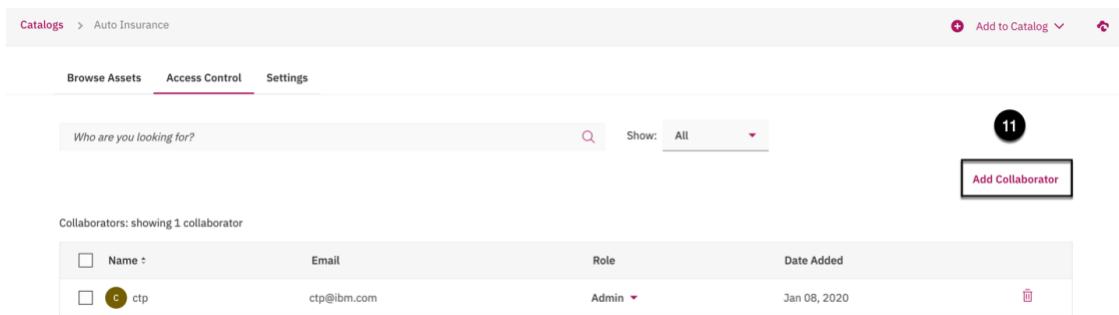
You will see a **Creating Auto Insurance** notification during catalog creation.



A screenshot of the Access Control tab in a catalog. At the top, there are tabs for 'Browse Assets' (highlighted with a red box), 'Access Control' (also highlighted with a red box), and 'Settings'. A circular badge with the number '10' is positioned above the tabs. Below the tabs, a message box displays the text 'Now you can add assets!' with an information icon, and a link 'Click here to get started.' A red box highlights the 'Access Control' tab.

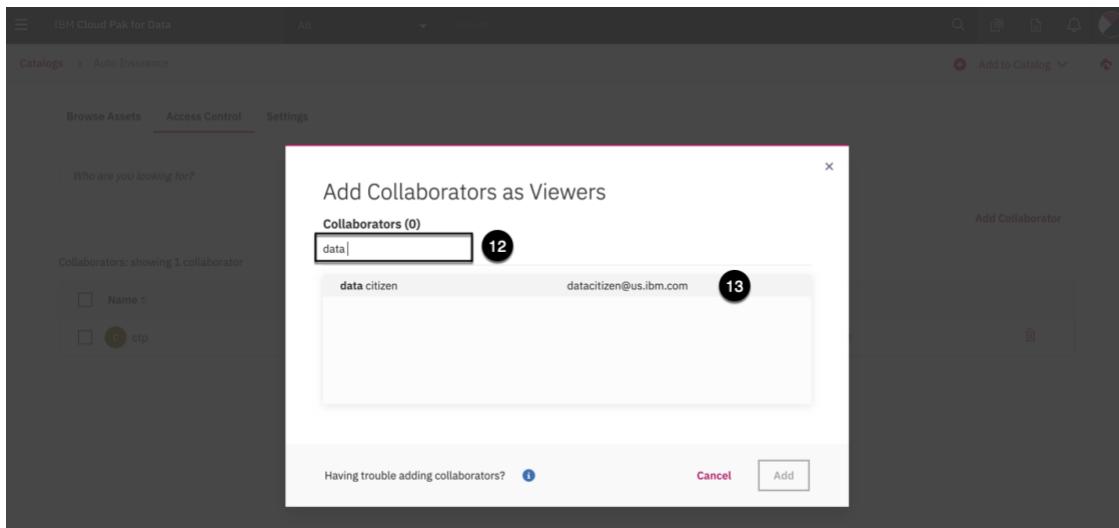
Once completed, you are brought into the newly created catalog. You will now add the **data citizen** user to the catalog as a *Viewer* so they can access the new catalog and use the data assets. You will log in as this user at the end of the lab to see how data protection rules are enforced.

#### 10. Click the **Access Control** tab.



A screenshot of the Access Control tab in a catalog. The 'Access Control' tab is highlighted with a red box. Below it, there is a search bar labeled 'Who are you looking for?' and a dropdown menu set to 'All'. A circular badge with the number '11' is positioned above the search bar. On the right side, a button labeled 'Add Collaborator' is highlighted with a red box. The main area shows a table of collaborators with one entry: 'ctp' (Email: ctp@ibm.com, Role: Admin, Date Added: Jan 08, 2020). A red box highlights the 'Add Collaborator' button.

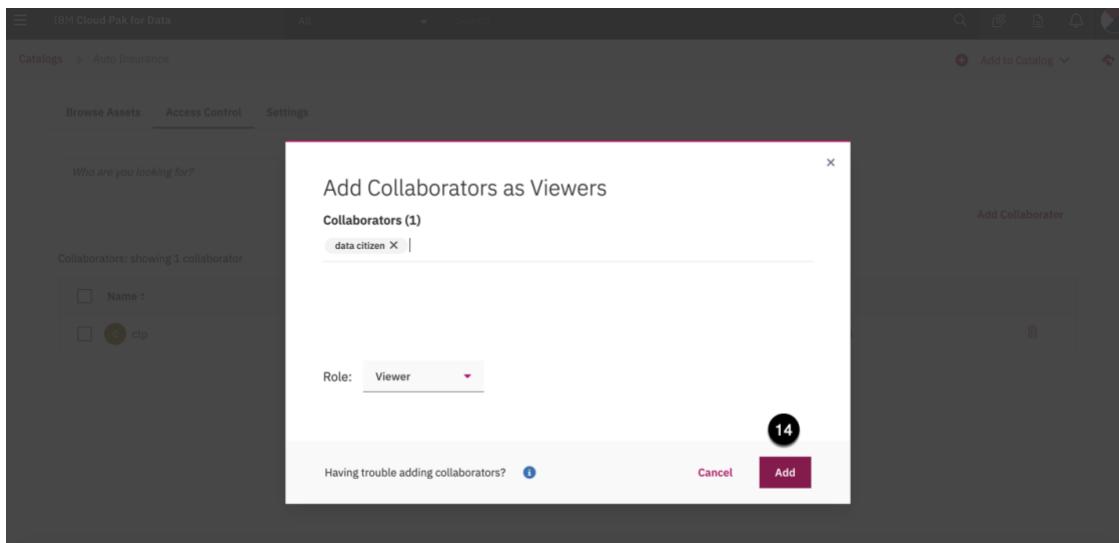
#### 11. Click the **Add Collaborator** button.



A screenshot of the 'Add Collaborators as Viewers' dialog box. The title is 'Add Collaborators as Viewers'. It shows a search input field containing 'data' (highlighted with a red box) and a list of users. One user, 'data citizen' (Email: datacitizen@us.ibm.com), is selected and highlighted with a red box. A circular badge with the number '12' is positioned above the search input, and another with '13' is positioned above the selected user. At the bottom, there are 'Cancel' and 'Add' buttons. A red box highlights the 'Add' button.

#### 12. Type the word **data** in the search area.

#### 13. Click on the **data citizen** user. The default role of Viewer is automatically assigned. Leave the role set to Viewer.



#### 14. Click Add.

Name	Email	Role	Date Added
ctp	ctp@ibm.com	Admin	Jan 08, 2020
data citizen	datacitizen@us.ibm.com	Viewer	Jan 11, 2020

You should now see the **data citizen** user added as a **Viewer**.

## Create Analytic Projects

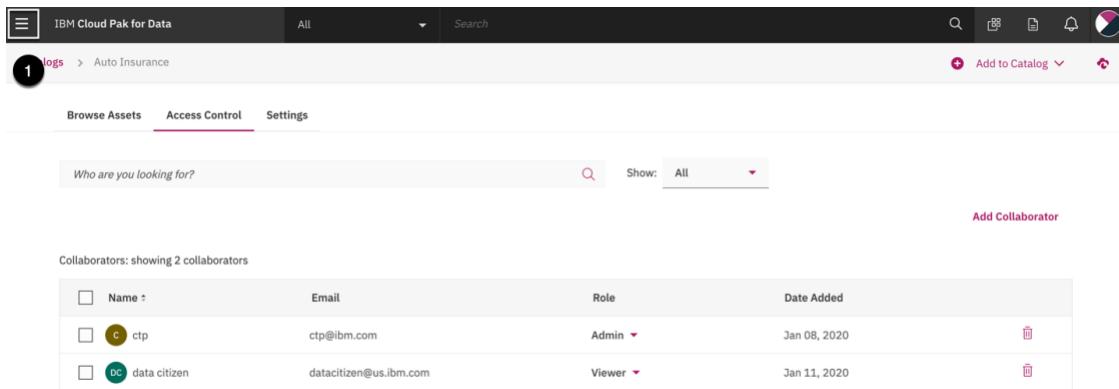
In this task you will create two Cloud Pak for Data analytics projects.

The first project, which will be named **Auto Insurance**, will be used by the auto insurance analytics team, to collaborate and build the analytic and AI assets; notebooks, models, data flows, dashboards etc. to analyze the auto insurance claims process. You will add auto insurance data assets from the **Auto Insurance** knowledge catalog to this project and do some shaping of the data using the data refinery to prepare the data for analytical insights.

The 2nd project, which will be named **Auto Discovery**, will be used to demonstrate the auto discovery capabilities of Watson Knowledge Catalog. When you catalog a **Connection**, you can choose the option to automatically discover data assets. The discovered assets are added to a Cloud Pak for Data analytics project as a temporary holding area for review. You will use a separate project for auto discovery, so it does not disrupt the data analytics project. Once data

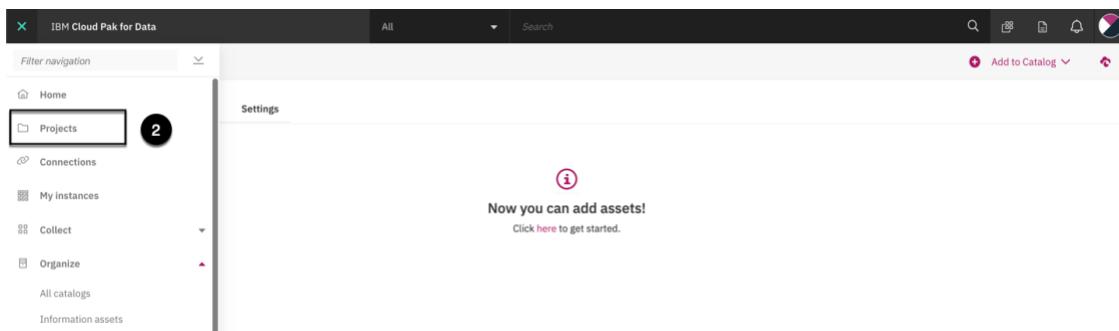
assets are discovered and added to a project, you can review them, determine which assets are relevant and then publish them to a catalog.

## Create the Auto Insurance Project



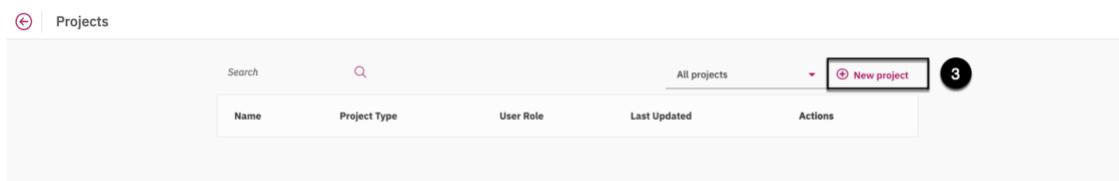
The screenshot shows the IBM Cloud Pak for Data interface. At the top, there's a navigation bar with the title "IBM Cloud Pak for Data". Below it, a breadcrumb trail shows "Logs > Auto Insurance". On the right side of the header are several icons: a magnifying glass for search, a document icon, a bell for notifications, and a profile picture. A pink button labeled "Add to Catalog" is also visible. The main content area has tabs for "Browse Assets", "Access Control" (which is currently selected), and "Settings". A search bar at the top of the content area contains the placeholder "Who are you looking for?". Below the search bar is a dropdown menu set to "All". To the right of the search bar is a pink "Add Collaborator" button. The main content area is titled "Collaborators: showing 2 collaborators". It lists two users: "ctp" (Admin role, added on Jan 08, 2020) and "data citizen" (Viewer role, added on Jan 11, 2020). Each user entry includes a checkbox, their name, email, role, date added, and a trash bin icon for removal.

1. Click the **IBM Cloud Pak for Data** navigation menu in the top left corner.



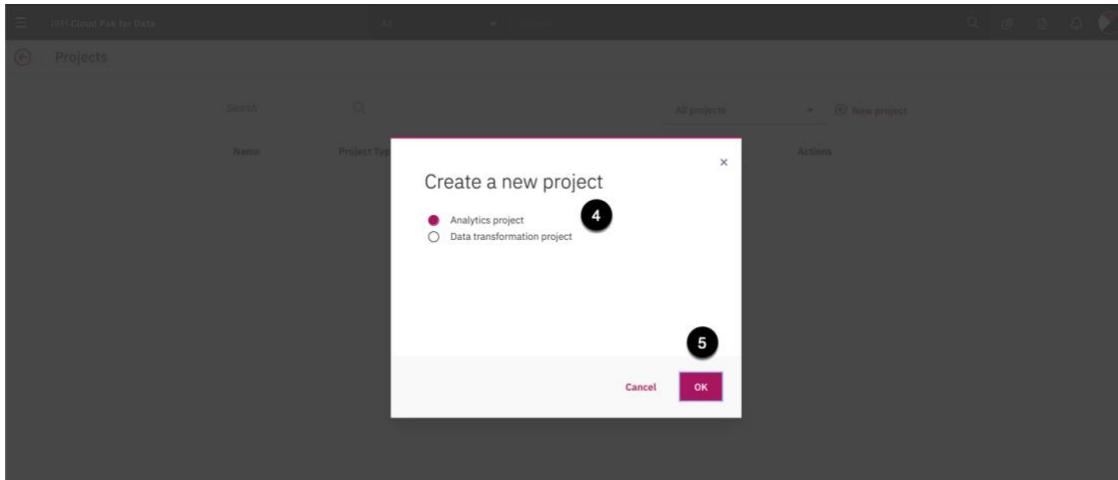
This screenshot shows the "Projects" menu item highlighted with a black box and a circled number "2". The menu also includes options like "Connections", "My instances", "Collect", and "Organize". To the right, there's a "Settings" section with a pink info icon and the text "Now you can add assets! Click here to get started." A pink "Add to Catalog" button is located at the top right of the main content area.

2. Click the **Projects** menu.



This screenshot shows the "Projects" page. A pink "New project" button is highlighted with a black box and a circled number "3". The page includes a search bar, a dropdown menu set to "All projects", and a table with columns for "Name", "Project Type", "User Role", "Last Updated", and "Actions".

3. Click the **New project** button.



4. Click the **Analytics project** radio button (usually selected by default).

5. Click the **OK** button.

Create a 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** 6

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 a Git repository.

**USE TO**

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

6. Click on **Create an empty project**.

New project

Define project details

Name 7

Auto Insurance

Description 8

Auto Insurance claims project

Choose project options

Integrate this project with Git 1

9

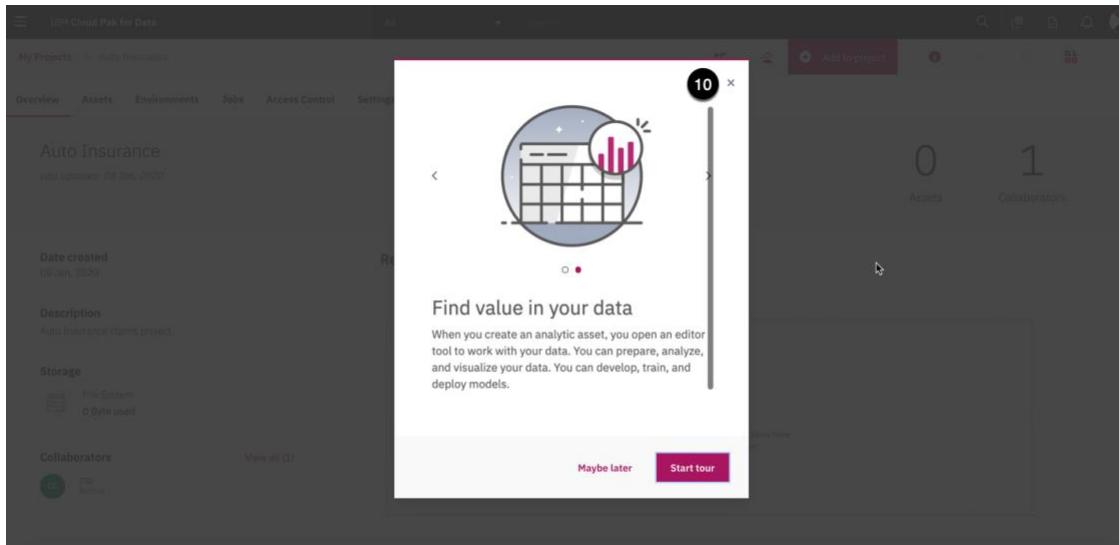
Cancel Create

7. Enter a Name of **Auto Insurance**.

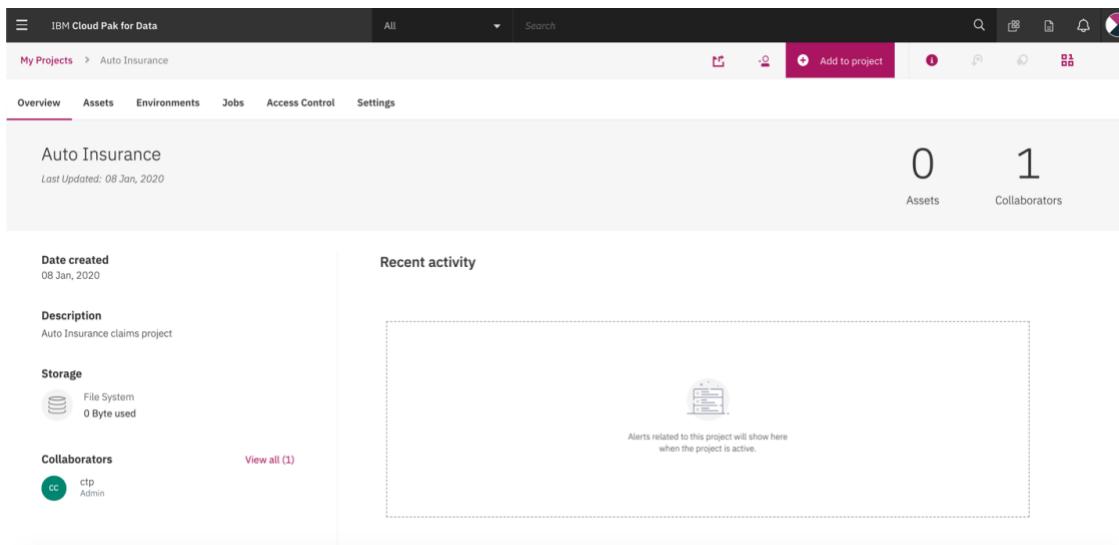
8. Enter a Description of **Auto Insurance claims project**.

9. Click the **Create** button.

The Create button will turn to **Creating...** so be patient and wait for the project to be created.



If the Getting Started tour dialog appears, click on the X in the top right corner to close it.



When the project creation is complete, you are brought into your newly created project and you will see the **Overview** section.

## Create the Auto Discovery 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' and a search bar. Below it, a breadcrumb trail shows '1 Projects > Auto Insurance'. The main content area displays the 'Auto Insurance' project details: 'Last Updated: 08 Jan, 2020'. On the right, there are counts for 'Assets' (0) and 'Collaborators' (1). Below this, there are tabs for 'Overview', 'Assets', 'Environments', 'Jobs', 'Access Control', and 'Settings'. A pink 'Add to project' button is located at the top right of the main content area.

1. Click the **IBM Cloud Pak for Data** navigation menu in the top left corner.

This screenshot shows the same interface as above, but with the 'Projects' menu item in the left sidebar highlighted with a black circle labeled '2'. The sidebar also includes other options like 'Connections', 'My instances', 'Collect', 'Organize', and 'All catalogs'. The main content area remains the same, showing the 'Auto Insurance' project details.

2. Click the **Projects** menu.

This screenshot shows the 'Projects' page. The 'New project' button in the top right is highlighted with a black circle labeled '3'. The table below lists one project: 'Auto Insurance' (Analytics, Admin, Last Updated: 8 Jan 2020, 4:05 PM). The left sidebar shows the 'Projects' menu item selected.

3. Click the **New project** button.

This screenshot shows the 'Create a new project' dialog box. It has a title 'Create a new project' with a close button. There are two radio buttons: 'Analytics project' (selected, indicated by a black circle labeled '4') and 'Data transformation project'. At the bottom are 'Cancel' and 'OK' buttons, with 'OK' being highlighted with a black circle labeled '5'.

4. Click the **Analytics project** radio button (usually selected by default).

**5. Click the **OK** button.**

The screenshot shows the 'Create a project' screen. It has two main sections: 'Create an empty project' (selected) and 'Create a project from a sample or file'. Each section includes an icon, a title, a brief description, and a 'USE TO' list. The 'Create an empty project' section is highlighted with a red circle labeled '6'.

**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** 6  
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 a Git repository.

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

**6. Click on **Create an empty project**.**

The screenshot shows the 'Define project details' step of the project creation wizard. It includes fields for 'Name' (Auto Discovery) and 'Description' (Auto Discovery project), both highlighted with red circles labeled '7' and '8'. At the bottom, there's a checkbox for 'Integrate this project with Git' and a 'Create' button highlighted with a red circle labeled '9'.

New project

Define project details

Name  
Auto Discovery 7

Description  
Auto Discovery project 8

Choose project options  
 Integrate this project with Git 9

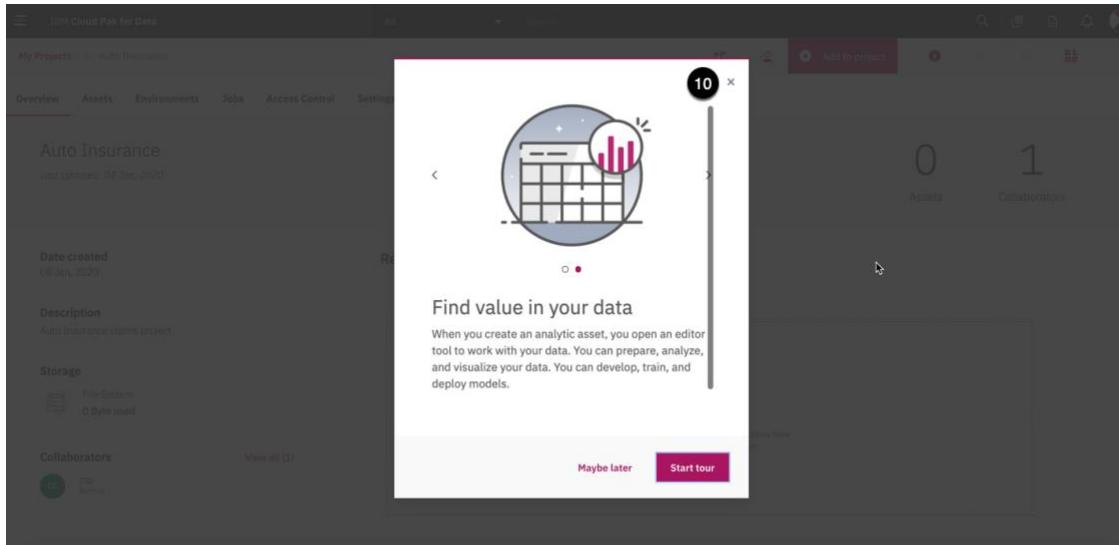
Cancel Create

**7. Enter a Name of **Auto Discovery**.**

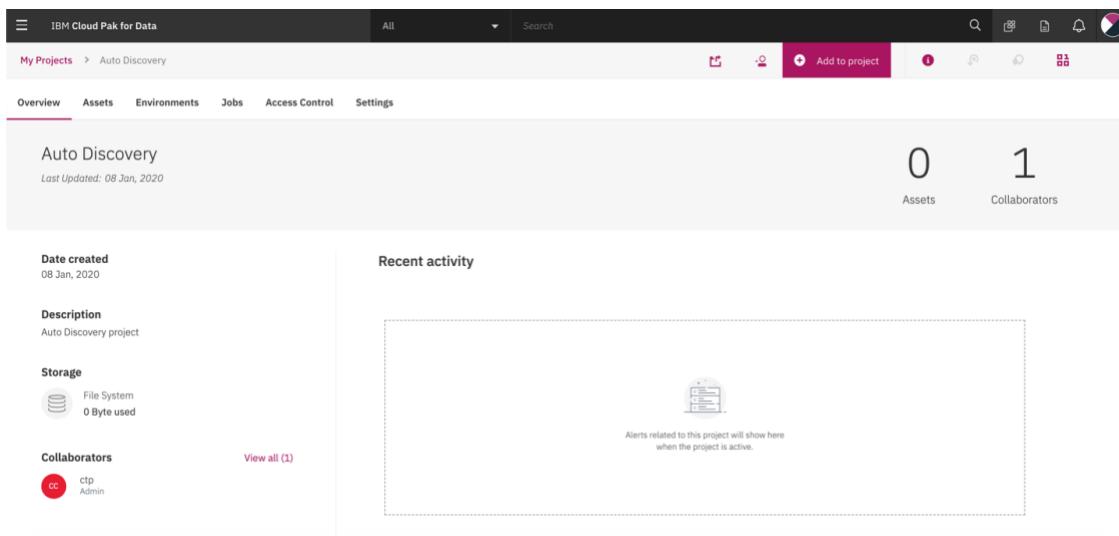
**8. Enter a Description of **Auto Discovery project**.**

**9. Click the **Create** button.**

The Create button will turn to **Creating...** so be patient and wait for the project to be created.



If the Getting Started tour dialog appears, Click on the X in the top right corner to close it.



When the project creation is complete, you are brought into your newly created project and you will see the **Overview** section.

## Discover and Catalog Data Assets

In this task, you will discover and catalog unstructured data assets from the local file system and structured data assets from a **Db2 Warehouse on Cloud** connection that you will create. This will introduce you to the three methods available to discover and catalog data assets; **Local files**, **Connected asset** and **Connection**. You will use these methods to catalog data assets into the newly created Knowledge Catalog and then tag them for users to easily find them, understand their content and make them available throughout IBM Cloud Pak for Data, for use during data preparation and within models, dashboards and notebooks.

**Note** - Some of the screen shots are in blue instead of magenta. They were taken from the Watson Knowledge Catalog on Cloud hands-on lab that walks through the exact same use case. I have not re-captured them in magenta for several reasons:

- 1 - The process and UI are identical in the Cloud and Cloud Pak for Data versions, and there were too many images that would have to be recaptured and replaced.
- 2 - The Cloud Pak for Data UI will be converted to Carbon X in 2020 which will change the UI color scheme to blue from magenta to match the Cloud version of Watson Knowledge Catalog.

You will see this theme throughout the remainder of the lab.

## Catalog Unstructured Data

The screenshot shows the IBM Cloud Pak for Data interface. At the top, there is a navigation bar with a menu icon, the text "IBM Cloud Pak for Data", a search bar, and various icons. Below the navigation bar, the main content area shows a single project named "Auto Discovery". The project details include "Last Updated: 08 Jan, 2020". On the right side of the project card, there are counts for "Assets" (0) and "Collaborators" (1). Below the project card, there are sections for "Recent activity" and "Date created" (08 Jan, 2020).

Click the **IBM Cloud Pak for Data** navigation menu in the top left corner.

The screenshot shows the IBM Cloud Pak for Data interface with the navigation menu open. The "Projects" option is selected. In the "Organize" section, the "All catalogs" option is highlighted with a red box. The main content area shows a summary of assets and collaborators, and a "Recent activity" section.

From the **Organize** section, select the **All catalogs** menu.

Your catalogs

New Catalog

Admin  
Auto Insurance

Creator: ctp  
Date created: Jan 08, 2020 1:37 PM

Auto Insurance Knowledge Catalog.

Admin  
Default Catalog

Creator: admin  
Date created: Nov 15, 2019 10:19 PM

Default catalog configured for sync

1. Click the **Auto Insurance** catalog.

Catalogs / Auto Insurance

Browse Assets Access Control Settings

Add to Catalog ▾

Local files (circled)

Connected asset

Connection

(i)

Now you can add assets!  
Click here to get started.

2. Click **Add to Catalog > Local files** from the catalog menu.

Catalogs / Auto Insurance

Add data assets from local files

Select File(s)  
Drop your files here  
or browse your files to add new files (up to 5 GB each)  
Files for assets are saved in the catalog's associated storage.

3 (circled)

Business Terms

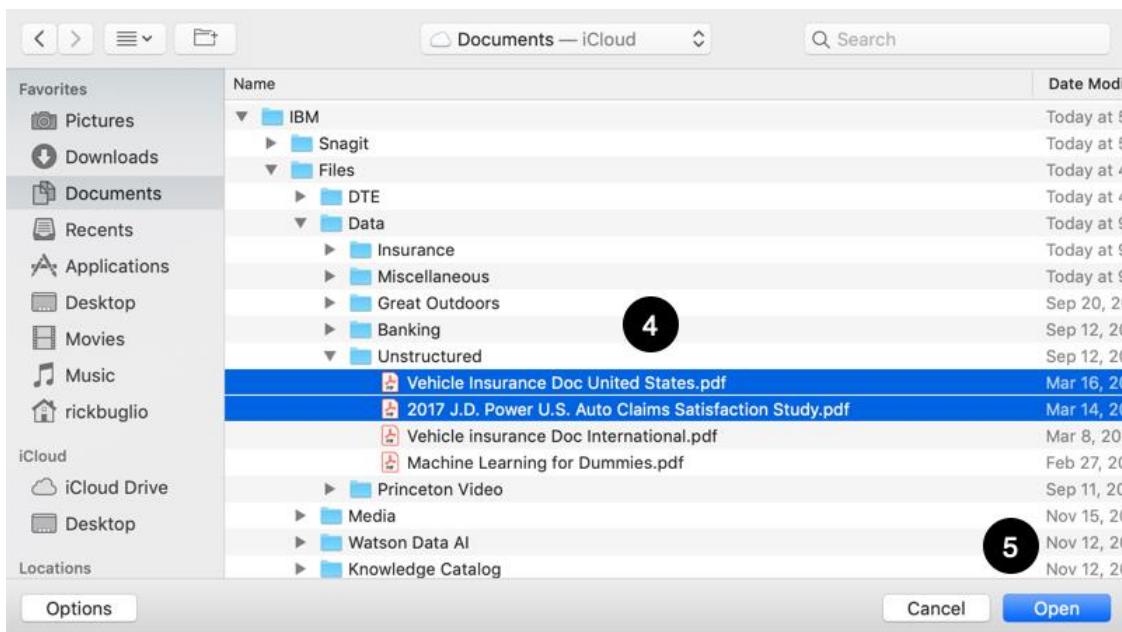
Search Business Terms

Tags

Start typing to add values +

Privacy

3. Click the **browse** link in the **Select File(s)** section to bring up the file selection dialog.



The MacOS **Finder** dialog is displayed. If you have a Windows system, it will look different, so depending on what system you are running, adjust to your system's method of selecting files.

4. Locate the “**Vehicle Insurance Doc United States.pdf**” and the “**2017 J.D. Power U.S. Auto Claims Satisfaction Survey.pdf**” files on your file system that you were instructed to download. Select both using the **Ctrl or Command key** on your keyboard (CTRL Click for Windows and Command Click for MacOS).

5. Click the **Open** button to begin cataloging the files.

Add data assets from local files

**Selected Files (2)\*** 6 Edit name and format

Continue adding files in the drop zone below or [browse](#) to select files

ASSET NAME	FORMAT	
Vehicle Insurance Doc United States.pdf	PDF	<span style="border: 1px solid #ccc; border-radius: 50%; padding: 2px;">×</span>
2017 J.D. Power U.S. Auto Claims Satisfactio...	PDF	<span style="border: 1px solid #ccc; border-radius: 50%; padding: 2px;">×</span>

6. Click the **pencil icon** next to the **Edit name and format** link.

This allows you to rename the data assets and change their file format. A default file format is inferred for you based on the file extension. In this case, they are PDF files, so **PDF** was auto selected. You **will not** change the format, but you will change their names by removing the file extension.

## Add data assets from local files

### Selected Files (2)\*

Asset Name	Format
Vehicle Insurance Doc United States	application/pdf
2017 J.D. Power U.S. Auto Claims Sat...	application/pdf

7      8      9

[Cancel](#) [Apply](#)

7. Click in the **Asset Name** area of the “**Vehicle Insurance Doc United States.pdf**” file. Go to the end of the filename and remove the **.pdf** extension.
8. Click in the **Asset Name** area of the “**2017 J.D. Power U.S. Auto Claims Satisfaction Survey.pdf**” file. Go to the end of the filename and remove the **.pdf** extension.
9. Click the **Apply** button to save the filename changes.

**Selected Files (2)\*** [Edit name and format](#)

Continue adding files in the drop zone below or [browse](#) to select files

Asset Name	Format
Vehicle Insurance Doc United States	PDF
2017 J.D. Power U.S. Auto Claims Satisfactio...	PDF

**Business Terms** [Search Business Terms](#)

**Tags** [Auto Insurance](#) [+](#) **12**

**Privacy**  Public  Private

All catalog members can find and use the asset.

**Target** [Auto Insurance](#)

**Members** [Add members](#) [+](#)

**Description**

Auto Insurance document **10**

10. Enter a Description of **Auto Insurance document**.
11. Enter a Tag of **Auto Insurance** into the **Tags** area.
12. Click the **+** sign next to the tag to add it.

Every time you enter a tag, you need to select the **+** sign to add the tag. The tags will appear as added tags in the tag area below the tag name. Once a tag is added, it can be used and selected for other data assets. Knowledge Catalog displays all available tags once they are added to the catalog. You will see this in action when you add the next file to the catalog.

**Selected Files (2)\***

Continue adding files in the drop zone below or [browse to select files](#)

Asset Name	Format	Actions
Vehicle Insurance Doc United States	PDF	
2017 J.D. Power U.S. Auto Claims Satisfactio...	PDF	

**Business Terms**

Search Business Terms

**Tags**

13 Document 14 Auto Insurance

**Privacy**

Public  Private

All catalog members can find and use the asset.

**Target**

Auto Insurance

13. Enter a Tag of **Document** into the **Tags** area.

14. Click the + sign next to the tag to add it.

**Selected Files (2)\***

Continue adding files in the drop zone below or [browse to select files](#)

Asset Name	Format	Actions
Vehicle Insurance Doc United States	PDF	
2017 J.D. Power U.S. Auto Claims Satisfactio...	PDF	

**Business Terms**

Search Business Terms

**Tags**

Start typing to add values 15 Auto Insurance Document

**Privacy**

Public  Private

All catalog members can find and use the asset.

**Description**

Auto Insurance document

Stay in the catalog until loading is complete! If you leave the catalog, the incomplete asset will be deleted.

**Members**

Add members

**Cancel** **Add**

The screenshot shows the **Auto Insurance** and **Document** tags that you should have entered for this asset. Make sure you have added them before you proceed to the next step that catalogs them.

15. Click the **Add** button to catalog the unstructured data assets.

Catalogs / Auto Insurance

Browse Assets Access Control Settings

2 files are being loaded into Auto Insurance.

Now you can add assets!  
Click [here](#) to get started.

A message is displayed notifying you that 2 assets are being loaded into the **Auto Insurance** catalog.

The screenshot shows the 'Auto Insurance' catalog page. At the top, there are tabs for 'Browse Assets', 'Access Control', and 'Settings'. A progress indicator shows '1 In progress' and 'Failed'. On the right, a message says 'No loads in progress' with a note to click the '+' action to add assets. Below this, the 'Recently Added' tab is highlighted (labeled 2). A search bar at the top has a placeholder 'What assets are you looking for?' and a dropdown labeled 'Any tag' (labeled 3). The main area displays two items: '2017 J.D. Power U.S. Auto ...' and 'Vehicle Insurance Doc Unit...'. Both items are data assets added on Aug 14, 2019, by Ricardo Buglio, with tags 'Auto Insurance' and 'Document'.

1. Click the X on the information dialog to close it if it remains open.
2. Click the **Recently Added** tab to view the contents.
3. Click in the **Any Tag** filter box to view the list of tags.

This screenshot shows the same catalog page after performing the steps in the previous list. The 'Any tag' filter dropdown is now open, displaying 'Auto Insurance' and 'Document' (labeled 1). The 'Recently Added' section still shows the two items from before. The bottom table also includes these items with their respective details.

Name	Owner	Tags	Type	Date Added
2017 J.D. Power U.S. Auto ...	Ricardo Buglio	Auto Insurance, Document	Data asset	Aug 14, 2019
Vehicle Insurance Doc Unit...	Ricardo Buglio	Auto Insurance, Document	Data asset	Aug 14, 2019

Upon completion, the data assets will automatically be added to the **Recently Added** section of the catalog asset browser. Scroll down and you will see the two newly added documents in the catalog with the tags you specified. Notice that the **Auto Insurance** and **Document** tags have been added to the **Tags** filter area.

## Auto Discover Data Assets

The screenshot shows the 'Auto Insurance' catalog page. At the top right, there is a blue button labeled 'Add to Catalog' with a dropdown arrow. A sub-menu is open, showing 'Local files', 'Connected asset', and 'Connection'. The 'Connection' option is highlighted with a blue background and has a black circle with the number '1' above it.

### 1. Click Add to Catalog > Connection from the catalog menu.

The screenshot shows the 'New connection' dialog. Under the 'IBM services' heading, the 'Db2 Warehouse' connector is selected and highlighted with a blue border and a black circle with the number '2' above it. Other options like 'Analytics Engine HDFS', 'Cognos Analytics', and 'Cloud Object Storage' are also listed.

**Notice** that the list of connectors to choose from is quite robust and includes all the IBM services and a generous number of Third-party services as well. Also, connection services are being added on a regular basis, so you may see more than the screenshot this tutorial is displaying.

### 2. Click on the Db2 Warehouse connector.

The screenshot shows the 'New connection (Db2 Warehouse - Db2 Warehouse)' configuration dialog. The process is numbered as follows:

3. In the 'Connection overview' section, 'Name' is set to 'Db2 Warehouse'.
4. In the 'Description' section, 'Knowledge Catalog Tutorial Db2 Warehouse' is entered.
5. In the 'Connection Details' section, 'Database' is set to 'BLUDB'.
6. In the 'Credentials' section, 'User name' is set to 'watsondemo'.
7. In the 'Connection Details' section, 'Host name or IP Address' is set to 'dashdb-enterprise-yp-dal13-16.services.dal.blu'.
8. In the 'Credentials' section, 'Password' is entered as '\*\*\*\*\*'.
9. In the 'Connection discovery' section, the checkbox 'Discover data assets' is checked.
10. In the 'Project for discovered assets' section, 'Auto Discovery' is selected.
11. At the bottom right, there are 'Test', 'Cancel', and 'Create' buttons, with 'Create' highlighted in red.

3. Enter a Name of **Db2 Warehouse**.
4. Enter a Description of **Knowledge Catalog Tutorial Db2 Warehouse**.
5. Enter a Database of **BLUDB**.
6. Enter a Username of **watsondemo**.
7. Copy and Paste **dashdb-enterprise-yp-dal13-16.services.dal.bluemix.net** into the Hostname or IP Address field.
8. Copy and Paste **WatsOnDataandAI!** into the Password field.
9. Click the **Discover data assets** check box under Connection discovery.
10. Select the **Auto Discovery** project from the “Project for discovered assets” selection list.
11. Click the **Test** button.

(+) New connection (Db2 Warehouse - Db2 Warehouse)

Connection overview		Connection Details		Connection discovery		
Name *	Db2 Warehouse	Database *	BLUDB	Host name or IP Address *	dashdb-enterprise-yp-dal13-16.services.dal.bluemix.net	Discover data assets
Description		Credentials			Project for discovered assets *	
Knowledge Catalog Tutorial Db2 Warehouse		<input type="radio"/> Personal	<input checked="" type="radio"/> Shared	Auto Discovery		
All project collaborators use the provided credentials for the connection.		User name *	WatsonDemo	Password *	WatsOnDataandAI!	Assets will be available in project Auto Discovery for you to publish to the Catalog
						<input type="button"/> New Project <input type="button"/> Reload

Enter information for the selected data source

Connection test passed →

When you see the green check mark and the message that the **Connection test passed**, click the **Create** button. If it does not pass the test, double check that you entered all the parameters correctly as stated in steps 3-10 above. If it still does not pass the test, notify the instructor.

You will receive a message that Knowledge Catalog is waiting for a response from the connection service (this is the auto discovery service) and a completion and redirection message. You will be brought back to the catalog asset browser and should see your newly added connection in the Data assets list.

Showing 3 of 3 items

<input type="checkbox"/>	Name	Owner	Tags	Business Terms	Type	Date Added
<input type="checkbox"/>	2017 J.D. Power U.S. Auto Claims Satisf...	Ricardo Buglio	Auto In... Docum...		Data asset	Aug 14, 2019
<input type="checkbox"/>	Db2 Warehouse	Ricardo Buglio			Connection	Aug 15, 2019

12. Click the **Recently Added** section of the asset browser.

**Notice** that all the data assets you added appear in this section. As assets are cataloged, they are added to the **Recently Added** section of the catalog asset browser in the sequence they were added, with the most recent appearing first in the list.

13. Click the **Any Type** filter.

**Notice** that the filter has a new asset type of **Connection**.

14. Click on the **Db2 Warehouse** asset in the **Recently Added** section.

**Overview**   **Access**   **Review**

**CONNECTION**  
Db2 Warehouse      [Remove](#)   [Download](#)   [Add to Project](#)

<b>Description</b> Knowledge Catalog Tutorial Db2 Warehouse	<b>Business Terms</b> There are no terms available for this asset.
Added: Aug 15, 2019 3:27 PM	<a href="#">Edit</a>
<b>Tags</b> 15	<b>Connection Preview</b> <a href="#">Click here to edit connection</a>
<b>Reviews</b> 0 reviews	
<b>Connection</b> Source type: Db2 Warehouse	
<b>Classification</b> None	

15. Hover next to the **Tags** section and Click the **pencil icon** to add tags to the connection.

**Description**  
Knowledge Catalog Tutorial Db2 Warehouse  
Added: Aug 15, 2019 3:27 PM

**Business Terms**  
There are no terms available for this asset.

**Tags** 16  
Start typing to add values +  
**Auto Insurance** X  
Document

**Connection Preview**  
Click [here](#) to edit connection

16. Click in the **Tags** section and select the **Auto Insurance** tag from the list of tags.

**Note** - If the list does not appear after you click in the Tags section, type in the letter **A**.

**Description**  
Knowledge Catalog Tutorial Db2 Warehouse  
Added: Aug 15, 2019 3:27 PM

**Business Terms**  
There are no terms available for this asset.

**Tags** 17  
Warehouse +  
No matches. Click + to add the text as a tag.  
**Auto Insurance** X

**Connection Preview**  
Click [here](#) to edit connection

17. Click in the **Tags** section, Enter the word **Warehouse** as a tag and Click the **+** sign next to the tag to add it.

**Description**  
Knowledge Catalog Tutorial Db2 Warehouse  
Added: Aug 15, 2019 3:27 PM

**Business Terms**  
There are no terms available for this asset.

**Tags** 18  
Start typing to add values +  
**Auto Insurance** X **Warehouse** X

**Connection Preview**  
Click [here](#) to edit connection

18. Click the **Apply** button.

The discovery process has been running as a service in the background, discovering and populating data assets into the **Auto Discovery** project. Let's examine the project to review the discovery results. We are interested in finding relevant auto insurance data, specifically

Customer, Policy and Claims data that will be used by the auto insurance claims web application.

The screenshot shows the IBM Cloud Pak for Data interface. At the top left is the navigation menu icon. Below it, the breadcrumb path reads 'Cloud Pak for Data > Auto Insurance > Db2 Warehouse'. The main content area displays a 'Db2 Warehouse' connection under the 'CONNECTION' section. Action buttons for 'Remove', 'Download', and 'Add to Project' are visible. A red arrow points to the navigation menu icon at the top left.

Click the **IBM Cloud Pak for Data** navigation menu in the top left corner.

The screenshot shows the IBM Cloud Pak for Data interface with the navigation menu expanded. The 'Projects' menu item is highlighted with a red box. Other options like 'Home', 'Connections', and 'My instances' are also visible. Action buttons for 'Remove', 'Download', and 'Add to Project' are at the bottom right. A red arrow points to the 'Projects' menu item.

Click on the **Projects** menu.

The screenshot shows the 'Projects' list interface. The 'Auto Discovery' project is selected and highlighted with a red box, labeled with the number 19. Other projects listed are 'Auto Insurance' and another unnamed project. The interface includes a search bar, a 'New project' button, and a table header for 'Name', 'Project Type', 'User Role', 'Last Updated', and 'Actions'.

19. From the **Projects** list, select the **Auto Discovery** project.

The screenshot shows the details of the 'Auto Discovery' project. The 'Assets' tab is selected and highlighted with a red box, labeled with the number 20. The project was last updated on '09 Jan, 2020'. Key statistics shown are 214 Assets and 1 Collaborator. The 'Recent activity' section shows two log entries: 'Discovery process has failed for connection Db2 Warehouse to project Auto Discovery' at 9:10 AM and 'Discovery process has started for connection Db2 Warehouse to project Auto Discovery' at 9:08 AM. The 'Storage' and 'Collaborators' sections are also visible.

**Notice** the number of data assets that were discovered. This discovery, when it was run, auto discovered 214 data assets and added them to the project. Your discovery may be different from this screenshot because this connection is a shared Db2 warehouse and data assets are being added and removed all the time by other users. Knowledge Catalog scanned the Db2

Warehouse on Cloud database instance and collected the metadata for all the user data assets that the “watsondemo” user **is authorized** to access.

Also, notice that the **Activity message** area has a notification that the “Discovery process has failed for connection **Db2 Warehouse** to project **Auto Discovery**”.

Don’t be alarmed. It may be a bit misleading, but it is **not an error**. The service is notifying you it failed because it was not able to discover all user data assets it found, and add them to the project, due to authorization constraints. The “watsondemo” user **does not** have select authority on some of the data assets in the warehouse, so they were not discoverable and added to the project.

**20.** Click the **Assets** tab to view the discovered assets.

**21.** Click in the **search area** and Enter the letters **custom** to find all data assets that start with the letter’s **custom**.

The list is filtered and displays the assets that meet the search criteria. You are looking for auto insurance customers but there are several assets that are related to customer information. The data asset named **CUSTOMERS** looks like it could be the right one. To verify it is related to auto insurance customers, you can preview the first 1000 records of the asset.

**22.** Click on the **CUSTOMERS** data asset to open the data previewer.

CUSTOMER	NAME	COUNTRY	LATITUDE	LONGITUDE	STREET_ADD...	CITY	STATE	STATE_C...
AB31813	Gayler Haburne	US	38.954005	-77.366611	12197 Sunset Hills R	Reston	Virginia	VA
AL96740	Emmott Ambrogi	US	41.31864251	-81.3461476	125 Barrington Town	Aurora	Ohio	OH
A074776	Bruce Coles	US	46.72891549	-122.9809094	1161 Harrison Ave.	Centralia	Washington	WA
AP23850	Quintilla Moreno	US	39.91090775	-74.9927597	1192 East Rt. 70	Cherry Hill	New Jersey	NJ
AP66226	Barnie MacGinlay	US	34.25722359	-118.5801846	20516 Devonshire St	Chatsworth	California	CA
AP67935	Leonid Elegood	US	42.90756255	-78.74992962	1 Walden Avenue	Cheektowaga	New York	NY
AP98768	Skell Crates	US	40.61252496	-80.05440962	10590 Perry Highway	Wexford	Pennsylvania	PA
AQ46199	Isadore Dougher	US	30.48057294	-97.83297956	1525 Cypress Creek I	Cedar Park	Texas	TX
AS99827	Rachel Brooks	US	33.807763	-118.273264	208 Sepulveda Blvd.	Carson	California	CA

The information panel on the right shows a tag of **AUTO\_INSURANCE**. This is the schema in the Db2 Warehouse instance that the table came from. Also, if you scroll to the right, you will see that there are columns related to auto insurance like **NUMBER\_OF\_POLICIES** and **NUMBER\_OF\_COMPLAINTS** etc. This is the auto insurance customers data asset that's needed.

The next several steps will demonstrate how you would publish this data asset to the **Auto Insurance** catalog. However, **you will not** publish it from the project. You will catalog it from the **Auto Insurance** catalog in a subsequent step, along with several other tables needed for the auto insurance analysis project, to demonstrate how you can add **Connected assets** from a Connection.

23. Click the **Auto Discovery** link at the top of the page to go back to the Auto Discovery project.

	NAME	TYPE	CREATED BY	LAST MODIFIED	ACTIONS
<input type="checkbox"/>	CUSTOMER_CHURN	Data Asset	ctp	9 Jan 2020, 9:10:12 am	
<input type="checkbox"/>	CUSTOMER_ACQUISITION	Data Asset	ctp	9 Jan 2020, 9:10:11 am	
<input checked="" type="checkbox"/>	CUSTOMERS	Data Asset	ctp	9 Jan 2020, 9:09:04 am	

24. Click on the **Assets** tab.

25. Click in the **search area** and Enter the letters **custom** to filter the Data assets list.

26. Click on the **checkbox** to the left of the **CUSTOMERS** data asset.

27. Notice that a **Publish** button appears at the top of the list. **Do not** click the button. This is the button you would select to publish the asset to a Knowledge Catalog, but you **will not** be publishing it to the catalog.

Click the **IBM Cloud Pak for Data** navigation menu in the top left corner.

The screenshot shows the Data Catalog interface. On the left, there's a sidebar with various navigation options like Home, Projects, Connections, My Instances, Collect, Organize, and All catalogs. The 'All catalogs' option is highlighted with a red box and has a black callout bubble with the number '28' above it. The main area shows a catalog entry for 'Auto Insurance' with details: Creator: ctp, Date created: Jan 08, 2020 1:37 PM. To the right, there's a file upload section with tabs for Load, Files, and Catalog, and a message: 'Drop files here or browse for files to upload.'

28. From the **Organize** section, select the **All catalogs** menu item.

The screenshot shows the 'Your catalogs' page. It lists two catalogs: 'Auto Insurance' and 'Default Catalog'. The 'Auto Insurance' catalog is highlighted with a red box and has a black callout bubble with the number '29' above it. Below each catalog entry, there are details like creator, date created, and a description. There's also a 'New Catalog' button at the top right.

29. Click on the **Auto Insurance** catalog.

### Catalog Structured Data

You have cataloged unstructured data files from the local file system and auto discovered data assets from a Db2 Warehouse connection. You will now catalog three tables from the Db2 Warehouse connection; **Claims**, **Customers** and **Policies** using the **Connected asset** catalog method. These tables are needed for the auto insurance claims analysis processing.

The screenshot shows the 'Auto Insurance' catalog page. At the top, there's a breadcrumb trail: Catalogs / Auto Insurance. Below it are tabs for Browse Assets, Access Control, and Settings. On the right, there's a 'Add to Catalog' button with a dropdown menu. The 'Connected asset' option is highlighted with a red box and has a black callout bubble with the number '1' above it. Other options in the menu include 'Local files' and 'Connection'.

1. Click **Add to Catalog > Connected asset** from the Catalog menu.

Catalogs / Auto Insurance

Add asset from connection

<b>Source*</b>	<b>Business Terms</b>
<b>Name*</b>	<b>Tags</b>
<b>Description</b>	<b>Privacy</b>
<b>Members</b>	

5. Select Source

2. Auto Insurance Claims

3. All U.S. auto insurance claims

4. Tags: Auto Insurance

6. Public

7. Add members

2. Enter a Name of **Auto Insurance Claims**.

3. Enter a Description of **All U.S. auto insurance claims**.

4. Click in the **Tags** area and select the **Auto Insurance** tag from the drop-down list.

**Note** - If the list does not appear after you click in the Tags section, type in the letter **A**.

5. Click the **Select Source** button to choose a Connection to add connected assets from.

Catalogs / Auto Insurance

Connections	Db2 Warehouse	AUTO_INSURANCE
Connections (1)	Schemas (20)	Tables (3)
Db2 Warehouse	AIOS	CLAIMS
	AUDIT	CUSTOMERS
6.	AUTO_INSURANCE	POLICIES
	BLUADMIN	
	DB2INST1	
	DTE_INSURANCE_DEMO	
	GOSALES	
	GOSALESDW	
	GOSALESHR	
	GOSALESMR	
	GOSALESRT	
	IBM_RTMON_DATA	
	JROY	
	NCR_FINANCIAL	

7. AUTO\_INSURANCE

8. CLAIMS

9. Select

6. Click on the **Db2 Warehouse** connection from the list of connections.

7. Click on the **AUTO\_INSURANCE** schema from the list of schemas.

8. Click on the **CLAIMS** table from the list of tables.

9. Click the **Select** button.

Catalogs / Auto Insurance

### Add asset from connection

**Source** Db2 Warehouse Change source

Db2 Warehouse : /AUTO\_INSURANCE/CLAIMS

**Name\***  
Auto Insurance Claims

**Description**  
All U.S. auto insurance claims

**Business Terms**  
Search Business Terms

**Tags**  
Start typing to add values +  
Auto Insurance X

**Privacy**  
 Public  Private  
All catalog members can find and use the asset.

**Members**  
Add members +

10 10

Cancel Add

**10.** Click the **Add** button.

You should see the table in the data asset list with the tag you supplied.

Catalogs / Auto Insurance

		+ Add to Catalog ^
		Local files
		Connected asset <span style="background-color: #005a9c; color: white; border: 1px solid #005a9c; padding: 2px 5px; border-radius: 3px;">11</span>
		Connection

**11.** Click **Add to Catalog > Connected asset** from the Catalog menu.

Catalogs / Auto Insurance

### Add asset from connection

**Source\***  
15 Select Source

**Name\***  
Auto Insurance Customers 12

**Description**  
All U.S. auto insurance customers 13

**Business Terms**  
Search Business Terms

**Tags**  
Start typing to add values +  
Auto Insurance X 14

**Classification\*** i  
None

**Privacy**  
 Public  Private  
All catalog members can find and use the asset.

**Members**

**12.** Enter a Name of **Auto Insurance Customers**.

**13.** Enter a Description of **All U.S. auto insurance customers**.

**14.** Click in the **Tags** area and select the **Auto Insurance** tag from the drop-down list.

**Note** - If the list does not appear after you click in the Tags section, type in the letter A.

**15.** Click the **Select Source** button to choose a Connection to add connected assets from.

The screenshot shows the 'Catalogs' interface with the path 'Catalogs / Auto Insurance'. On the left, there's a sidebar with 'Connections' and 'Db2 Warehouse' selected. The main area shows 'Schemas (20)' under 'Db2 Warehouse'. A list of schemas is displayed, with 'AUTO\_INSURANCE' highlighted by a callout labeled 17. To the right, under 'AUTO\_INSURANCE', there are three tables: 'CLAIMS', 'CUSTOMERS', and 'POLICIES', each with a callout labeled 18. At the bottom right are 'Cancel' and 'Select' buttons, with 'Select' highlighted by a callout labeled 19.

16. Click on the **Db2 Warehouse** connection from the list of connections.
17. Click on the **AUTO\_INSURANCE** schema from the list of schemas.
18. Click on the **CUSTOMERS** table from the list of tables.
19. Click the **Select** button.

The screenshot shows the 'Add asset from connection' dialog. It has a 'Source' section with 'Db2 Warehouse' selected and 'Db2 Warehouse : /AUTO\_INSURANCE/CUSTOMERS' shown. Below it are fields for 'Name\*' (Auto Insurance Customers) and 'Description' (All U.S. auto insurance customers). To the right, there are sections for 'Business Terms' (Search Business Terms), 'Tags' (Start typing to add values, with 'Auto Insurance' listed), 'Classification\*' (None), 'Privacy' (Public selected), and 'Members' (with a 'Public' radio button and a note that all catalog members can find and use the asset). At the bottom right are 'Cancel' and 'Add' buttons, with 'Add' highlighted by a callout labeled 20.

20. Click the **Add** button.
- You should see the table in the data asset list with the tag you supplied.

The screenshot shows the 'Catalogs' interface with the path 'Catalogs / Auto Insurance'. The top navigation bar includes 'Add to Catalog ^' and a search bar. Below, there are tabs for 'Browse Assets', 'Access Control', and 'Settings'. A dropdown menu is open at the top right, with 'Connected asset' highlighted by a callout labeled 21.

21. Click **Add to Catalog > Connected asset** from the Catalog menu.

Catalogs / Auto Insurance

Add asset from connection

<b>Source*</b>	<b>Business Terms</b>
<b>Name*</b>	<b>Tags</b>
<b>Description</b>	<b>Classification*</b>
<b>Tags</b>	<b>Privacy</b>
<b>Members</b>	

22. **Name\***: Auto Insurance Policies  
 23. **Description**: All U.S. auto insurance policies  
 24. **Tags**: Start typing to add values  
 25. **Select Source**  
 26. **Classification\***: None  
 27. **Public**  
 28. **Private**  
 29. **All catalog members can find and use the asset.**

22. Enter a Name of **Auto Insurance Policies**.

23. Enter a Description of **All U.S. auto insurance policies**.

24. Click in the **Tags** area and select the **Auto Insurance** tag from the drop-down list.

**Note** - If the list does not appear after you click in the Tags section, type in the letter A.

25. Click the **Select Source** button to choose a Connection to add connected assets from.

Catalogs / Auto Insurance

Connections	Db2 Warehouse	AUTO_INSURANCE
Connections (1)	Schemas (20)	Tables (3)
Db2 Warehouse	AIOS	CLAIMS
	AUDIT	CUSTOMERS
26.	AUTO_INSURANCE	POLICIES
	BLUADMIN	
	DB2INST1	
	DTE_INSURANCE_DEMO	
	GOSALES	
	GOSALESDW	
	GOSALESHR	
	GOSALESMR	
	GOSALESR	
	IBM_RTMON_DATA	
	JROY	
	NCR_FINANCIAL	

27. **Select**

26. Click on the **Db2 Warehouse** connection from the list of connections.

27. Click on the **AUTO\_INSURANCE** schema from the list of schemas.

28. Click on the **POLICIES** table from the list of tables.

29. Click the **Select** button.

Catalogs / Auto Insurance

Add asset from connection

**Source** Db2 Warehouse **Change source**

**Name\*** Auto Insurance Policies

**Description** All U.S. auto insurance policies

**Business Terms** Search Business Terms

**Tags** Start typing to add values **+ Auto Insurance X**

**Classification\*** **None**

**Privacy**  Public  Private

All catalog members can find and use the asset.

**Members** 30

**Add**

30. Click the **Add** button.

Catalogs / Auto Insurance

Watson Recommends Highly Rated Recently Added 31

Watson Recommends	Highly Rated	Recently Added	Connection
Data asset Auto Insurance Policies Owner: Ricardo Buglio Added: Aug 15, 2019 5:49 PM Tags: Auto ...  ★★★★★ 0 reviews	Data asset Auto Insurance Customers Owner: Ricardo Buglio Added: Aug 15, 2019 5:44 PM Tags: Auto ...  ★★★★★ 0 reviews	Data asset Auto Insurance Claims Owner: Ricardo Buglio Added: Aug 15, 2019 5:37 PM Tags: Auto ...  ★★★★★ 0 reviews	Db2 Warehouse Owner: Ricardo Buglio Added: Aug 15, 2019 3:27 PM Tags: Auto ... Waren...  ★★★★★ 0 reviews

31. Click the **Recently Added** tab.

You should see the cataloged tables and connection in the **Recently Added** section as a data asset with the tags you supplied.

## Understand and Socialize Data Assets

As data assets are cataloged, they are automatically profiled and classified so data consumers can have a better understanding of their content. They can then be enriched using Knowledge Catalog's social capabilities.

In this task, you will visit the **Profile** section of a structured data asset to examine the profiling and classification features provided. You will also visit the **Review** section to experience how you can rate and review assets to allow others to easily identify and evaluate them based on their ranking and comments.

Catalogs > Auto Insurance

Browse Assets Access Control Settings

## Auto Insurance

What assets are you looking for?

Any type Any source Any name

Watson Recommends Highly Rated Recently Added

**1**

**2**

Data asset **Auto Insurance Policies**

Owner: ctp Added: Jan 09, 2020 3:02 PM...PM Tags: Auto ...

Watson Recommends Highly Rated Recently Added

**1**

**2**

Data asset **Auto Insurance Customers**

Owner: ctp Added: Jan 09, 2020 10:06 AM...AM Tags: Auto ...

Data asset **Auto Insurance Claims**

Owner: ctp Added: Jan 09, 2020 10:00 AM...AM Tags: Auto ...

Connection **Db2 Warehouse**

Owner: ctp Added: Jan 08, 2020 6:06 PM...PM Tags: Auto ... Wareh...

Data asset **Vehicle Ins**

Owner: ctp Added: Jan 08 Tags: Auto ...

**3**

1. Click on the **Recently Added** section of the data asset browser.
2. Click on the **Auto Insurance Customers** asset to view its properties.

Catalogs > Auto Insurance > Auto Insurance Customers

Add to Catalog

Overview Access Review Profile Lineage

**1**

**2**

**3**

### DATA ASSET

#### Auto Insurance Customers

Description All U.S. auto insurance customers

Added: Jan 09, 2020 10:06 AM...AM  
Format: application/octet-stream  
Size: 312 KB

Business Terms There are no terms available for this asset.

Tags Auto Insurance

Reviews ★★★★☆ 0 reviews

Connection Source: Db2 Warehouse  
Source type: Db2 Warehouse

Classification None

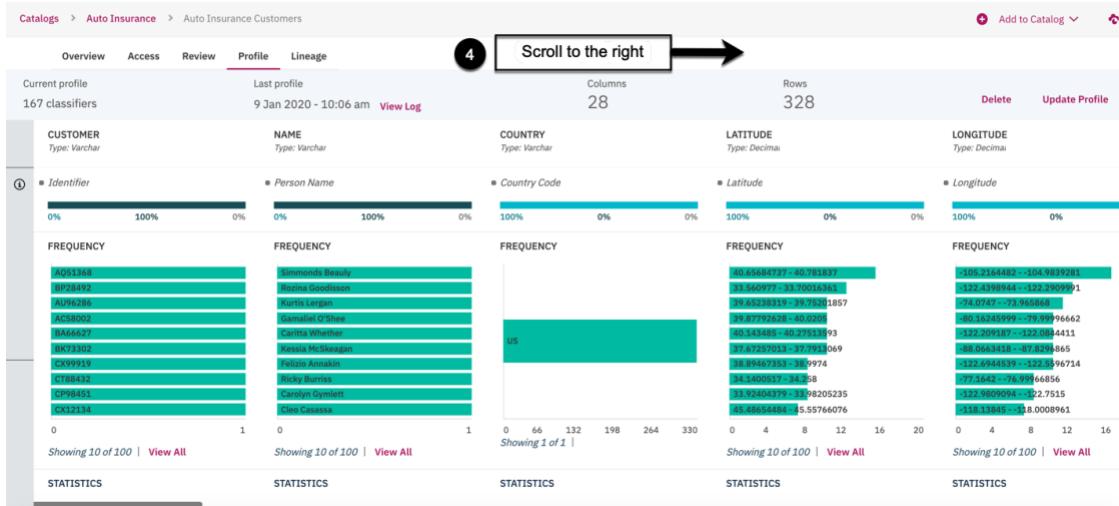
Schema: 28 Columns 328 Rows

CUSTOMER	NAME	COUNTRY	LATITUDE	LONGITUDE	STREET_ADDR...	CITY	STATE	STATE_CO...	ZIP_CODE
AB72731	Adolph Skitch	US	41.75113981	-88.0127658	1001 W 75th Street	Woodridge	Illinois	IL	60517
AB96670	Wynnie Dunnett	US	39.2781	-120.1203	100 Northstar Dr	Truckee	California	CA	96161
AC58002	Waylan Treleven	US	47.76121	-122.3464	18325 Aurora Ave N	Shoreline	Washington	WA	98133
AH99727	Penny Duckhous	US	33.88081187	-118.0288381	16610 Valley View Av	La Mirada	California	CA	90638
AN36757	Malory Gantz	US	34.02091598	-84.31698227	1530 Old Alabama Rc	Roswell	Georgia	GA	30076
AP53665	Sianna Kissell	US	41.77025751	-88.20481022	1951 W Jefferson Ave	Naperville	Illinois	IL	60540
AR47849	Janine McCreath	US	40.9581	-74.0747	177 Route 17 South	Paramus	New Jersey	NJ	7652
AS97690	Milicent Caveau	US	40.8591	-73.9694	136-38 Linwood Plaza	Fort Lee	New Jersey	NJ	7024
AW77988	Agnes Woodfield	US	38.9974	-105.0672	1101 E Hwy 24	Woodland Park	Colorado	CO	80863
AY40674	Jamaal Duddle	US	35.080383	-81.70435	102 Lemmons Lane	Gaffney	South Carolina	SC	29341

You are brought to the **Overview** section of the asset where you can view a 1000 row sample of the data and metadata about the asset, including column level classifications if it's a data asset. You can modify its name and description, add tags and assign business terms and classifications at the asset or column level.

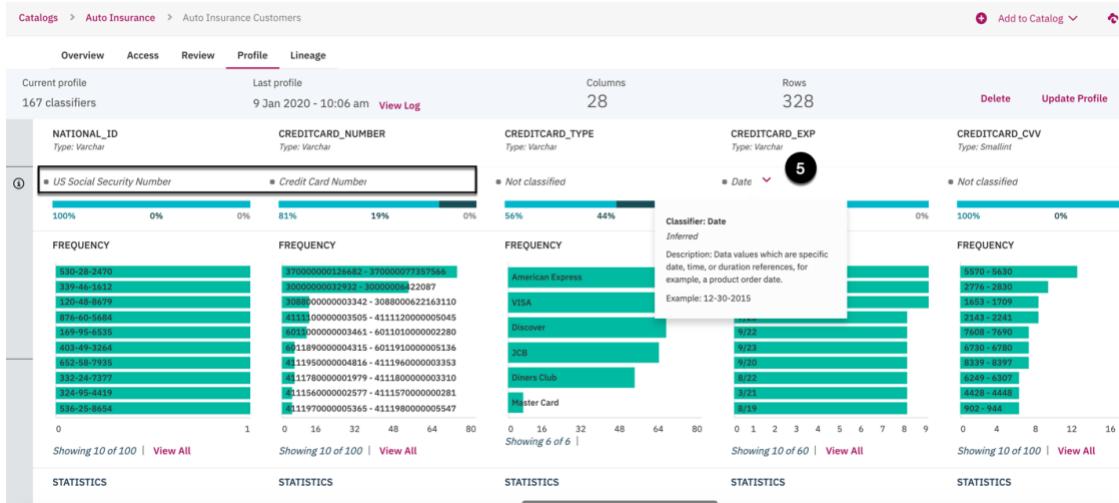
3. Click on the **Profile** section of the data asset.

The profile should automatically appear. If not, and you are presented with a method to create or update the profile, follow the instructions to do so.



The profile of a data asset that contains relational or structured data, shows information about each column in the data set, based on the first 5,000 rows of data. The profile shows the frequency of the inferred attribute classifiers and statistics about the data for each column. [Attribute classifiers](#) describe the contents of the data in the column: for example, city, account number or credit card number. Attribute classifiers are necessary to [anonymize data](#) with data policies. The attribute classifiers appear for each column on the asset's **Overview** and **Profile** page.

#### 4. Scroll to the right until you see the NATIONAL\_ID column statistics.

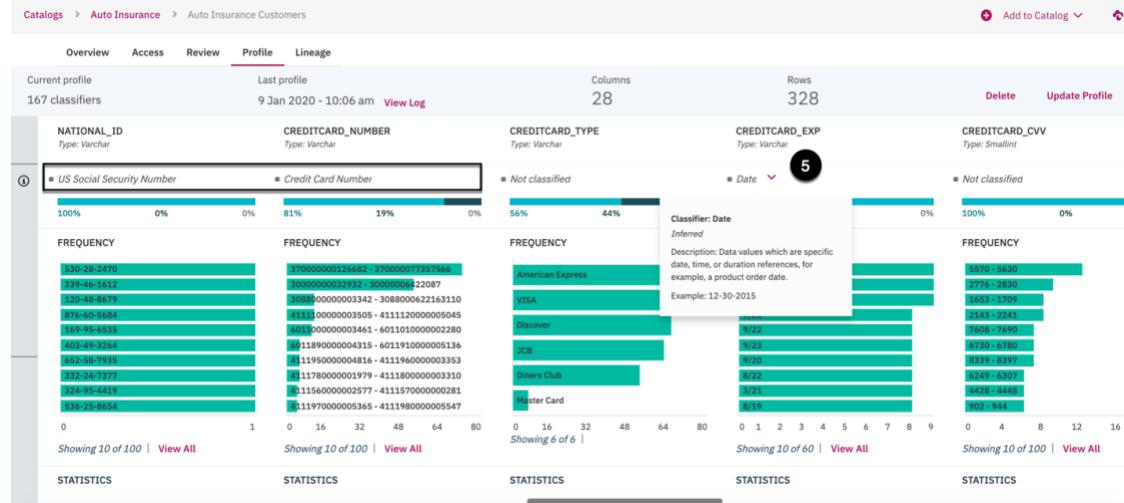


**Note** - You will notice that the NATIONAL\_ID and CREDITCARD\_NUMBER columns are classified correctly but the CREDITCARD\_EXP and CREDITCARD\_CVV columns are not. You will also notice they look like sensitive information and should be protected by data protection rules.

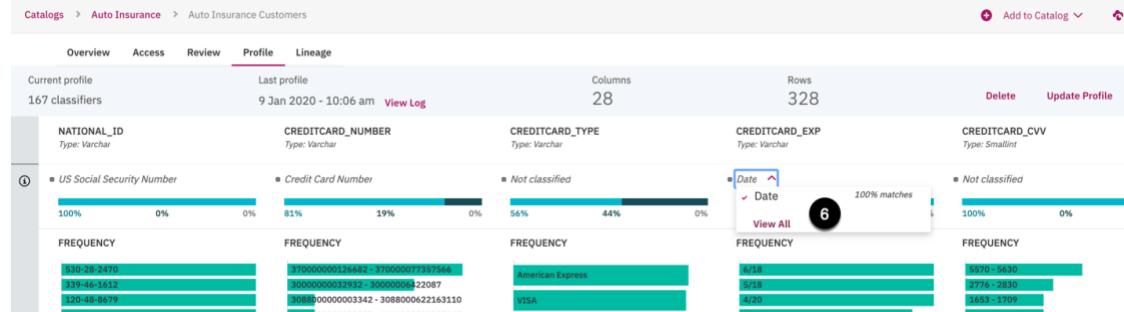
There are no data protection rules created in the glossary of this Cloud Pak for Data cluster to protect sensitive information. Even if there were, they would not be enforced and stop you from viewing the data content because you added the data asset to the catalog and are the owner. If another user were to log in and view the data, and rules were active, they would be enforced.

Time permitting, you will get a chance to experience how data protection works at the end of the lab so you can see it in action.

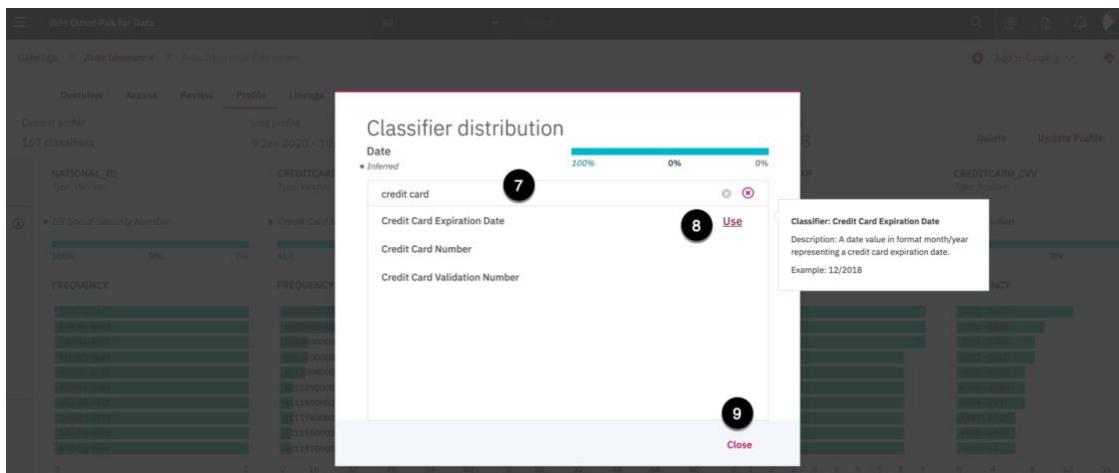
In any case, in order for any data protection rule based on a data class to be enforced, the data class assignments have to be correct. To protect the Credit Card Expiration Date and Validation Number of a Credit Card, the data classes for the **CREDITCARD\_EXP** and **CREDITCARD\_CVV** columns have to be set properly so you will do that now.



5. Click on the down arrow next to the **Date** classification of the **CREDITCARD\_EXP** column.



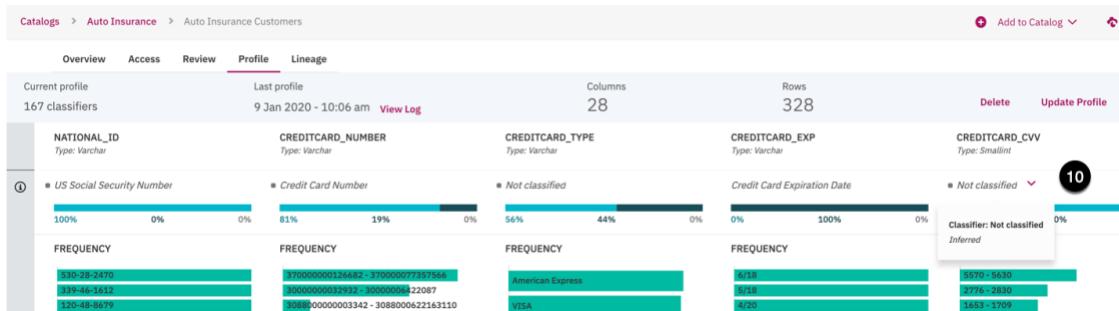
6. Click the **View All** menu item.



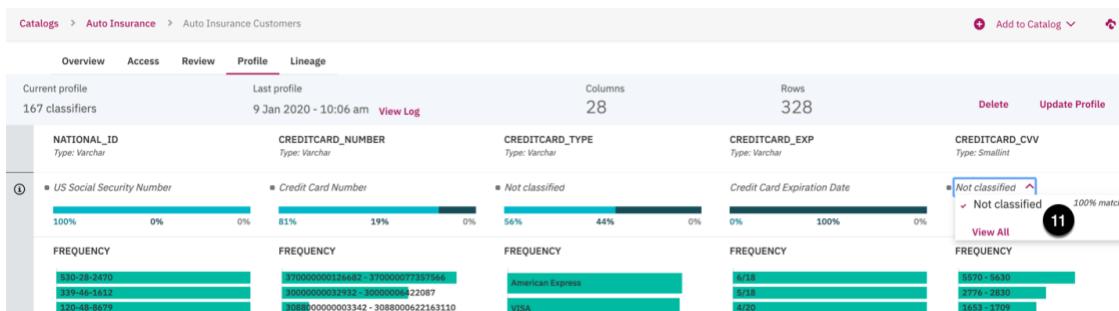
7. Type in **credit card** in the search area.

8. Hover next to the **Credit Card Expiration Date** data class and click **Use**.

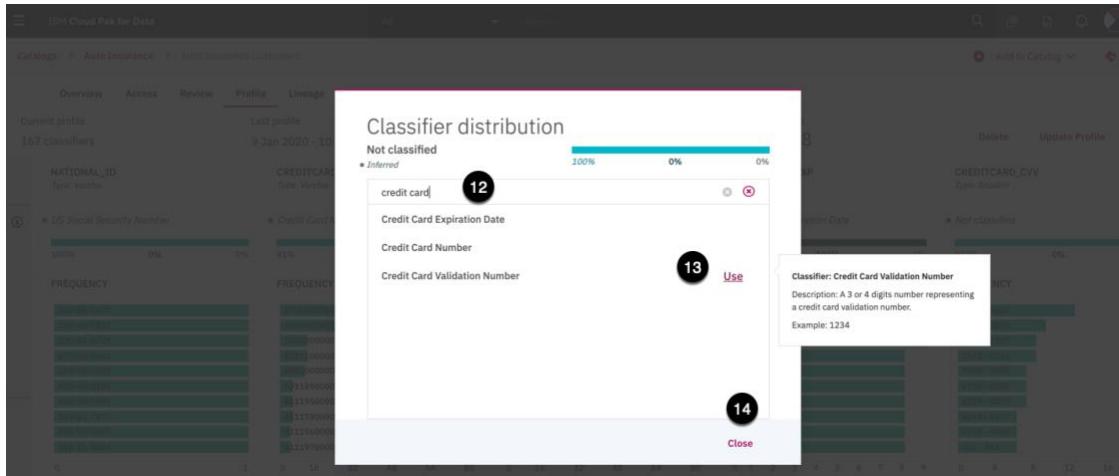
9. Click the **Close** button.



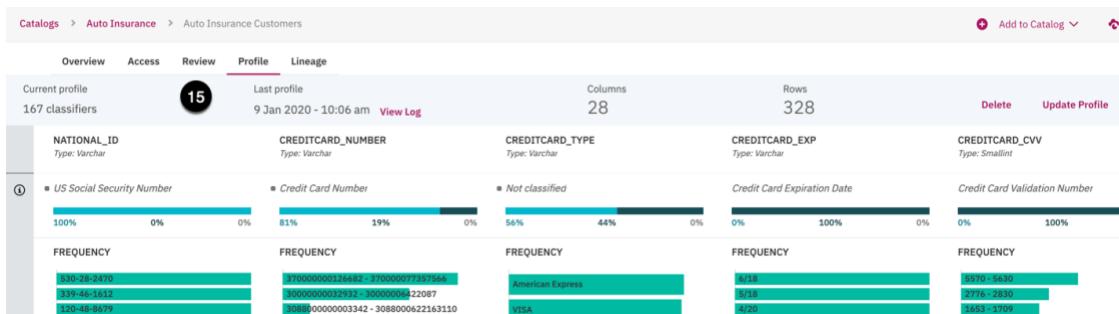
10. Click on the down arrow next to the **Not classified** classification of the CREDITCARD\_CVV column.



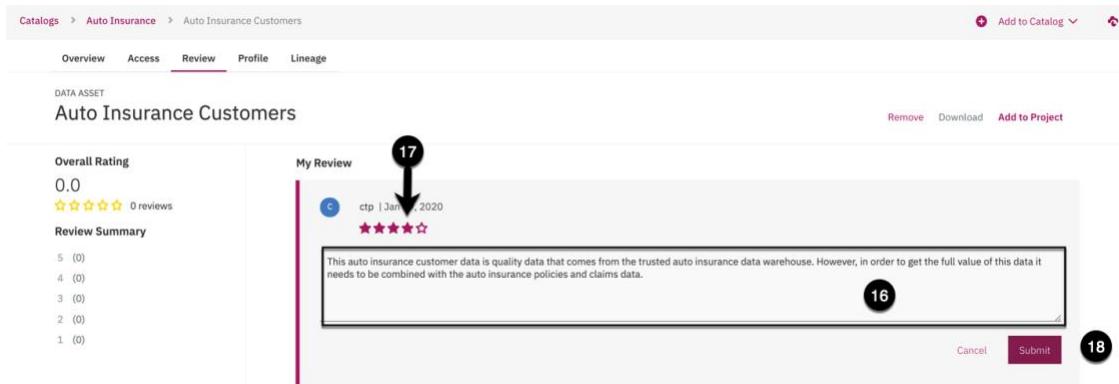
11. Click the **View All** menu item.



12. Type in **credit card** in the search area.
13. Hover next to the **Credit Card Validation Number** data class and click **Use**.
14. Click the **Close** button.



15. Click the **Review** section to rate and review the data asset.



16. Copy and Paste the following bolded text into the Description: **This auto insurance customer data is quality data that comes from the trusted auto insurance data warehouse. However, in order to get the full value of this data it needs to be combined with the auto insurance policies and claims data.**
17. Click the **4th star** from the left to give the asset a 4-star rating.

18. Click the **Submit** button.

The screenshot shows the 'Review' tab of the 'Auto Insurance Customers' data asset. On the left, there's a 'Review Summary' section with an overall rating of 4.0 and 1 review. On the right, there's a 'My Review' section by user 'ctp' dated Jan 09, 2020, with a 4-star rating and a comment: 'This auto insurance customer data is quality data that comes from the trusted auto insurance data warehouse. However, in order to get the full value of this data it needs to be combined with the auto insurance policies and claims data.' There are 'Edit' and 'Delete' buttons below the review.

**Notice** that you now have one review with an Overall Rating of 4.0.

19. Click the **Auto Insurance** link at the top of the page to go back to the catalog asset browser.

The screenshot shows the 'Highly Rated' section of the 'Auto Insurance' asset browser. It lists several data assets and connections, each with a star rating and review count. The 'Auto Insurance Customers' asset is highlighted with a 4-star rating and 1 review. Below the list is a table showing detailed properties for the selected asset.

Name	Owner	Tags	Business Terms	Type	Date Added
2017 J.D. Power U.S. Auto Claims Satisfaction Study	ctp	Auto Insu... Document		Data asset	Jan 09, 2020
Auto Insurance Claims	ctp	Auto Insu...		Data asset	Jan 09, 2020

20. Click on the **Highly Rated** section and notice that the **Auto Insurance Customers** data asset is the most highly rated data asset with 1 review.

21. From the data asset list below, click on the **2017 J.D. Power U.S. Auto Claims Satisfaction Study** asset to view its properties.

Catalogs > Auto Insurance > 2017 J.D. Power U.S. Auto Claims ... Add to Catalog

**23** Overview Access Review Lineage

**DATA ASSET**  
2017 J.D. Power U.S. Auto Claims Satisfaction Study

Description  
Auto Insurance document  
Added: Jan 09, 2020 6:15 PM..PM  
Format: PDF  
Size: 1 MB

Business Terms  
There are no terms available for this asset.

Tags  
Auto Insurance Document

Reviews  
★ ★ ★ ★ 0 reviews

Classification  
None

d2cdbeec-6b58-46c2-8364-0c52c9531fc1 1 / 6



Remove Download Add to Project

The **Overview** section displays the document and allows you to view its contents. Notice that numerous viewing controls appear along with action buttons to print, download and rotate the document. If you do not see the controls, place your cursor inside the document viewing area towards the top.

22. Scroll down to view the content of the document.

23. Click the **Review** section to rate and review the data asset.

Catalogs > Auto Insurance > 2017 J.D. Power U.S. Auto Claims ... Add to Catalog

Overview Access **Review** Lineage

**DATA ASSET**  
2017 J.D. Power U.S. Auto Claims Satisfaction Study

Overall Rating  
0.0  0 reviews

Review Summary  
5 (0)  
4 (0)  
3 (0)  
2 (0)  
1 (0)

My Review  
**25**   
ctp | Jan 09, 2020   
Very interesting survey of auto claims satisfaction but will not be useful for our auto claims analytics project. **24**

Cancel Submit **26**

24. Copy and Paste the following bolded text into the Description **Very interesting survey of auto claims satisfaction but will not be useful for our auto claims analytics project.**

25. Click the **3rd star** from the left to give the asset a 3-star rating.

26. Click the **Submit** button.

The screenshot shows the 'Auto Insurance' asset details page. At the top, there's a breadcrumb navigation: Catalogs > Auto Insurance > 2017 J.D. Power U.S. Auto Claims ... A 'Review' tab is highlighted. Below the tabs, the title '2017 J.D. Power U.S. Auto Claims Satisfaction Study' is displayed. On the left, there's a 'Review Summary' section with an overall rating of 3.0 and a 4-star review. On the right, a review by 'ctp' dated Jan 09, 2020, is shown, stating: 'Very interesting survey of auto claims satisfaction but will not be useful for our auto claims analytics project.' Buttons for 'Edit' and 'Delete' are at the bottom right of the review.

**Notice** that you now have one review with an Overall Rating of 3.0.

27. Click the **Auto Insurance** link at the top of the page to go back to the catalog asset browser.

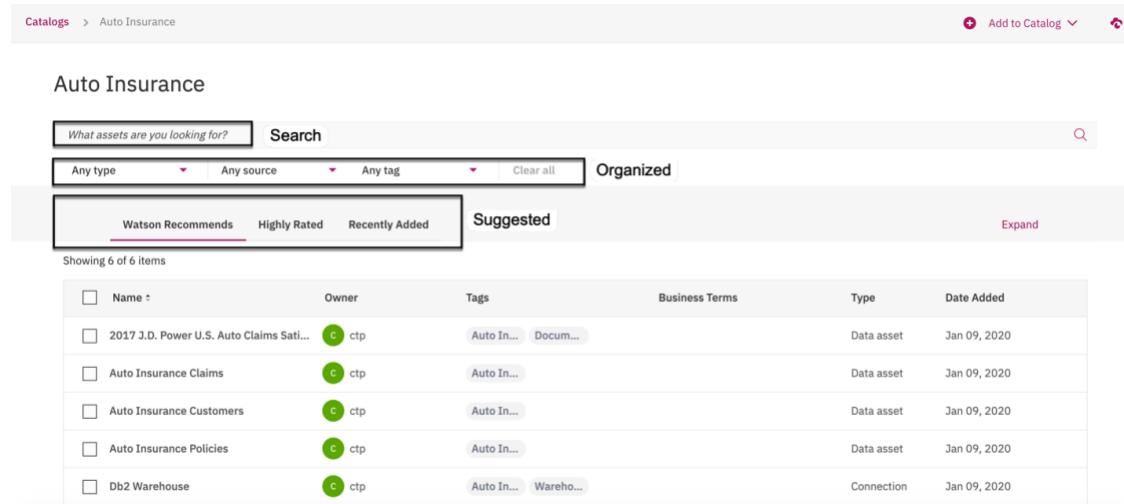
The screenshot shows the 'Auto Insurance' asset browser. At the top, there are tabs for 'Browse Assets', 'Access Control', and 'Settings'. The 'Browse Assets' tab is selected. In the center, there's a search bar and a 'Watson Recommends' section. The 'Highly Rated' section is highlighted with a red box. It lists four data assets: 'Auto Insurance Customers', '2017 J.D. Power U.S. Auto ...', 'Auto Insurance Claims', and 'Vehicle Insurance Doc Uni...'. The '2017 J.D. Power U.S. Auto ...' asset is the second item in the list. On the far right of this section, there's a 'Collapse' button.

28. Click the **Highly Rated** section.

Notice that the **2017 J.D. Power U.S. Auto Claims Satisfaction Study** data asset is now showing as the 2nd highest rated data asset with 1 review.

29. Click the **Collapse** button to the far right of where the suggestions are to close the area in preparation for the next task.

## Shop for Data



The screenshot shows the Knowledge Catalog interface for the 'Auto Insurance' catalog. At the top, there's a navigation bar with 'Catalogs' and 'Auto Insurance'. On the right, there are buttons for 'Add to Catalog' and a search icon. Below the navigation is a search bar with placeholder text 'What assets are you looking for?' and a 'Search' button. Underneath the search bar are filter dropdowns for 'Any type', 'Any source', and 'Any tag', followed by a 'Clear all' button and an 'Organized' button. Below these filters are three tabs: 'Watson Recommends' (which is selected), 'Highly Rated', and 'Recently Added'. To the right of these tabs is a 'Suggested' section and an 'Expand' button. The main area displays a table titled 'Showing 6 of 6 items' with columns: Name, Owner, Tags, Business Terms, Type, and Date Added. The table lists the following assets:

Name	Owner	Tags	Business Terms	Type	Date Added
2017 J.D. Power U.S. Auto Claims Sat...	ctp	Auto In... Docum...		Data asset	Jan 09, 2020
Auto Insurance Claims	ctp	Auto In...		Data asset	Jan 09, 2020
Auto Insurance Customers	ctp	Auto In...		Data asset	Jan 09, 2020
Auto Insurance Policies	ctp	Auto In...		Data asset	Jan 09, 2020
Db2 Warehouse	ctp	Auto In... Wareho...		Connection	Jan 09, 2020

In this task, you will leverage Knowledge Catalog's intelligent **Shop for Data** AI-powered **Search and Suggest** experience that guides you to the most relevant assets in the catalog, based on understanding of relationships between assets, usage of those assets and social connections between the users of those assets.

You will also use the **Filter** section of the Knowledge Catalog that is automatically built and **Organized** by Asset Type and Tag as you catalog assets. Tagging is essential when cataloging assets, it expedites the process for consumers to easily search and find what they are looking for.

### Shop using Suggestions

You just experienced this type of search in the previous task. You can easily search for data using the **Watson Recommends**, **Highly Rated** and **Recently Added** suggestion categories to find relevant data. These categories are automatically populated by Knowledge Catalog as you catalog, curate and enrich data assets.

### Shop using Search

In this section you will shop for data by specifying search criteria using the **Search area** (Where it reads *What assets are you looking for?*) of the Knowledge Catalog asset browser. Note that search criteria are not case sensitive.

Catalogs / Auto Insurance Add to Catalog

Browse Assets Access Control Settings

### Auto Insurance

**1**

Name	Owner	Tags	Business Terms	Type	Date Added
2017 J.D. Power U.S. Auto Claims Satisf...	Ricardo Buglio	Auto In... Docum...		Data asset	Aug 14, 2019
Vehicle Insurance Doc United States	Ricardo Buglio	Auto In... Docum...		Data asset	Aug 14, 2019

**2**

1. Inside the Knowledge Catalog search area type in **document**.

Data assets are displayed that have a tag, column name, asset name or description that contains the consecutive letters of **document**.

2. Click the **x** at the far right of the search area to clear the search.

Catalogs / Auto Insurance Add to Catalog

Browse Assets Access Control Settings

### Auto Insurance

**3**

Name	Owner	Tags	Business Terms	Type	Date Added
Db2 Warehouse	Ricardo Buglio	Auto In... Wareho...		Connection	Aug 15, 2019

**4**

3. Inside the Knowledge Catalog search area type in **warehouse**.

Data assets are displayed that have a tag, column name, asset name or description that contains the consecutive letters of **warehouse**.

4. Click the **x** at the far right of the search area to clear the search.

Browse Assets Access Control Settings

### Auto Insurance

**5**

Name	Owner	Tags	Business Terms	Type	Date Added
2017 J.D. Power U.S. Auto Claims Satisf...	Rick Buglio	Auto In... Docum...		Data asset	Nov 28, 2018
Auto Insurance Claims	Rick Buglio	Auto In...		Data asset	Sep 05, 2019
Auto Insurance Policies	Rick Buglio	Auto In...		Data asset	Sep 05, 2019

**6**

5. Inside the Knowledge Catalog search area type in **claim**.

Data assets are displayed that have a tag, column name, asset name or description that contains the consecutive letters of **claim**.

Why is the **Auto Insurance Policies** table in the result set? This is an example of the search finding an asset that has a column that contains the consecutive characters **claim** in its name. This table has three columns named **LAST CLAIM**, **DENIED CLAIMS** and **CLAIMS FILED** that meet the criteria. And remember, search is not case sensitive. You will see these columns when you prepare the data in the next task.

6. Click the **x** at the far right of the search area to clear the search.

Using the same search method, type in the following searches in the **search area**. Clear the search area after each search to get the correct results:

- Enter the characters **all** in the search area and view the results.
- Enter the characters **U.S** in the search area and view the results.
- Enter the characters **study** in the search area and view the results.
- Enter the characters **unit** in the search area and view the results.

Time permitting, you can experiment on your own and type in different criteria in the **search area** to get more experience with how search works.

### Search using Filters

In this section, you shop for data using the **Filter** area that is automatically built by Knowledge Catalog as assets are added to the catalog. You can use one to many filters in combination with each other to get the desired results you are looking for. You may also get an empty search result depending on the combinations you specify.

The screenshot shows the Knowledge Catalog interface. At the top, there's a navigation bar with 'Catalogs / Auto Insurance'. On the right, there are buttons for 'Add to Catalog' and a share icon. Below the navigation is a search bar with the placeholder 'What assets are you looking for?'. To the right of the search bar are several filter dropdowns: 'Filter' (set to 'Any type'), 'Any tag', 'Modified on', 'Modified by', and a 'Clear all' button. Below the search bar, there are three collapsed suggestion categories: 'Watson Recommends', 'Highly Rated', and 'Recently Added'. On the far right of this section is a blue 'Expand' button, which is highlighted with a red box. At the bottom of the screenshot, there's a footer bar with icons for Home, Search, Catalogs, and Help.

Before you begin, make sure the **search area** is cleared out and that the suggestion categories are collapsed. You should see an **Expand** button if they are collapsed, like in the screenshot above. If you see a **Collapse** button, select it to collapse the section to gain more viewing real estate.

Browse Assets    Access Control    Settings

### Auto Insurance

Q What assets are you looking for?

Any type    Any tag    Clear all

Connection  
 Data asset  
 Document

Highly Rated    Recently Added    Expand

Showing 6 of 6 items

Name	Owner	Tags	Business Terms	Type	Date Added
2017 J.D. Power U.S. Auto Claims Satisf...	Rick Buglio	Auto In... Docum...		Data asset	Nov 28, 2018

In the **Filter** area of the catalog browser:

1. Click in the **Type** filter area and select the **Data asset** type and view the results.

Browse Assets    Access Control    Settings

### Auto Insurance

Q What assets are you looking for?

Type    Any tag    Clear all

Type  
 Any tag

Showing 5 of 5 items

Name	Owner	Tags	Business Terms	Type	Date Added
2017 J.D. Power U.S. Auto Claims Satisf...	Rick Buglio	Auto In... Docum...		Data asset	Nov 28, 2018
Auto Insurance Claims	Rick Buglio	Auto In...		Data asset	Sep 05, 2019
Auto Insurance Customers	Rick Buglio	Auto In...		Data asset	Sep 05, 2019
Auto Insurance Policies	Rick Buglio	Auto In...		Data asset	Sep 05, 2019
Vehicle Insurance Doc United States	Rick Buglio	Auto In... Docum...		Data asset	Nov 28, 2018

Only the 5 assets that have an asset type of **Data asset** are displayed; 2 files and 3 tables. The other types of assets that can be cataloged are **Connections, Models, Notebooks and Dashboards**.

Browse Assets    Access Control    Settings

### Auto Insurance

Q What assets are you looking for?

Type    Any tag    Clear all

Type  
 Any tag

Showing 5 of 5 items

Name	Owner	Tags	Business Terms	Type	Date Added
2017 J.D. Power U.S. Auto Claims Satisf...	Rick Buglio	Auto In... Docum...		Data asset	Nov 28, 2018

Type  
 Document

2. Click in the **Tag** filter area and select the **Document** tag and view the results.

Browse Assets   Access Control   Settings

## Auto Insurance

What assets are you looking for?

Type   Tag   Clear all   3

Showing 2 of 2 items

Name	Owner	Tags	Business Terms	Type	Date Added
2017 J.D. Power U.S. Auto Claims Satisfaction Survey	Rick Buglio	Auto Ins... Document		Data asset	Nov 28, 2018
Vehicle Insurance Doc United States	Rick Buglio	Auto Ins... Document		Data asset	Nov 28, 2018

Only **Data assets** that have a tag of **Document** are displayed.

- Click the **Clear All** button to clear all filters.

## Prepare Data for Analytics and AI

In this task, you will take the structured data you cataloged in the **Auto Insurance** catalog into the **Auto Insurance** analytics project you created and prepare the data for analytics and AI. You will gain an understanding of the data preparation capabilities within a Cloud Pak for Data analytic project and how it can also help you understand and visualize the data before and after preparation.

### Add Cataloged Data Assets to a Project

In order to refine data, the data needs to be in a project. You will add three auto insurance data assets from the **Auto Insurance** catalog to the Auto Insurance project to prepare it for analytics and AI. There are two ways to add cataloged assets to a project; from the catalog and from a project. You will add the cataloged assets to the **Auto Insurance** project from the **Auto Insurance** catalog.

Catalogs > Auto Insurance

Add to Catalog

Browse Assets   Access Control   Settings

## Auto Insurance

What assets are you looking for?

Any type   Any source   Any tag   Clear all

Watson Recommends   Highly Rated   Recently Added   Expand

4 Add to Project   Remove   3 of 6 items selected   Cancel

Name	Owner	Tags	Business Terms	Type	Date Added
2017 J.D. Power U.S. Auto Claims Satisfaction Survey	ctp	Auto Ins... Document		Data asset	Jan 09, 2020
<input checked="" type="checkbox"/> Auto Insurance Claims	ctp	Auto Insu...		Data asset	Jan 09, 2020
<input checked="" type="checkbox"/> Auto Insurance Customers	ctp	Auto Insu...		Data asset	Jan 09, 2020
<input checked="" type="checkbox"/> Auto Insurance Policies	ctp	Auto Insu...		Data asset	Jan 09, 2020
<input type="checkbox"/> Db2 Warehouse	ctp	Auto Insu... Warehouse		Connection	Jan 09, 2020

- Click the **check box** next to the **Auto Insurance Claims** data asset.
- Click the **check box** next to the **Auto Insurance Customers** data asset.

3. Click the **check box** next to the **Auto Insurance Policies** data asset.

4. Click the **Add to Project** button at the top of the data asset list.

Asset Name	Catalog	Connection
Auto Insurance Claims	Auto Insurance	Db2 Warehouse
Auto Insurance Customers	Auto Insurance	Db2 Warehouse
Auto Insurance Policies	Auto Insurance	Db2 Warehouse

Notice on the right that the **Db2 Warehouse** connection was also included to be added to the Auto Insurance project; The connection is needed by the project to access the data from the Db2 Warehouse.

5. Select the **Auto Insurance** project from the list of Target projects.

6. Click the **Add** button.

You are brought back into the **Auto Insurance** Knowledge Catalog after the data assets are added to the project. A message at the top of the catalog will inform you that the assets were successfully added to the project.

Click the **IBM Cloud Pak for Data** navigation menu in the top left corner.

The screenshot shows the 'Settings' page of the IBM Cloud Pak for Data interface. The left sidebar is open, showing various navigation options like Home, Projects (which is highlighted), Connections, My instances, Collect, Organize, All catalogs, Information assets, Data and AI governance, Categories, Business terms, Classifications, Data classes, Reference data, Policies, and Rules. A callout bubble with the number '7' is positioned over the 'Projects' button.

## 7. Click the Projects menu.

The screenshot shows the 'Projects' list interface. The left sidebar is closed, and the main area displays a table of projects. The columns are Name, Project Type, User Role, Last Updated, and Actions. There are two projects listed: 'Auto Insurance' (Analytics, Admin, 8 Jan 2020, 4:05 PM) and 'Auto Discovery' (Analytics, Admin, 9 Jan 2020, 9:07 AM). A callout bubble with the number '8' is positioned over the 'Auto Insurance' project row.

## 8. From the Projects list, Select the Auto Insurance project.

The screenshot shows the 'My Projects' overview page. The left sidebar includes tabs for Overview, Assets, Environments, Jobs, Access Control, and Settings. The main area has sections for 'Data assets' (listing Auto Insurance Policies, Auto Insurance Customers, Auto Insurance Claims, and Big2 Warehouse) and 'Data Refinery flows' (listing Auto Insurance Data Flow). A central dialog box titled 'Gather your resources' contains text about adding team and data assets, and a 'Start tour' button. A large callout bubble with the letter 'X' is positioned at the top right corner of this dialog box.

If you see the Getting Started dialog appear, click on the X in the top right corner to close it.

## Refine the Data

The screenshot shows the 'Assets' tab selected in the navigation bar. Below the search bar, there's a section titled 'Data assets' with a list of four items: 'Auto Insurance Customers', 'Auto Insurance Policies', 'Auto Insurance Claims', and 'Db2 Warehouse'. To the right of the list is a 'New data asset' button and an 'Actions' dropdown menu. Step 1 is marked with a circle at the top left of the navigation bar. Step 2 is marked with a circle inside the ellipsis menu. Step 3 is marked with a circle inside the 'Refine' option in the dropdown.

1. Click on the **Assets** tab at the top of the project page.
2. Select the **ellipsis...** to the right of the **Auto Insurance Customers** data asset to view the data asset action menu.
3. Select the **Refine** menu item.

You are brought into the **data preparation** component of the analytic project to begin shaping the **Auto Insurance Customers** data. In the subsequent steps, you will use some of the data preparation operations to shape the auto insurance data you added to the project and create a newly shaped dataset that you will put back to the project as a **CSV** file that will be used by the analytics project team.

The screenshot shows the 'Data Refinery Flow Details' page for the 'Auto Insurance Customers' dataset. It includes sections for 'DATA SOURCE', 'LOCATION', 'DATA REFINERY FLOW NAME', 'STEPS', and 'DATA REFINERY FLOW OUTPUT'. Step 4 is marked with a circle on the left side of the table, and step 6 is highlighted with a red box and a 'Next' button.

Country	Latitude	Longitude
US	41.75113981	-88.0127658
US	39.2781	-120.1203
US	47.76121	-122.3464
US	33.88081187	-118.0288881
US	34.02091598	-84.31698227
US	41.77025751	-88.20481022
US	40.9581	-74.0747
US	40.8591	-73.9694
US	38.9974	-105.0672
US	35.080383	-81.70435
US	42.339208	-71.134786
US	42.43478788	-83.43361611
US	39.58050569	-105.1353397
US	33.98205235	-118.2491117
US	41.913853	-88.31240528

If you see the Getting Started dialog appear, click on the X in the top right corner to close it.

The screenshot shows the Data Refinery interface. On the left, there is a table titled 'Auto Insurance Customers' with columns: CUSTOMER, NAME, COUNTRY, LATITUDE, and LONGITUDE. The table contains 11 rows of data. On the right, there is a sidebar titled 'DATA REFINERY FLOW DETAILS' with sections for 'Data Source' (set to 'Auto Insurance Customers'), 'LOCATION' (set to 'Auto Insurance'), 'DATA REFINERY FLOW NAME' (set to 'Auto Insurance Customers\_flow'), and 'STEPS' (set to 0). A large 'Edit' button is prominently displayed at the bottom of the sidebar. A circled number '4' is located in the top right corner of the sidebar.

#### 4. Click in the **Edit** button to edit the Data Flow details.

The screenshot shows the 'Edit' dialog for the Data Refinery flow. It has two main sections: 'DATA REFINERY FLOW DETAILS' on the left and 'DATA REFINERY FLOW OUTPUT' on the right. In the 'DATA REFINERY FLOW DETAILS' section, the 'LOCATION' is set to 'Auto Insurance'. The 'DATA REFINERY FLOW NAME' field is highlighted with a circled number '5'. In the 'DATA REFINERY FLOW OUTPUT' section, the 'LOCATION' is set to 'Auto Insurance/Data assets', and the 'DATA SET NAME' is set to 'Auto Insurance Customer...'. A checkbox for overwriting existing data is checked. A note at the bottom says 'File format: CSV'. A circled number '6' is located in the top right corner of the 'DATA REFINERY FLOW DETAILS' section.

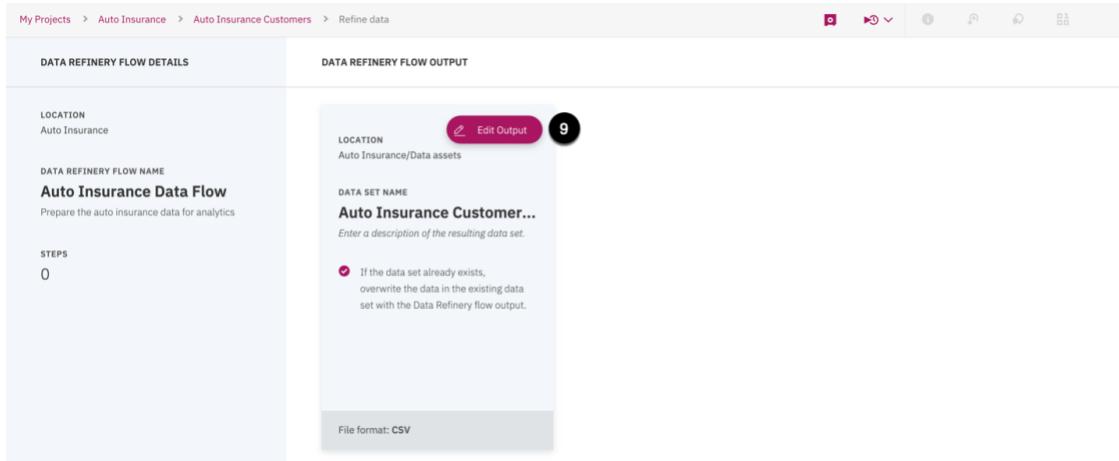
#### 5. Click the **pencil icon** in the DATA REFINERY FLOW NAME area of the DATA REFINERY FLOW DETAILS section.

The screenshot shows the 'Edit' dialog after renaming the flow. The 'DATA REFINERY FLOW DETAILS' section now shows the 'DATA REFINERY FLOW NAME' field with the value 'Auto Insurance Data Flow' (circled with '6'). The 'DESCRIPTION' field contains the text 'Prepare the auto insurance data for analytics' (circled with '7'). The 'DATA REFINERY FLOW OUTPUT' section remains the same. A circled number '8' is located in the bottom right corner of the 'DATA REFINERY FLOW DETAILS' section.

#### 6. Rename the Data Flow to **Auto Insurance Data Flow** with the proper case, and spaces between the words.

7. Copy and paste, or enter, this bolded text **Prepare the auto insurance data for analytics** into the DESCRIPTION field.

8. Click the **Apply** button.

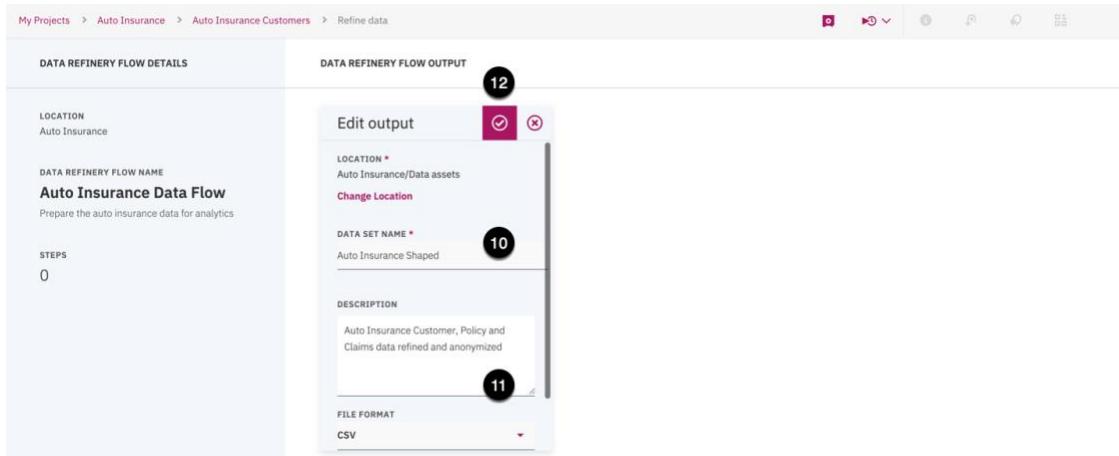


Hover over the **pencil icon** in the LOCATION area of the DATA REFINERY FLOW OUTPUT section.

9. Click the **Edit Output** button to change the DATA SET NAME.

In this section, you can change the data flows output target location. You can choose any connector, supported as a target connector, that is available as part of the Cloud Pak for Data common fabric. However, only connectors defined to the project you are in, that can be targets, will be displayed to select from.

For this tutorial, you will change the DATA SET NAME but **not** the LOCATION. The target location will be the default location, the **Auto Insurance** project.



10. Rename the Data Set to **Auto Insurance Shaped** with the proper case, spaces between the words, and removal of the **.csv** extension.

11. Copy and paste, or enter, this bolded text **Auto Insurance Customer, Policy and Claims data combined and refined** into the DESCRIPTION field.

12. Click the **Save** button (looks like a check mark) on the toolbar to save the changes.

The screenshot shows the Data Refinery interface. On the left, the 'DATA REFINERY FLOW DETAILS' panel displays the location as 'Auto Insurance', the flow name as 'Auto Insurance Data Flow', and a note that it prepares auto insurance data for analytics. It also shows '0 STEPS'. On the right, the 'DATA REFINERY FLOW OUTPUT' panel shows the location as 'Auto Insurance/Data assets', the dataset name as 'Auto Insurance Shaped', and a note that it contains 'Auto Insurance Customer, Policy and Claims data refined and anonymized'. It includes an option to overwrite existing data if it already exists. The file format is set to CSV. At the bottom, there is a note to review the details and a 'Done' button.

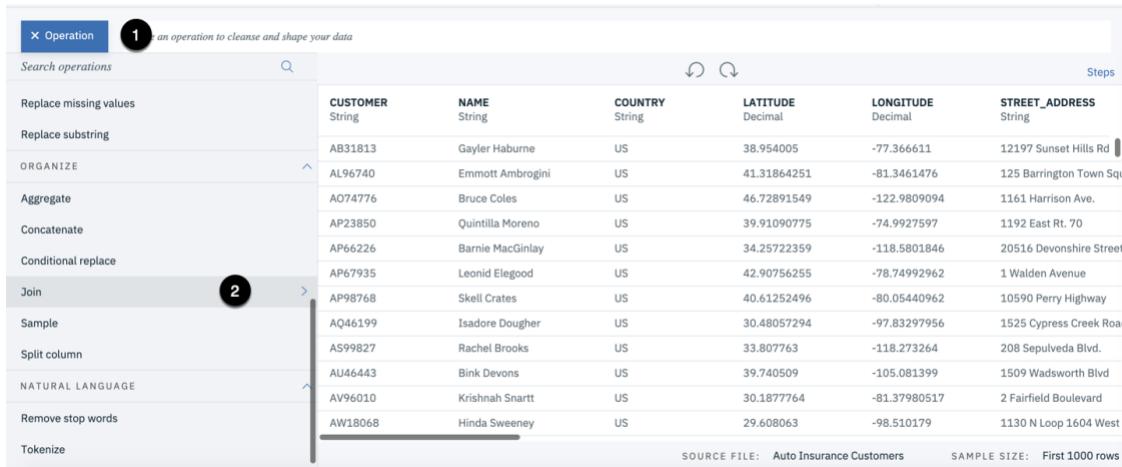
13. Click the **Done** button.

The screenshot shows the Data Refinery interface. On the left, the 'Operation' panel displays a table of customer data with columns: CUSTOMER, NAME, COUNTRY, LATITUDE, and LONGITUDE. The table lists 15 rows of data. Below the table are buttons for 'SOURCE FILE:' (Auto Insurance Customers) and 'SAMPLE SIZE:' (First 328 rows). On the right, the 'Details' panel is open, showing the 'Edit' tab selected. It displays the 'DATA REFINERY FLOW DETAILS' section with the flow name 'Auto Insurance Data Flow' and a note about preparing auto insurance data for analytics. It also shows the 'DATA REFINERY FLOW OUTPUT' section with the dataset name 'Auto Insurance Shaped' and a note about containing refined and anonymized data. The 'Steps' section indicates 0 steps.

14. Click the **Save** button on the toolbar to save the Data Flow.

15. Click the **X** on the Details panel to close the panel and maximize the shaper real estate.

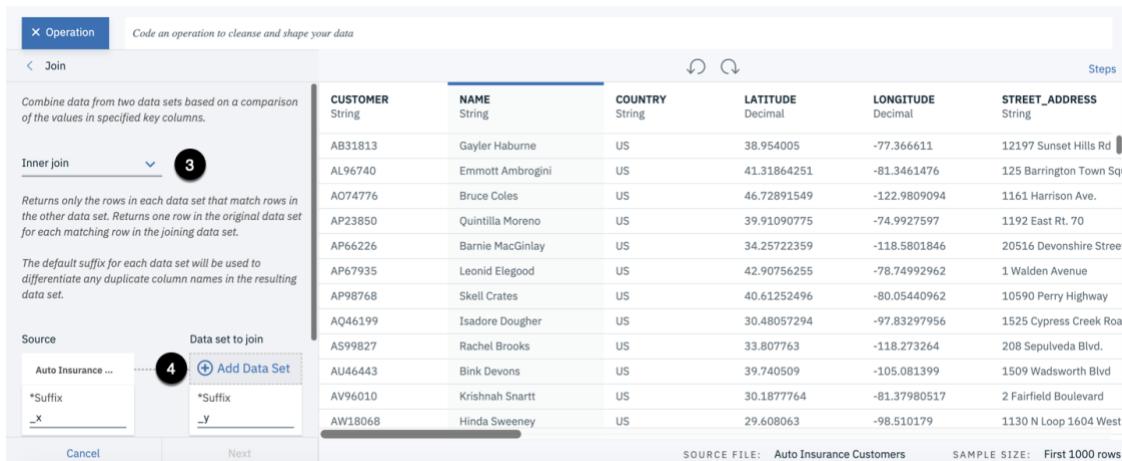
## Combine Data



The screenshot shows the Data Operations interface. At the top left, there is a button labeled "Operation" with a circled number "1" next to it. Below it is a search bar with the placeholder "Search operations". A list of operations is displayed, including "Replace missing values", "Replace substring", "ORGANIZE", "Aggregate", "Concatenate", "Conditional replace", "Join" (which is highlighted with a circled number "2"), "Sample", "Split column", "NATURAL LANGUAGE", "Remove stop words", and "Tokenizer". To the right of the list is a preview table with columns: CUSTOMER, NAME, COUNTRY, LATITUDE, LONGITUDE, and STREET\_ADDRESS. The table contains 10 rows of sample data. At the bottom of the interface, there are buttons for "SOURCE FILE: Auto Insurance Customers" and "SAMPLE SIZE: First 1000 rows".

1. Click the **Operation** button to view the shaping operations menu.

2. Scroll down and click the **Join** operation.



The screenshot shows the "Join" operation configuration dialog. At the top left, there is a "Code an operation to cleanse and shape your data" input field with the placeholder "Combine data from two data sets based on a comparison of the values in specified key columns." Below it is a dropdown menu showing "Inner join" (with a circled number "3"). A note below the dropdown states: "Returns only the rows in each data set that match rows in the other data set. Returns one row in the original data set for each matching row in the joining data set. The default suffix for each data set will be used to differentiate any duplicate column names in the resulting data set." On the right side, there is a preview table identical to the one in the previous screenshot. At the bottom, there are buttons for "Cancel" and "Next" (with a circled number "4") and a "Data set to join" section containing "Auto Insurance ...", "\*Suffix", and "x". There is also a "+ Add Data Set" button.

3. Select the **Inner join** method from the join method list.

4. Click the **+ Add Data Set** button in the **Data set to join** section.

Data set to join with Auto Insurance Customers

Auto Insurance

Data assets

- Assets (2)
- Connections
- Data assets (3)
- > Auto Insurance Claims
- > Auto Insurance Customers
- 5 > Auto Insurance Policies
- 6

Cancel Apply

5. Click on the **Data assets** section.
6. Select the **Auto Insurance Policies** data asset.
7. Click the **Apply** button.

**X Operation** Code an operation to cleanse and shape your data

Join

The default suffix for each data set will be used to differentiate any duplicate column names in the resulting data set.

Source Data set to join

Auto Insurance ... Auto Insurance ... 8

\*Suffix X

Auto Insurance ... 9

\*Suffix Y

JOIN KEYS 10

CUSTOMER	NAME	COUNTRY	LATITUDE	LONGITUDE	STREET_ADDRESS
AB31813	Gayler Haburne	US	38.954005	-77.366611	12197 Sunset Hills Rd
AL96740	Emmett Ambrogini	US	41.31864251	-81.3461476	125 Barrington Town Squ
A074776	Bruce Coles	US	46.72891549	-122.9809094	1161 Harrison Ave.
AP23850	Quintilla Moreno	US	39.91090775	-74.9927597	1192 East Rt. 70
AP66226	Barnie MacGinlay	US	34.25722359	-118.5801846	20516 Devonshire Street
AP67935	Leonid Elegood	US	42.90756255	-78.74992962	1 Walden Avenue
AP98768	Skell Crates	US	40.61252496	-80.05440962	10590 Perry Highway
AQ46199	Isadore Douther	US	30.48057294	-97.83297956	1525 Cypress Creek Roa
AS9827	Rachel Brooks	US	33.807763	-118.273264	208 Sepulveda Blvd.
AU46443	Bink Devons	US	39.740509	-105.081399	1509 Wadsworth Blvd
AV96010	Krishnah Snartt	US	30.1877764	-81.37980517	2 Fairfield Boulevard
AW18068	Hinda Sweeney	US	29.608063	-98.510179	1130 N Loop 1604 West

Steps

Cancel Next

SOURCE FILE: Auto Insurance Customers SAMPLE SIZE: First 1000 rows

Scroll down in the Join properties area until you see the JOIN KEYS section.

8. Click in the **Auto Insurance Customers** JOIN KEYS column selection list on the left and select the **CUSTOMER** column as the join key column.
9. Click in the **Auto Insurance Policies** JOIN KEYS column selection list on the right and select the **CUSTOMER** column as the join key column.
10. Click the **Next** button.

**Join**

CUSTOMER	NAME	COUNTRY	LATITUDE	LONGITUDE	STREET_ADDRESS
AB31813	Gayler Haburne	US	38.954005	-77.366611	12197 Sunset Hills Rd
AL96740	Emmott Ambrogini	US	41.31864251	-81.3461476	125 Barrington Town Sq
AO74776	Bruce Coles	US	46.72891549	-122.9809094	1161 Harrison Ave.
AP23850	Quintilla Moreno	US	39.91090775	-74.9927597	1192 East Rt. 70
AP66226	Barnie MacGinlay	US	34.25722359	-118.5801846	20516 Devonshire Street
AP67935	Leonid Elegood	US	42.90756255	-78.7492962	1 Walden Avenue
AP98768	Skell Crates	US	40.61252496	-80.05440962	10590 Perry Highway
AQ46199	Isadore Dougher	US	30.48057294	-97.83297956	1525 Cypress Creek Roa
AS9827	Rachel Brooks	US	33.807763	-118.273264	208 Sepulveda Blvd.
AU46443	Bink Devons	US	39.740509	-105.081399	1509 Wadsworth Blvd
AV96010	Krishnah Snartt	US	30.1877764	-81.37980517	2 Fairfield Boulevard
AW18068	Hinda Sweeney	US	29.608063	-98.510179	1130 N Loop 1604 West

SOURCE FILE: Auto Insurance Customers SAMPLE SIZE: First 1000 rows

Back Apply

## 11. Scroll down the column list and **uncheck** the following columns:

CREDITCARD\_NUMBER, CREDITCARD\_TYPE, CREDITCARD\_EXP, CREDITCARD\_CVV, INCOME

Unchecking columns excludes them from the join result.

## 12. Click the **Apply** button.

**Policies Columns**

POLICY_ID	COVERAGE	EFFECTIVE_TO...	POLICY_TYPE	POLICY	RENEW_OFFER	SALES
1 AZ2668345	Basic	2019-01-19	Personal Auto	Personal L2	offer3	Call Ce
2 CQ7350150	Extended	2019-01-20	Personal Auto	Personal L1	offer4	Agent
3 RB9672905	Extended	2019-02-16	Personal Auto	Personal L3	offer1	Branch
4 GC4500212	Basic	2019-01-10	Personal Auto	Personal L1	offer1	Agent
5 CJ8259258	Basic	2019-02-18	Personal Auto	Personal L3	offer1	Branch
6 E36624580	Extended	2019-01-20	Personal Auto	Personal L1	offer4	Web
7 PL6745054	Basic	2019-01-28	Personal Auto	Personal L3	offer3	Branch
8 OO1895498	Basic	2019-01-01	Personal Auto	Personal L2	offer1	Agent
9 OV9148795	Basic	2019-02-25	Personal Auto	Personal L2	offer1	Web
10 MT9162281	Premium	2019-01-16	Personal Auto	Personal L3	offer2	Branch
11 ZF2778838	Extended	2019-01-14	Personal Auto	Personal L3	offer3	Branch
12 HG7711818	Premium		Personal Auto	Personal L2	offer2	Branch

SOURCE FILE: Auto Insurance Customers SAMPLE SIZE: First 463 rows

The join should complete successfully. **Scroll** to the right to see that the Auto Insurance Policies table columns are now appended at the end of the Auto Insurance Customers table in the shaper.

**Notice** that the Steps panel appears with the Join as the first shaping step. The Steps panel lets you view your shaping operations, in the order they are performed, and allows for the modification and removal of steps to back out shaping operations done in error or no longer needed.

13. Click the **Operation** button to view the shaping operations menu.

14. Scroll down and Click the **Join** operation.

15. Select the **Inner join** method from the join method list.

16. Click the **+ Add Data Set** button in the **Data set to join** section.

17. Click on the **Data assets** section.

18. Select the **Auto Insurance Claims** data asset.

19. Click the **Apply** button.

The screenshot shows the 'Join' configuration screen. On the left, under 'Source', 'Auto Insurance ...' is selected. On the right, 'Data set to join' is set to 'Auto Insurance ...'. Below these, two dropdown menus labeled 'X' and 'Y' are shown, each with a suffix '(S)' and a 'Join Key' selection list containing 'POLICY\_ID'. A large preview table on the right lists data from both datasets, showing columns like LATITUDE, STREET\_ADDRESS, CITY, STATE, POLICY\_ID, COVERAGE, EFFECTIVE\_TO..., and POLICY\_TYPE. The table has 562 rows. At the bottom, there are 'Cancel' and 'Next' buttons, with 'Next' being highlighted.

Scroll down in the Join properties area to view a full list of columns.

20. Click in the **Auto Insurance Customers** JOIN KEYS column selection list on the left and select the **POLICY\_ID** column as the join key column.

21. Click in the **Auto Insurance Claims** JOIN KEYS column selection list on the right and select the **POLCY\_ID** column as the join key column.

22. Click the **Next** button.

The screenshot shows the 'Join' configuration screen after applying the changes. The left sidebar now lists 'LATITUDE', 'LONGITUDE', 'STREET\_ADDRESS', 'CITY', 'STATE', 'STATE\_CODE', 'ZIP\_CODE', 'EMAIL\_ADDRESS', 'PHONE\_NUMBER', 'GENDER', and 'NATIONAL\_ID' with checkboxes checked. The right side shows the joined dataset with columns: POLICY\_ID, COVERAGE, EFFECTIVE\_TO..., POLICY\_TYPE, and a partially visible 'P' column. The table has 463 rows. At the bottom, there are 'Back' and 'Apply' buttons, with 'Apply' being highlighted.

23. Click the **Apply** button.

The screenshot shows the Data Flow interface with the 'Claims Columns' table selected. The table has columns: CLAIM\_ID, FIRST\_NOTICE..., RESPONSE, CLAIM\_REASON, and INCIDENT\_SUMMARY. The 'Steps' panel on the right shows two steps: 'Data Source' (Auto Insurance Customers) and 'Join' (inner-joined data from Auto Insurance Policies based on columns CUSTOMER,CUSTOMER). Another 'Join' step is listed under 'JUST ADDED'.

	CLAIM_ID	FIRST_NOTICE...	RESPONSE	CLAIM_REASON	INCIDENT_SUMMARY
1	5450146950	2017-09-02	Yes	Hail	
2	3104086084	2016-10-06	No	Collision	
3	5729027190	2017-01-03	No	Other	
4	8180057168	2016-08-27	No	Collision	
5	2328025206	2016-07-09	No	Hail	
6	2326986146	2017-07-25	Yes	Hail	
7	807078258	2017-06-01	Yes	Collision	
8	8678029010	2016-03-20	No	Scratch/Dent	
9	9185344075	2017-01-14	No	Hail	
10	3261067196	2016-06-05	No	Collision	
11	9886704676	2016-02-16	No	Collision	
12	4702720486	2017-03-20	No	Collision	

SOURCE FILE: Auto Insurance Customers SAMPLE SIZE: First 143 rows

The join should complete successfully. **Scroll** to the right to see that the Auto Insurance Claims table columns are now appended at the end of the Auto Insurance Customers and Policies table in the shaper.

Notice that the Steps panel appears with the two Joins shaping operations.

24. Click the **Save** button on the toolbar to save the data flow.

Frequently saving a data flow is a good best practice and ensures you will not lose any of your work. Auto saving will be implemented in a future release.

## Rename and Remove Data

The screenshot shows the Data Flow interface with the 'Operation' button selected. The 'Operations' menu is open, showing 'FREQUENTLY USED' options: Calculate, Convert column type, Filter, Math, Remove, Rename (selected), Sort ascending, Sort descending, Substitute, and Text. The 'Steps' panel on the right shows the same two-step join process as the previous screenshot.

1. Click the **Operation** button to view the shaping operations menu.

2. Click the **Rename** operation.

To begin, select a column.

**CUSTOMER\_X**

CLAIM_ID	FIRST_NOTICE_DATE	RESPONSE	CLAIM_REASON
5450146950	2017-09-02	Yes	Hail
3104086084	2016-10-06	No	Collision
5729027190	2017-01-03	No	Other
8180057168	2016-08-27	No	Collision
2328025206	2016-07-09	No	Hail
2326986146	2017-07-25	Yes	Hail
807078258	2017-06-01	Yes	Collision
8678029010	2016-03-20	No	Scratch/Dent
9185344075	2017-01-14	No	Hail
3261067196	2016-06-05	No	Collision
9886704676	2016-02-16	No	Collision
4702720486	2017-03-20	No	Collision

SOURCE FILE: Auto Insurance Customers SAMPLE SIZE: First 143 rows

Cancel Next

3. Click in the column selection list area and start typing the letters **cust**. Select the **CUSTOMER.x** column from the column selection list.

4. Click the **Next** button.

Code an operation to cleanse and shape your data

Rename column

Selected column: CUSTOMER\_X

Rename the column.

New column name \*

**CUSTOMER**

CUSTOMER_X
AA71604
AA71604
AA71604
AB13432
AB13432
AB13432
AB60627
AB60627
AB60627
AE60813
AE60813
AE60813

SOURCE FILE: Auto Insurance Customers SAMPLE SIZE: First 143 rows

Cancel Apply

5. Enter a **New column name of CUSTOMER** in uppercase.

6. Click the **Apply** button.

Code an operation to cleanse and shape your data

Search operations

FREQUENTLY USED

- Calculate
- Convert column type
- Filter
- Math
- Remove
- Rename
- Sort ascending
- Sort descending
- Substitute
- Text

**CLEANSE**

CUSTOMER	NAME	COUNTRY	LATITUDE
AB72731	Adolph Skitch	US	41.75113981
AB72731	Adolph Skitch	US	41.75113981
AB72731	Adolph Skitch	US	41.75113981
AB96670	Wynnie Dunnett	US	39.2781
AB96670	Wynnie Dunnett	US	39.2781
AB96670	Wynnie Dunnett	US	39.2781
AC58002	Waylan Trelevan	US	47.76121
AC58002	Waylan Trelevan	US	47.76121
AC58002	Waylan Trelevan	US	47.76121
AH99727	Penny Duckhouse	US	33.88081187
AH99727	Penny Duckhouse	US	33.88081187

SOURCE FILE: Auto Insurance Customers SAMPLE SIZE: First 1000 rows

3 STEPS

- Auto Insurance Customers
- Join
  - inner-joined data from Auto Insurance Policies based on columns CUSTOMER,CUSTOMER
- Join
  - inner-joined data from Auto Insurance Claims based on columns POLICY\_ID,POLICY\_ID

Rename column

Renamed column CUSTOMER\_X to CUSTOMER

7. Click the **Operation** button to view the shaping operations menu.

8. Click the **Remove** operation.

SOURCE FILE: Auto Insurance Customers SAMPLE SIZE: First 1000 rows

9. Click in the column selection list area and start typing the letters **cust**. Select the **CUSTOMER.y** column from the list.

SOURCE FILE: Auto Insurance Customers SAMPLE SIZE: First 1000 rows

10. Click the **Next** button.

My Projects / Auto Insurance / Auto Insurance Data Flow

**X Operation** Code an operation to cleanse and shape your data

Change Column Selection

Selected column: CUSTOMER\_Y

Remove the column.

No input is needed for this operation.

CUSTOMER_Y	CUSTOMER_X
AA71604	
AA71604	
AA71604	
AB13432	
AB13432	
AB13432	
AB60627	
AB60627	
AB60627	
AE60813	
AE60813	
AE60813	

Cancel **Apply** SOURCE FILE: Auto Insurance Customers SAMPLE SIZE: First 143 rows

3 STEPS

Data Source  
Auto Insurance Customers

Join  
inner-joined data from Auto Insurance Policies based on columns CUSTOMER,CUSTOMER

Join  
inner-joined data from Auto Insurance Claims based on columns POLICY\_ID,POLICY\_ID

Rename column JUST ADDED  
Renamed column CUSTOMER\_X to

11. Click the **Apply** button.

12. Go to the toolbar and Click the **Save** button to save the data flow.

## Create New Data

**X Operation** 1 Code an operation to cleanse and shape your data

Search operations

Replace missing values

Replace substring

**ORGANIZE**

Aggregate

Concatenate 2

Conditional replace

Join

Sample

Split column

NATURAL LANGUAGE

Remove stop words

Tokenize

CUSTOMER	NAME	COUNTRY	LATITUDE
AB72731	Adolph Skitch	US	41.75113981
AB72731	Adolph Skitch	US	41.75113981
AB72731	Adolph Skitch	US	41.75113981
AB96670	Wynnie Dunnett	US	39.2781
AB96670	Wynnie Dunnett	US	39.2781
AB96670	Wynnie Dunnett	US	39.2781
AB96670	Wynnie Dunnett	US	39.2781
AC58002	Waylan Trelevan	US	47.76121
AC58002	Waylan Trelevan	US	47.76121
AC58002	Waylan Trelevan	US	47.76121
AH99727	Penny Duckhouse	US	33.88081187
AH99727	Penny Duckhouse	US	33.88081187

SOURCE FILE: Auto Insurance Customers SAMPLE SIZE: First 1000 rows

4 STEPS

Auto Insurance Customers

Join  
inner-joined data from Auto Insurance Policies based on columns CUSTOMER,CUSTOMER

Join  
inner-joined data from Auto Insurance Claims based on columns POLICY\_ID,POLICY\_ID

Rename column  
Renamed column CUSTOMER\_X to CUSTOMER

1. Click the **Operation** button to view the shaping operations menu.

2. Scroll down to the ORGANIZE section and Click the **Concatenate** operation.

3. Select the **STREET\_ADDRESS** column from the column selection list.

4. Click the **Next** button.

5. Click in the *Select Column* area and select the **CITY** column from the list of columns to concatenate.

6. Click in the *Select Column* area and select the **STATE\_CODE** column from the list of columns to concatenate.

**Operation** Code an operation to cleanse and shape your data

< Concatenate

Change Column Selection

Selected column: STREET\_ADDRESS

Select the columns to concatenate with STREET\_ADDRESS in the order you want.

1. STREET\_ADDRESS

2. CITY

3. STATE\_CODE ⊖

STATE ZIP\_CODE 7

EMAIL\_ADDRESS PHONE\_NUMBER

SOURCE FILE: Auto Insurance Customers SAMPLE SIZE: First 1000 rows

Cancel Apply

4 STEPS

Auto Insurance Customers

Join

inner-joined data from Auto Insurance Policies based on columns CUSTOMER,CUSTOMER

Join

inner-joined data from Auto Insurance Claims based on columns POLICY\_ID,POLICY\_ID

Rename column

Renamed column CUSTOMER\_x to CUSTOMER

7. Click in the *Select Column* area and select the **ZIP\_CODE** column from the list of columns to concatenate.

**Operation** Code an operation to cleanse and shape your data

< Concatenate

Change Column Selection

2. CITY

3. STATE\_CODE

4. ZIP\_CODE

Select column\*

Type in a single space 8

Separator

Select or enter a column separator

Name of the concatenated column

SOURCE FILE: Auto Insurance Customers SAMPLE SIZE: First 1000 rows

Cancel Apply

4 STEPS

Auto Insurance Customers

Join

inner-joined data from Auto Insurance Policies based on columns CUSTOMER,CUSTOMER

Join

inner-joined data from Auto Insurance Claims based on columns POLICY\_ID,POLICY\_ID

Rename column

Renamed column CUSTOMER\_x to CUSTOMER

8. Click in the Separator area and **type in a single space** character.

**Operation** Code an operation to cleanse and shape your data

< Concatenate

Change Column Selection

2. CITY

3. STATE\_CODE

4. ZIP\_CODE

Select column\*

Separator

Custom()

Select or enter a column separator

Name of the concatenated column\* ADDRESS 10

SOURCE FILE: Auto Insurance Customers SAMPLE SIZE: First 1000 rows

Cancel Apply

4 STEPS

Auto Insurance Customers

Join

inner-joined data from Auto Insurance Policies based on columns CUSTOMER,CUSTOMER

Join

inner-joined data from Auto Insurance Claims based on columns POLICY\_ID,POLICY\_ID

Rename column

Renamed column CUSTOMER\_x to CUSTOMER

**Note** - Before proceeding to the next step, to name the new concatenated column, you should see a **Custom( )** entry appear in the Separator field. This was added by Data Refinery because you typed a **space** in the Separator field.

9. Enter **ADDRESS** as the Name of the concatenated column.

10. Click the **Apply** button.

The screenshot shows the Data Refinery interface. On the left, there are two tables: 'Auto Insurance Customers' and 'Auto Insurance Policies'. The 'Auto Insurance Customers' table has columns: NAME (String), COUNTRY (String), LATITUDE (Decimal), and LONGITUDE (Decimal). The 'Auto Insurance Policies' table has a single column: ADDRESS (String). The right side of the interface shows the 'Steps' panel, which details the concatenation process:

- Step 1: 'Auto Insurance Customers' (inner-joined from Auto Insurance Policies based on columns CUSTOMER,CUSTOMER)
- Step 2: 'Join' (inner-joined from Auto Insurance Claims based on columns POLICY\_ID,POLICY\_ID)
- Step 3: 'Rename column' (Renamed column CUSTOMER\_x to CUSTOMER)
- Step 4: 'Address' (concatenated ADDRESS = STREET\_ADDRESS + CITY + STATE\_CODE + POSTAL\_CODE)

SOURCE FILE: Auto Insurance Customers SAMPLE SIZE: First 1000 rows

As a result of the concatenate operation, you will now see that you have a new column named **ADDRESS** that contains the concatenation of the columns **STREET\_ADDRESS**, **CITY**, **STATE\_CODE** and **POSTAL\_CODE** separated by a space.

**Note** - The **ADDRESS** column now appears to the left of the **STREET\_ADDRESS** column in the shaper result set. However, when the result set is created as a CSV file the **ADDRESS** column will be placed as the last column in the row.

## Anonymize Data

The screenshot shows the Data Refinery interface with the 'Anonymize Data' step selected. The left pane displays a table with columns: PHONE\_NUMBER, GENDER, NATIONAL\_ID, EDUCATION, EMPLOYMENT..., and MARITAL\_STAT... . The 'NATIONAL\_ID' column has a context menu open, with the 'Ellipses...' option highlighted. The right pane shows the 'Steps' panel with five steps: 'Data Source' (Auto Insurance Customers), 'Join' (inner-joined from Auto Insurance Policies based on columns CUSTOMER,CUSTOMER), 'Replace' (inner-replaced from Auto Insurance Claims based on columns POLICY\_ID,POLICY\_ID), 'Join' (inner-joined from Auto Insurance Claims based on columns POLICY\_ID,POLICY\_ID), and 'Rename column' (Renamed column CUSTOMER\_x to CUSTOMER).

SOURCE FILE: Auto Insurance Customers SAMPLE SIZE: First 1000 rows

Scroll to the right and locate the **NATIONAL\_ID** column.

1. Select the **Ellipses...** in the top right corner of the **NATIONAL\_ID** column to view the column action menu.

## 2. Select the **Substitute** menu item.

The screenshot shows the 'Data' tab selected in the top navigation bar. The main area displays a preview of a dataset with columns: PHONE\_NUMBER, GENDER, NATIONAL\_ID, EDUCATION, EMPLOYMENT\_STATUS, and MARITAL\_STATUS. The NATIONAL\_ID column contains sensitive SSN values. To the right, a 'Steps' panel is visible, showing a sequence of operations: 'Data Source' (Auto Insurance Customers), 'Join' (inner-joined data from Auto Insurance Policies based on columns CUSTOMER,CUSTOMER), and 'Rename column' (Renamed column CUSTOMER\_x to). The source file is listed as 'Auto Insurance Customers' and the sample size is 'First 1000 rows'.

The **NATIONAL\_ID** column contains a U.S. SSN, which is classified as sensitive information that business users should not have access to. The **Substitute** operation anonymizes the data and replaces the original value with a unique and consistent substituted value to protect the privacy of the information. This column was intentionally included in the join to demonstrate how this operation works. This is another way within IBM Cloud Pak for Data, combined with the data governance capabilities of Knowledge Catalog, to protect sensitive, confidential or personally identifiable information.

### Sort Data

The screenshot shows the 'Visualizations' tab selected in the top navigation bar. The main area displays a preview of a dataset with columns: CUSTOMER, NAME, COUNTRY, LATITUDE, LONGITUDE, and ADDRESS. A context menu is open over the CUSTOMER column, with the 'Sort ascending' option highlighted. Numbered callouts indicate: 1) the ellipsis in the top right corner of the CUSTOMER column, 2) the 'Sort ascending' menu item, 3) the Save button in the toolbar, and 4) the 'Steps' panel. The source file is listed as 'Auto Insurance Customers' and the sample size is 'First 1000 rows'. A note at the bottom left says 'Scroll all the way to the left!' with an arrow pointing to the CUSTOMER column.

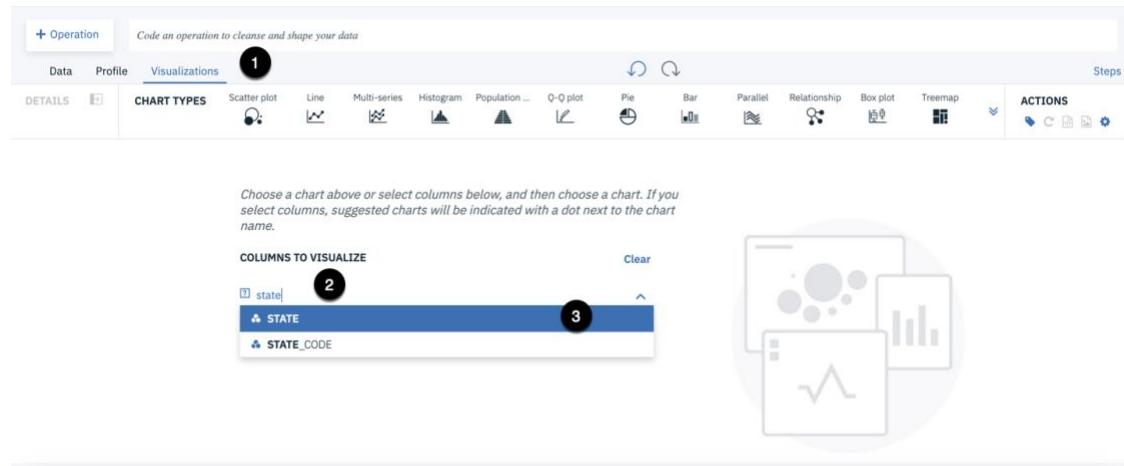
Scroll all way to the left to the **CUSTOMER** column.

1. Select the **ellipsis...** in the top right corner of the **CUSTOMER** column to view the column action menu.
2. Select the **Sort ascending** menu item.
3. Go to the toolbar and select the **Save** button.

4. Click the **Steps** button to hide the steps for more real estate to get ready for the next section.

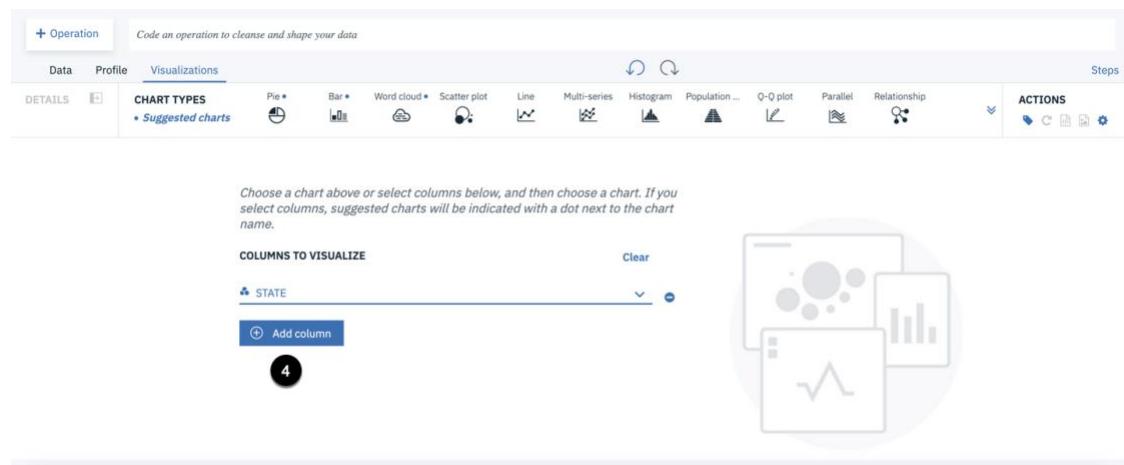
## Understand Data

The Data Refinery has built in **Visualization** to quickly and easily build charts and graphs to better understand data content before shaping and to validate shaping results.



The screenshot shows the Data Refinery interface with the 'Visualizations' tab selected. The 'CHART TYPES' section shows various chart types like Scatter plot, Line, Multi-series, Histogram, etc. Below it, the 'COLUMNS TO VISUALIZE' section has a text input field containing 'state'. A dropdown menu is open, showing 'STATE' and 'STATE\_CODE' as options. A large circular icon on the right displays three sample charts: a bubble chart, a bar chart, and a line chart.

1. Click the **Visualizations** tab.
2. Click in the **COLUMNS TO VISUALIZE** area and begin typing the letters **state**.
3. Select the **STATE** column from the list of columns.



The screenshot shows the Data Refinery interface with the 'Visualizations' tab selected. The 'CHART TYPES' section shows various chart types like Pie, Bar, Word cloud, Scatter plot, etc. Below it, the 'COLUMNS TO VISUALIZE' section has a text input field containing 'state'. A button labeled '+ Add column' is visible at the bottom of the input field. A large circular icon on the right displays three sample charts: a bubble chart, a bar chart, and a line chart.

4. Click the **Add Column** button.

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.

**COLUMNS TO VISUALIZE**

Clear

STATE

claims (5)

DENIED CLAIMS

CLAIMS FILED (6)

5. Click in the second line of the **COLUMNS TO VISUALIZE** area and begin typing the letters **claims**.

6. Select the **CLAIMS\_FILED** column from the list of columns.

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.

**COLUMNS TO VISUALIZE**

Clear

STATE

CLAIMS FILED (7)

Add column

7. Click on the **Bar** chart from the **CHART TYPES** toolbar.

Code an operation to cleanse and shape your data

**CHART TYPES**

Scatter plot • Line • Multi-series • Population ... • Pie • Bar • Box plot • Error bar • Dual Y-axes • Histogram • Q-Q plot

**ACTIONS**

Bar chart (12)

Category \*

STATE (8)

Order based on

Category name (9)

Category value (10)

Category order

As read (9)

Ascending (10)

Descending

Summary

Count (11)

Sum

Mean

STATE

COUNT

STATE	COUNT
California	163
New York	~150
Oregon	~120
Virginia	~100
Minnesota	~80
South Carolina	~60
Massachusetts	~40
Washington	~30
Mississippi	~20
Louisiana	~15
New Mexico	~10
Rhode Island	~5

A Bar chart is displayed showing the number of Claims Filed by State.

8. Click the **Category value** radio button in the **Order based on** section.

9. Click the **Descending** radio button in the **Category order** section.

10. Click the **Count** radio button in the **Summary** section.

Notice that **California** has the highest number of claims filed.

11. Hover over the very top of the **California** bar in the chart to see the number of claims filed.

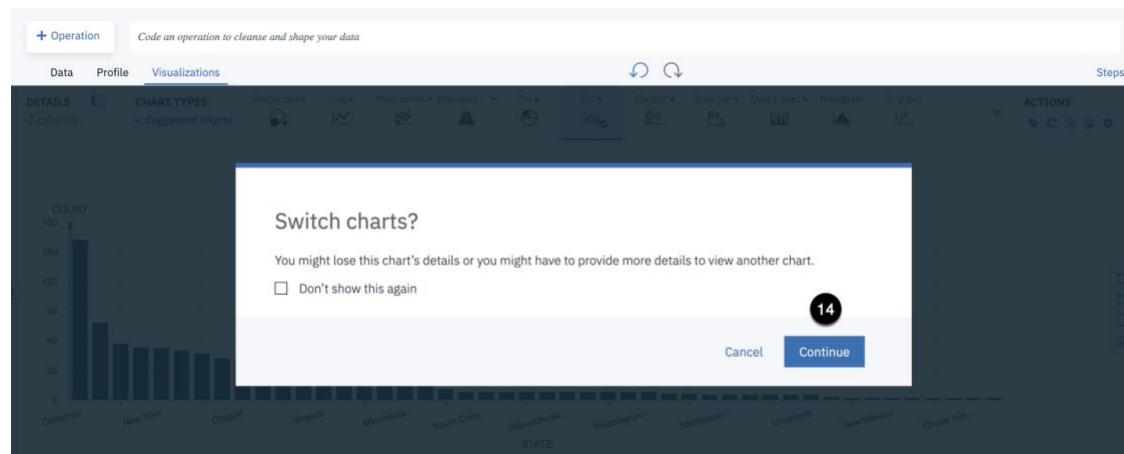
**Note** - The number of claims filed you see may be different than what is displayed on the screenshot because this is a multi-tenant environment with updates, deletes and inserts happening on a regular basis.

12. Click the **Details** section button to hide the Details section to gain more real estate.

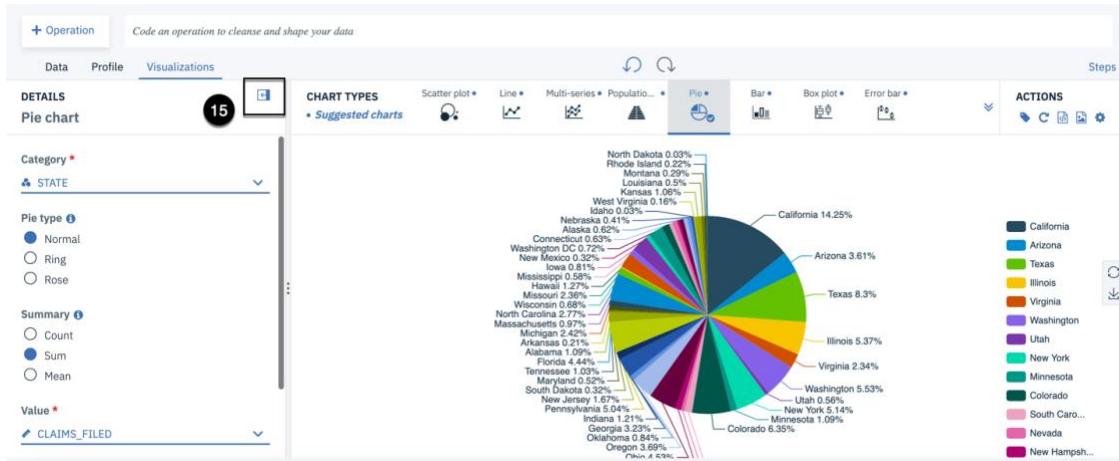


You can change the chart type, and keep the same properties, by selecting any of the suggested chart types (those that have a **dot** next to them).

13. Click on the **Pie** chart from the **CHART TYPES** toolbar.

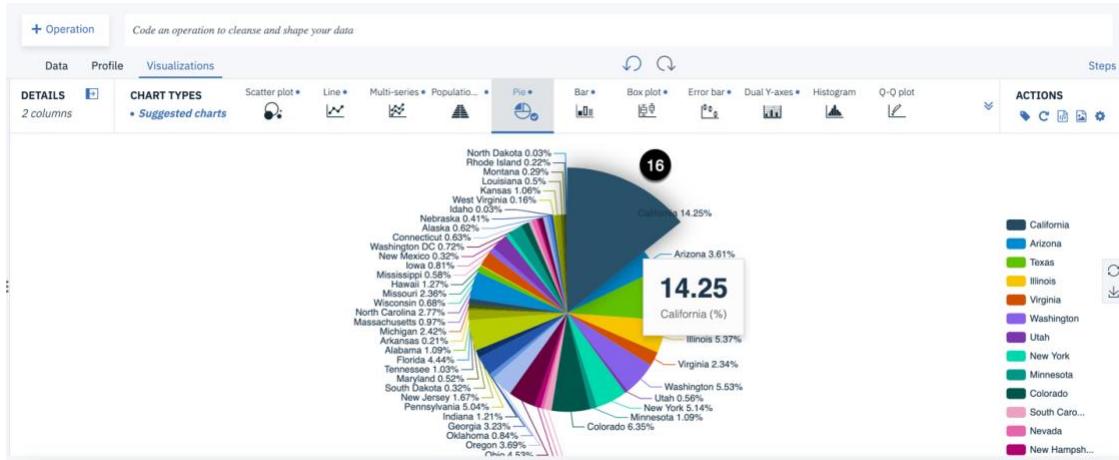


14. When the **Switch charts?** dialog appears, Click the **Continue** button.



The visualization is changed to a **Pie** chart to render the result.

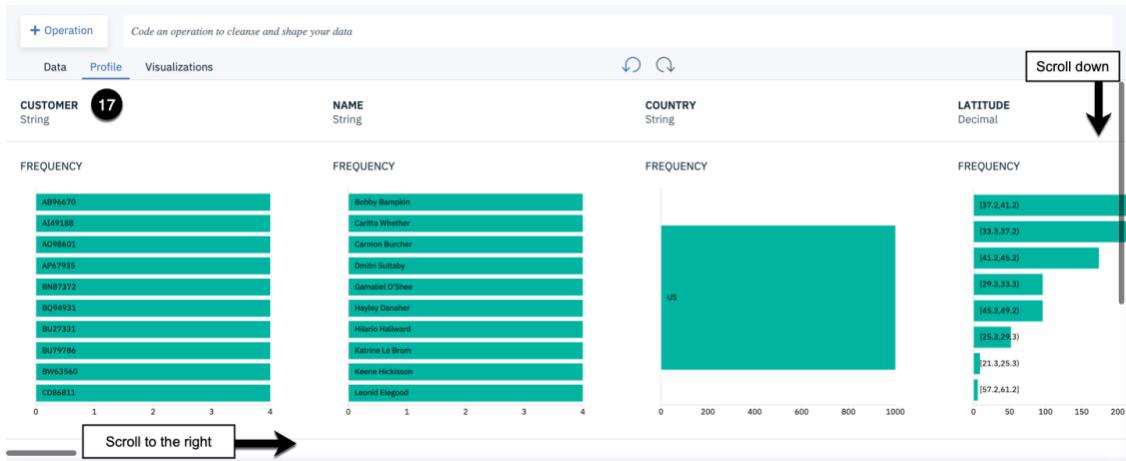
15. Click the **Details** section button to hide the Details section to gain more real estate.



16. Hover over the state of **California** slice in the pie chart to see the claims filed percentage compared to the other states. This further clarifies that the state of California is the leader in claims filed.

Feel free to experiment more with the **Visualizations** feature of Data Refinery on your own, using the data provided, or your own, to get more experience with the robust capabilities and see how it can assist in better understanding data before and after shaping.

Last, but surely not least, the data refinery has built in **Profiling** to allow for examination and exploration of data statistics and data type classifications to assist users in better understanding data content before and after data shaping.

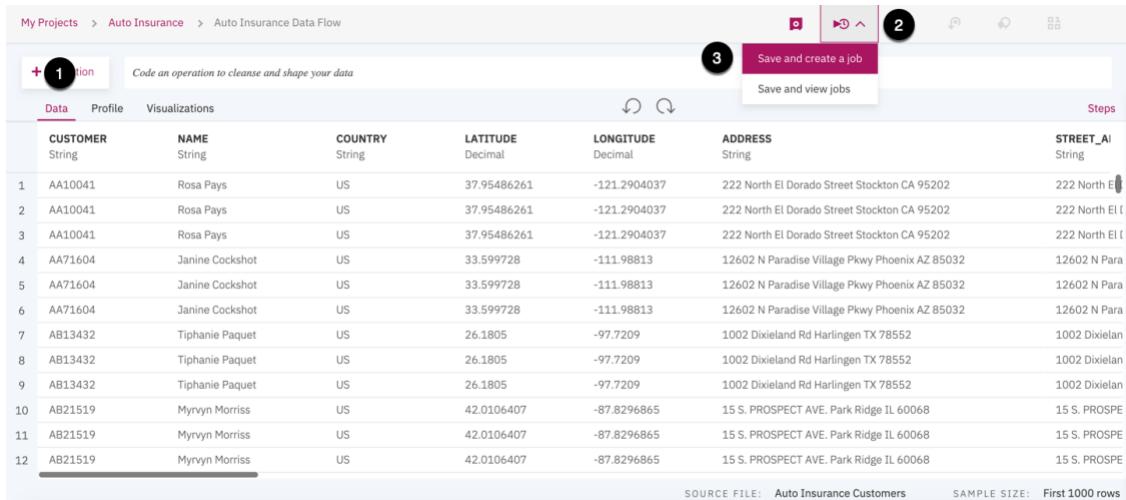


## 17. Click the **Profile** tab to view the profile.

**Scroll** to the right and down to view all the column statistics provided.

### Run the Data Flow

In order to process the shaping operations, you just performed, you need to create a **Job** and run it. The job will use the data flow's output data set name, target location and format type to place and create the data flow output. Based on the changes you specified, the job will create a CSV file named **Auto Insurance Shaped** in your **Auto Insurance** project.



## 1. Click on the **Data** tab to go back to the data view.

## 2. Click the **Jobs** button on the toolbar.

## 3. Select the **Save and create a job** menu item.

Create a job

Create a job to specify how and when to run an analytical asset. Select the analytic asset and set up a schedule or run the job immediately.

Job Name Auto Insurance Shaped	<b>4</b>	INPUT Auto Insurance Customers	OUTPUT Auto Insurance Shaped	CSV
Description (Optional) Prepare the auto insurance data for analytics	<b>5</b>	<input type="checkbox"/> Schedule to run		
Associated asset DATA REFINERY FLOW Auto Insurance Data Flow 7 Steps Edit	<b>6</b>			
Select runtime Default Data Refinery XS	<b>7</b>	Cancel Create Create and Run		

4. Enter a Job Name of **Auto Insurance Shaped** with the proper case, and spaces between the words.
5. Copy and paste, or enter, this bolded text **Prepare the auto insurance data for analytics** into the job Description field.
6. Use the **Default Data Refinery XS** runtime, it should be pre-selected.
7. Click the **Create and Run** button.

My Projects > Auto Insurance > Auto Insurance Shaped

Auto Insurance Shaped Prepare the auto insurance data for analytics	Associated Asset DATA REFINERY FLOW Auto Insurance Data Flow 7 Steps													
Scheduled to run No Schedule Created	Environment definition Default Data Refinery XS	INPUT Auto Insurance Customers	OUTPUT Auto Insurance Shaped	CSV										
<b>Runs</b> <table border="1"> <thead> <tr> <th>Start Time</th> <th>Status</th> <th>Duration</th> <th>Started By</th> <th>Action</th> </tr> </thead> <tbody> <tr> <td>Jan 10, 2020, 10:39:32 AM</td> <td>Starting</td> <td>---</td> <td>ctp</td> <td></td> </tr> </tbody> </table>					Start Time	Status	Duration	Started By	Action	Jan 10, 2020, 10:39:32 AM	Starting	---	ctp	
Start Time	Status	Duration	Started By	Action										
Jan 10, 2020, 10:39:32 AM	Starting	---	ctp											

The status will change from **Queued** to **Starting** to **Running** to **Completed**.

My Projects > Auto Insurance > Auto Insurance Shaped

**8**

### Auto Insurance Shaped

Prepare the auto insurance data for analytics

Scheduled to run	Edit	Environment definition	Edit	INPUT	OUTPUT	CSV
No Schedule Created		Default Data Refinery XS		Auto Insurance Customers	Auto Insurance Shaped	

**Associated Asset**  
DATA REFINERY FLOW  
**Auto Insurance Data Flow** 7 Steps

**Runs**

Start Time	Status	Duration	Started By	Action
Jan 10, 2020, 10:39:32 AM	Completed	43 seconds	ctp	

You can use your browser's refresh function to refresh the page to see the data flow status updates. Wait until the data flow status changes to **Completed** before proceeding to the next step. It should take a minute or less to finish.

8. Click on the **Auto Insurance** project navigation link on the toolbar to get back to the sections of the project.

My Projects > Auto Insurance

**9**

What are you looking for?

**Assets**

0 asset selected.

<input type="checkbox"/>	NAME	TYPE	CREATED BY	LAST MODIFIED	ACTIONS
<input type="checkbox"/>	Auto Insurance Shaped <b>10</b>	Data Asset	ctp	10 Jan 2020, 9:16:06 am	⋮
<input type="checkbox"/>	Auto Insurance Customers	Data Asset	ctp	10 Jan 2020, 8:36:06 am	
<input type="checkbox"/>	Auto Insurance Claims	Data Asset	ctp	10 Jan 2020, 8:36:06 am	
<input type="checkbox"/>	Auto Insurance Policies	Data Asset	ctp	10 Jan 2020, 8:36:06 am	
<input type="checkbox"/>	Db2 Warehouse	Connection	ctp	10 Jan 2020, 8:36:05 am	

**10**

**▼ Data Refinery flows** **New Data Refinery flow**

9. You should be taken to the **Assets** tab of the project. If not, click on the **Assets** tab to view the project assets.

**Notice** that you now have a new data asset named **Auto Insurance Shaped**. This is the CSV dataset the Data Refinery generated based on your shaping recipe.

10. Hover over the **Auto Insurance Shaped** data asset and click on it to preview the data and verify the data flow results.

Preview   Profile   Lineage

Schema: 54 Columns  
Preview: 1000 rows Last refresh: 35 minutes ago Refresh

**11**

Data Asset  
Auto Insurance Shaped

Description  
No description available for this asset

Tags  
No tags available for this asset

Added: 03:16 PM UTC, 2020/01/10  
Size: 549.154 KB

CUSTOMER	NAME	COUNTRY	LATITUDE	LONGITUDE	ADDRESS	STREET_ADD...	CITY	STATE
Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String
AA10041	Rosa Pays	US	37.95486261	-121.2904037	222 North El Dor	222 North El Dorado :	Stockton	California
AA10041	Rosa Pays	US	37.95486261	-121.2904037	222 North El Dor	222 North El Dorado :	Stockton	California
AA10041	Rosa Pays	US	37.95486261	-121.2904037	222 North El Dor	222 North El Dorado :	Stockton	California
AA71604	Janine Cockshot	US	33.599728	-111.98813	12602 N Paradis	12602 N Paradise Vil	Phoenix	Arizona
AA71604	Janine Cockshot	US	33.599728	-111.98813	12602 N Paradis	12602 N Paradise Vil	Phoenix	Arizona
AB13432	Tiphanie Paquet	US	26.1805	-97.7209	1002 Dixieland R	1002 Dixieland Rd	Harlingen	Texas
AB13432	Tiphanie Paquet	US	26.1805	-97.7209	1002 Dixieland R	1002 Dixieland Rd	Harlingen	Texas
AB13432	Tiphanie Paquet	US	26.1805	-97.7209	1002 Dixieland R	1002 Dixieland Rd	Harlingen	Texas
AB21519	Myrvyn Morris	US	42.0106407	-87.8296865	15 S. PROSPECT	15 S. PROSPECT AVE	Park Ridge	Illinois
AB21519	Myrvyn Morris	US	42.0106407	-87.8296865	15 S. PROSPECT	15 S. PROSPECT AVE	Park Ridge	Illinois

11. Click the X on the data asset info panel to close it and maximize the viewing real estate.

Preview   Profile   Lineage

Schema: 54 Columns  
Preview: 1000 rows Last refresh: 40 minutes ago Refresh

**12**

Scroll to the right

**Refine**

CUSTOMER	NAME	COUNTRY	LATITUDE	LONGITUDE	ADDRESS	STREET_ADD...	CITY	STATE	STATE_CO...	ZIP_CODE	EMAIL_
Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String
AA10041	Rosa Pays	US	37.95486261	-121.2904037	222 North El Dorado	222 North El Dorado :	Stockton	California	CA	95202	rpaysp8
AA10041	Rosa Pays	US	37.95486261	-121.2904037	222 North El Dorado	222 North El Dorado :	Stockton	California	CA	95202	rpaysp8
AA10041	Rosa Pays	US	37.95486261	-121.2904037	222 North El Dorado	222 North El Dorado :	Stockton	California	CA	95202	rpaysp8
AA71604	Janine Cockshot	US	33.599728	-111.98813	12602 N Paradise Vil	12602 N Paradise Vill	Phoenix	Arizona	AZ	85032	jcocksh
AA71604	Janine Cockshot	US	33.599728	-111.98813	12602 N Paradise Vil	12602 N Paradise Vill	Phoenix	Arizona	AZ	85032	jcocksh
AA71604	Janine Cockshot	US	33.599728	-111.98813	12602 N Paradise Vil	12602 N Paradise Vill	Phoenix	Arizona	AZ	85032	jcocksh
AB13432	Tiphanie Paquet	US	26.1805	-97.7209	1002 Dixieland R	1002 Dixieland Rd	Harlingen	Texas	TX	78552	tpaquet
AB13432	Tiphanie Paquet	US	26.1805	-97.7209	1002 Dixieland R	1002 Dixieland Rd	Harlingen	Texas	TX	78552	tpaquet
AB13432	Tiphanie Paquet	US	26.1805	-97.7209	1002 Dixieland R	1002 Dixieland Rd	Harlingen	Texas	TX	78552	tpaquet
AB21519	Myrvyn Morris	US	42.0106407	-87.8296865	15 S. PROSPECT AVE	15 S. PROSPECT AVE	Park Ridge	Illinois	IL	60068	mmorris
AB21519	Myrvyn Morris	US	42.0106407	-87.8296865	15 S. PROSPECT AVE	15 S. PROSPECT AVE	Park Ridge	Illinois	IL	60068	mmorris

At first glance, it looks like you have duplicate rows, but that is not the case. There can be multiple claims per policy and multiple policies per customer. You did an inner join, so the customer and policy data is repeated for every claim per policy per customer. Each claim has a unique **CLAIM\_ID**.

- Notice the new name for the **CUSTOMER** column and that the data is sorted by **CUSTOMER**.
- Notice the new **ADDRESS** column. Place your cursor near the end of the column until you see a solid bar and drag it open to see the full content.

12. Scroll to the right.

Schema: 54 Columns

Preview: 1000 rows Last refresh: 51 minutes ago Refine

EMAIL_ADDR...	PHONE_NUM...	GENDER	NATIONAL_ID	EDUCATI...	EMPLOYMENT_STA...	MARITAL_STA...	CUSTOMER_LIFETIME_VA...	NUMBER_OF_POLI...
Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String	Type: String
rpaysp8@homestea...	865-749-5448	Female	73724a8a5e731018fb3c5f958	Bachelor	Employed	Married	9421.101961	3
rpaysp8@homestea...	865-749-5448	Female	73724a8a5e731018fb3c5f958	Bachelor	Employed	Married	9421.101961	3
rpaysp8@homestea...	865-749-5448	Female	73724a8a5e731018fb3c5f958	Bachelor	Employed	Married	9421.101961	3
jcockshotqc@wikim...	808-976-1894	Female	d0642f60ceacb3fe75483cc4e1	Master	Employed	Married	2802.621642	1
jcockshotqc@wikim...	808-976-1894	Female	d0642f60ceacb3fe75483cc4e1	Master	Employed	Married	2802.621642	1
jcockshotqc@wikim...	808-976-1894	Female	d0642f60ceacb3fe75483cc4e1	Master	Employed	Married	2802.621642	1
tpaquet54@gmpg.or...	612-256-1393	Female	47a2bc15177610791ea72cde1	Bachelor	Unemployed	Single	10628.06415	3
tpaquet54@gmpg.or...	612-256-1393	Female	47a2bc15177610791ea72cde1	Bachelor	Unemployed	Single	10628.06415	3
tpaquet54@gmpg.or...	612-256-1393	Female	47a2bc15177610791ea72cde1	Bachelor	Unemployed	Single	10628.06415	3
mmorissbm@word...	203-751-1286	Male	21959d6efec5b293355a9905b	College	Employed	Married	2705.987629	1
mmorissbm@word...	203-751-1286	Male	21959d6efec5b293355a9905b	College	Employed	Married	2705.987629	1

- Notice that the **NATIONAL\_ID** column was anonymized and that the **CUSTOMER.y** column has been removed.

13. Click on the **Auto Insurance** project navigation link on the toolbar to go back to the project home page.

What are you looking for?

**14**

**15** Scroll down

**New data asset**

**New Data Refinery flow**

NAME	TYPE	CREATED BY	LAST MODIFIED	ACTIONS
Auto Insurance Shaped	Data Asset	ctp	10 Jan 2020, 9:16:06 am	
Auto Insurance Customers	Data Asset	ctp	10 Jan 2020, 8:36:06 am	
Auto Insurance Claims	Data Asset	ctp	10 Jan 2020, 8:36:06 am	
Auto Insurance Policies	Data Asset	ctp	10 Jan 2020, 8:36:06 am	
Db2 Warehouse	Connection	ctp	10 Jan 2020, 8:36:05 am	

**14**

**15** Scroll down

**New data asset**

**New Data Refinery flow**

**14**

**15** Scroll down

**New data asset**

**New Data Refinery flow**

14. You should be taken to the **Assets** tab of the project. If not, click on the **Assets** tab to view the project assets.

15. Scroll down to the **Data Refinery flows** section of the Assets tab.

	NAME	TYPE	CREATED BY	LAST MODIFIED	ACTIONS
<input type="checkbox"/>	Auto Insurance Shaped	Data Asset	ctp	10 Jan 2020, 9:16:06 am	
<input type="checkbox"/>	Auto Insurance Customers	Data Asset	ctp	10 Jan 2020, 8:36:06 am	
<input type="checkbox"/>	Auto Insurance Claims	Data Asset	ctp	10 Jan 2020, 8:36:06 am	
<input type="checkbox"/>	Auto Insurance Policies	Data Asset	ctp	10 Jan 2020, 8:36:06 am	
<input type="checkbox"/>	Db2 Warehouse	Connection	ctp	10 Jan 2020, 8:36:05 am	

**New Data Refinery flow**

NAME	TYPE	CREATED BY	LAST MODIFIED	ACTIONS
Auto Insurance Data Flow	Data Refinery flow	ctp	10 Jan 2020, 9:14:39 am	<span style="color: #0070C0;">...</span> <b>16</b> <ul style="list-style-type: none"> <li><a href="#">Clone</a></li> <li><a href="#">Create job</a></li> <li><a href="#">View job</a></li> <li><a href="#">Remove</a></li> </ul>

Once a data flow is saved, it is placed in the **Data Refinery flows** section of the project. You should see your saved **Auto Insurance Data Flow**.

16. Select the **ellipses...** to the far right under the data flow **ACTIONS** column.

**Note** - You may have to scroll down in the UI to see the entire menu depending on the browser you are using.

You can perform the following actions from the data flow actions menu:

- **Clone** - Creates a copy of the data flow. The flow is added to the Data Refinery flows list as “original-name copy 1”.
- **Create job** - Opens the job creation dialog to create a new job.
- **View job** - Opens the job page for data flow.
- **Remove** - Deletes the data flow from your project.

You can also click on the **data flow** to open it in the data flow shaper to modify it and view the data flow's result set and recipe steps.

## Protect Sensitive Information

In this section you will learn how to protect sensitive information by creating **Data Protection Rules**. You will create data protection rules to obfuscate (i.e. Mask) **US Social Security Numbers** and redact **Credit Card Information** and then validate that they are being enforced.

## Create Data Protection Rules

The screenshot shows the top navigation bar of the IBM Cloud Pak for Data interface. It includes the 'IBM Cloud Pak for Data' logo, a search bar, and various project management icons. Below the navigation bar, there's a breadcrumb trail 'My Projects > Auto Insurance'. The main menu has tabs for 'Overview', 'Assets' (which is selected), 'Environments', 'Jobs', 'Access Control', and 'Settings'. A search bar at the bottom allows users to look for specific assets.

Click the **IBM Cloud Pak for Data** navigation menu in the top left corner.

This screenshot shows the 'Organize' section of the navigation menu. It includes categories like 'All catalogs', 'Information assets', 'Data and AI governance', 'Categories', 'Business terms', 'Classifications', 'Data classes', 'Reference data', 'Policies', and 'Rules'. The 'Rules' option is highlighted with a red box and a number '1' below it. To the right, a preview window shows a table of data assets with columns for TYPE, CREATED BY, LAST MODIFIED, and ACTIONS. The table contains three entries: 'Shaped' (Data Asset, ctp, 10 Jan 2020, 9:16:06 am), 'Customers' (Data Asset, ctp, 10 Jan 2020, 8:36:06 am), and 'Claims' (Data Asset, ctp, 10 Jan 2020, 8:36:06 am).

1. Click **Organize > Data and AI Governance > Rules** from the menu.

This screenshot shows the 'Rules' page. At the top, there are tabs for 'Published' (selected) and 'Draft'. On the right, there's a 'Create rule' button with a number '2' below it. Below the tabs, there are search and filter options: 'Find rules', 'Sort by Name', and 'Show All rule types'. The main area displays a table of rules, with one row highlighted.

2. Click the **Create rule** button.

This screenshot shows the 'New rule' creation page. At the top, there's a 'Create rule' button with a number '2' above it. Below it, there's a 'Find rules' input field and a 'Sort by Name' dropdown. The main area is titled 'New rule' and asks 'Select the type of rule to create.' It shows two options: 'Data protection rule' (selected, with a number '3' next to its description) and 'Governance rule'. The 'Data protection rule' description states: 'A rule to mask sensitive data values or to deny access to data assets.'

3. Click the **Data protection rule** type.

The screenshot shows the 'Rule builder' interface with the following details:

- Details:**
  - Name\***: Protect Credit Card Information (marked with a circled '4')
  - Type\***: Access
  - Business definition\***: Protect all credit card numbers, expiration dates and validation numbers using the data redaction method. (marked with a circled '5')
- Rule builder - Criteria \*** (marked with a circled '1'):
  - CONDITION 1** (marked with a circled '6'):
    - If: Data class (marked with a circled '7')
    - contains any (marked with a circled '8'): Credit Card Number, Credit Card Number.
- Action \*** (marked with a circled '9'):
  - then: deny access to data

4. Enter a Name of **Protect Credit Card Information**.
5. Copy and Paste the following bolded text into the Business definition: **Protect all credit card numbers, expiration dates and validation numbers using the data redaction method**
6. In the Condition 1 area, for the **If** statement, select **Data Class**.
7. In the *Search for a data class* area type **credit card number**
8. Select the **Credit Card Number** data class from the list.

The screenshot shows the 'Rule builder' interface with the following details:

- Details:**
  - Name\***: Protect Credit Card Information
  - Type\***: Access
  - Business definition\***: Protect all credit card numbers, expiration dates and validation numbers using the data redaction method.
- Rule builder - Criteria \*** (marked with a circled '1'):
  - CONDITION 1** (marked with a circled '9'):
    - If: Data class (marked with a circled '10'): Credit Card Number, credit card expir
- Action \*** (marked with a circled '11'):
  - then: deny access to data

9. In the *Search for a data class* area type **credit card expir**
10. Select the **Credit Card Expiration Date** data class from the list.

**Details**

**Name\*** Protect Credit Card Information

**Type\*** Access

**Business definition\*** Protect all credit card numbers, expiration dates and validation numbers using the data reduction method.

**Rule builder**

**Criteria \***

**CONDITION 1**

If Data class contains any Credit Card Number X Credit Card Expiration Date X credit card valid

11 Credit Card Validation Number A 3 or 4 digits number representing a credit card validation number.

12

**Action \***

then deny access to data

11. In the *Search for a data class* area type **credit card valid**

12. Select the **Credit Card Validation Number** data class from the list.

**Details**

**Name\*** Protect Credit Card Information

**Type\*** Access

**Business definition\*** Protect all credit card numbers, expiration dates and validation numbers using the data reduction method.

**Rule builder**

**Criteria \***

**CONDITION 1**

If Data class contains any Credit Card Number X Credit Card Expiration Date X Credit Card Validation Number X

+ Add condition

**Action \***

then deny access to data  
mask data

13

13. In the Action area, for the **then** clause, select **mask data**.

**Details**

**Name\*** Protect Credit Card Information

**Type\*** Access

**Business definition\*** Protect all credit card numbers, expiration dates and validation numbers using the data reduction method.

**Rule builder**

**Criteria \***

**CONDITION 1**

If Data class contains any Credit Card Number X Credit Card Expiration Date X Credit Card Validation Number X

+ Add condition

**Action \***

then mask data where in columns containing

14

14. In the Action area, for the **where** clause, select **in columns containing**.

The screenshot shows the 'New data protection rule' configuration page. In the 'Action' section, under 'Search for a data class', the user has typed 'credit card number'. A dropdown menu lists several options, with 'Credit Card Number' highlighted in red and circled in black as step 16.

15. In the *Search for a data class* area type **credit card number**

16. Select the **Credit Card Number** data class from the list.

The screenshot shows the 'New data protection rule' configuration page. In the 'Action' section, under 'Search for a data class', the user has typed 'credit card expr'. A dropdown menu lists several options, with 'Credit Card Expiration Date' highlighted in red and circled in black as step 18.

17. In the *Search for a data class* area type **credit card expr**

18. Select the **Credit Card Expiration Date** data class from the list.

The screenshot shows the 'New data protection rule' configuration page. In the 'Action' section, under 'Search for a data class', the user has typed 'credit card valid'. A dropdown menu lists several options, with 'Credit Card Validation Number' highlighted in red and circled in black as step 20.

19. In the *Search for a data class* area type **credit card valid**

20. Select the **Credit Card Validation Number** data class from the list.

**Details**

**Name\*** Protect Credit Card Information

**Type\*** Access

**Business definition\*** Protect all credit card numbers, expiration dates and validation numbers using the data redaction method.

**Action \***

then mask data in columns containing

Credit Card Number X Credit Card Expiration Date X Credit Card Validation Number X

Select how to mask data:

- Redact** ✓ Before 452-821-1120 Replace data with Xs
- Substitute** Before 452-821-1120 Replace data with values that don't match the original format
- Obfuscate** Before 452-821-1120 Replace data with similarly formatted values

**21**

**Create**

## 21. Click **Create**.

**Rules** > Protect Credit Card Information

**22** | Protect Credit Card Information

**Business definition**  
Protect all credit card numbers, expiration dates and validation numbers using the data redaction method.

**Creator** ctp  
**Date created** 1/11/2020  
**Last editor** ctp  
**Last modified** 1/11/2020

**Criteria**

**Condition 1**  
If Data class contains any Credit Card Number, Credit Card Expiration Date, Credit Card Validation Number

**Action**  
Then Redact data in columns containing:  
Credit Card Number, Credit Card Expiration Date, Credit Card Validation Number

**23**

## 22. Click the **Rules** bread crumb to go back to the rules section.

Rules

Extract Import Create rule

Published Draft

Find rules

Sort by Name Show All rule types

**23**

## 23. Click the **Create rule** button.

Rules > New rule

New rule

Select the type of rule to create.

**Data protection rule** (?)  
A rule to mask sensitive data values or to deny access to data assets.

**Governance rule**  
A rule to describe the criteria for compliance with business objectives.

24. Click the **Data protection rule** type.

Rules > New data protection rule

Details	Rule builder
<p>Name* <b>Protect US Social Security Numbers</b> <small>(25)</small></p> <p>Type* <b>Access</b></p> <p>Business definition* <small>(26)</small> Protect all US Social Security Numbers using the data masking obfuscation method replacing data with similarly formatted but fictional values.</p>	<p>Criteria * <small>(1)</small></p> <p><b>CONDITION 1</b> <small>(27)</small></p> <p>If Data class <small>(28)</small> contains any <small>(29)</small></p> <p><b>US Social Security Number</b> <small>(30)</small> In the United States, a Social Security number (SSN) is a unique nine-digit number issued to U.S. citizens, permanent residents, and temporary (working) residents.</p> <p>Action <b>US Social Security Number Last 4</b> The last four digits of a United States Social Security Number then <b>(SSN)</b> deny access to data</p>

25. Enter a Name of **Protect US Social Security Numbers**.

26. Copy and Paste the following bolded text into the Business definition: **Protect all US Social Security Numbers using the data masking obfuscation method replacing data with similarly formatted but fictional values**

27. In the Condition 1 area, for the If statement, select **Data Class**.

28. In the *Search for a data class* area type **us social**

29. Select the **US Social Security Number** data class from the list.

Rules > New data protection rule

**Details**

**Name\***  
Protect US Social Security Numbers

**Type\***  
Access

**Business definition\***  
Protect all US Social Security Numbers using the data masking obfuscation method replacing data with similarly formatted but fictional values.

**Criteria \***

**CONDITION 1**

If Data class contains any  
US Social Security Number X

**Action \***

then deny access to data  
deny access to data  
mask data **30**

30. In the Action area, for the **then** clause, select **mask data**.

Rules > New data protection rule

**Details**

**Name\***  
Protect US Social Security Numbers

**Type\***  
Access

**Business definition\***  
Protect all US Social Security Numbers using the data masking obfuscation method replacing data with similarly formatted but fictional values.

**Criteria \***

**CONDITION 1**

If Data class contains any  
US Social Security Number X

**Action \***

then mask data where **in columns containing** **31**

31. In the Action area, for the **where** clause, select **in columns containing**.

Rules > New data protection rule

**Details**

**Name\***  
Protect US Social Security Numbers

**Type\***  
Access

**Business definition\***  
Protect all US Social Security Numbers using the data masking obfuscation method replacing data with similarly formatted but fictional values.

**Action \***

then mask data where **in columns containing** **32**

**us social**

**US Social Security Number**  
In the United States, a Social Security number (SSN) is a unique nine-digit number issued to U.S. citizens, permanent residents, and temporary (working) residents.

**US Social Security Number Last 4**  
The last four digits of a United States Social Security Number (SSN).  
452-821-1120      452-821-1120      452-821-1120

Replace data with Xs	Replace data with values that don't match the original format	Replace data with similarly formatted values
----------------------	---	--

**33**

32. In the *Search for a data class* area type **us social**

33. Select the **US Social Security Number** data class from the list.

**Details**

**Name\***  
Protect US Social Security Numbers

**Type\***  
Access

**Business definition\*** ⓘ  
Protect all US Social Security Numbers using the data masking obfuscation method replacing data with similarly formatted but fictional values.

**Action \*** ⓘ

then **mask data** in columns containing **US Social Security Number X**

Select how to mask data:

- Redact** Before 452-821-1120 Replace data with Xs
- Substitute** Before 452-821-1120 Replace data with values that don't match the original format
- Obfuscate** Before 452-821-1120 After 008-219-6240 Replace data with similarly formatted values

34

### 34. Click the **Obfuscate** masking method.

**Details**

**Name\***  
Protect US Social Security Numbers

**Type\***  
Access

**Business definition\*** ⓘ  
Protect all US Social Security Numbers using the data masking obfuscation method replacing data with similarly formatted but fictional values.

**Action \*** ⓘ

then **mask data** in columns containing **US Social Security Number X**

Select how to mask data:

- Redact** Before 452-821-1120 Replace data with Xs
- Substitute** Before 452-821-1120 Replace data with values that don't match the original format
- Obfuscate** Before 452-821-1120 After 008-219-6240 Replace data with similarly formatted values

35

**Create**

### 35. Click **Create**.

**Rules** > Protect US Social Security Num...

**36** Rules **Protect US Social Security Numbers** Edit Delete

**Business definition**  
Protect all US Social Security Numbers using the data masking obfuscation method replacing data with similarly

**Creator** ctp **Date created** 1/11/2020 **Last editor** ctp **Last modified** 1/11/2020

**Criteria**

**Condition 1**

If Data class contains any  
US Social Security Number

**Action**

Then Obfuscate data in columns containing:  
US Social Security Number

### 36. Click the **Rules** breadcrumb from the menu.

The screenshot shows the 'Rules' section of the IBM Cloud Pak for Data interface. At the top, there are tabs for 'Published' and 'Draft', with 'Published' selected. Below the tabs are search and filter options: 'Find rules' with a magnifying glass icon, 'Sort by Name', and 'Show All rule types'. The main content area displays two data protection rules:

- Protect Credit Card Information** (active): Protect all credit card numbers, expiration dates and validation numbers using the data redaction method. Last modified: Jan 10, 2020.
- Protect US Social Security Numbers** (active): Protect all US Social Security Numbers using the data masking obfuscation method replacing data with similarly formatted but fictional values. Last modified: Jan 11, 2020.

At the bottom left, it says 'Showing 2 of 2 rules'. On the right side, there is a vertical sidebar with a user profile picture labeled 'ctp', 'Profile and settings', and links to 'Getting Started', 'About', 'Community', and 'Support'. At the bottom of the sidebar is a 'Log out' button with a counter '38'.

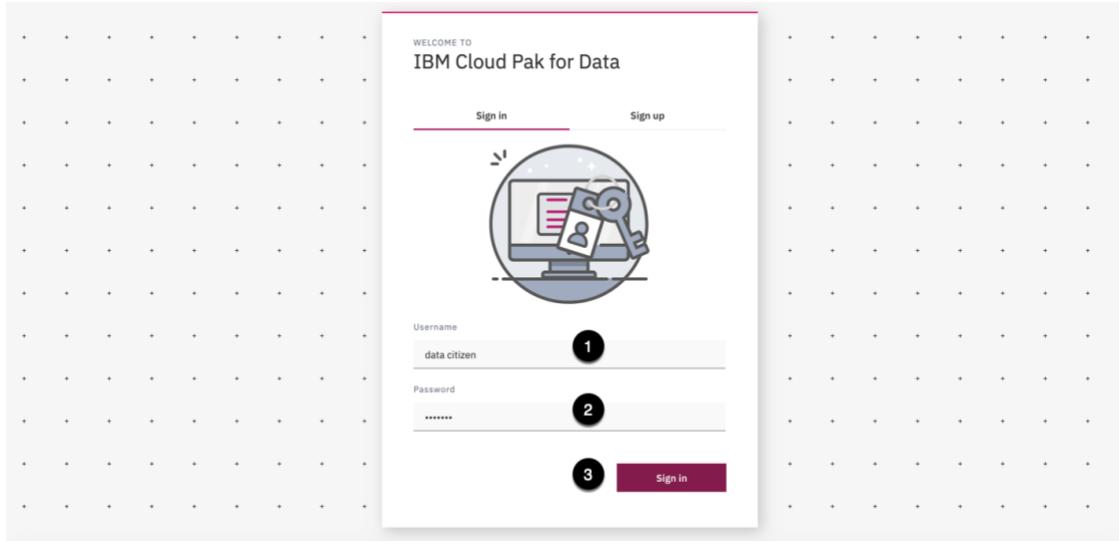
You should see the two data protection rules in the published tab. If not, refresh the page using your browser's refresh method.

37. Click on the **Profile and settings** icon in the top right corner.

38. Click **Log out**.

### Validate Data Protection Rules

You will now log in as a data citizen to validate that you can search and find the data you are looking for and that the data protection rules are being enforced as defined.



1. Enter **data citizen** as the **Username**.
2. Enter **citizen** as the **Password**.
3. Click the **Sign in** button.

WELCOME, data citizen!

## Let's get started!

Use these resources to make the most of your IBM Cloud Pak for Data experience.

Collect and organize      Analyze

IBM Cloud Pak for Data: Collect and organize

Build your enterprise data catalog and ensure that your data is mapped to a standard set of business terms and follows information governance policies and rules.

[Explore catalogs](#)

[Explore information assets](#)

Go to your home page

Hide on log in

#### 4. Enter **auto insurance** in the search area and press the enter key.

<input type="checkbox"/> Name	Type	Tags	Modified by	Date modified
<input type="checkbox"/> 2017 J.D. Power U.S. Auto Claims Satisfaction Study	Data asset	Auto ... Docu...	Unavailable	Jan 09, 2020
<input type="checkbox"/> All catalogs > Auto Insurance Auto Insurance document				
<input type="checkbox"/> Auto Insurance Policies	Data asset	Auto ...	Unavailable	Jan 09, 2020
<input type="checkbox"/> All catalogs > Auto Insurance All U.S. auto insurance policies				
<input type="checkbox"/> Vehicle Insurance Doc United States	Data asset	Auto ... Docu...	Unavailable	Jan 09, 2020
<input type="checkbox"/> All catalogs > Auto Insurance Auto Insurance document				
<input type="checkbox"/> Auto Insurance Claims	Data asset	Auto ...	Unavailable	Jan 09, 2020
<input type="checkbox"/> All catalogs > Auto Insurance All U.S. auto insurance claims				
<input type="checkbox"/> Auto Insurance Customers	<b>5</b> Data asset	Auto ...	ctp	Jan 09, 2020
<input type="checkbox"/> All catalogs > Auto Insurance All U.S. auto insurance customers				
<input type="checkbox"/> Db2 Warehouse	Connection	Auto ... Ware...	ctp	Jan 09, 2020
<input type="checkbox"/> Knowledge Catalog Tutorial Db2 Warehouse				
<input type="checkbox"/> Coordination of benefits priority count	Business term		admin	Dec 05, 2019
<input type="checkbox"/> Categories > udmh				

#### 5. Click on the **Auto Insurance Customers** data asset from the Auto Insurance catalog.

Search Results > Catalogs > Auto Insurance > Auto Insurance Customers

Overview Access Review Profile Lineage

DATA ASSET

### Auto Insurance Customers

Remove Download Add to Project

Description	All U.S. auto insurance customers	Schema:	28 Columns 328 Rows 6 4 Columns masked		
Added:	Jan 09, 2020 6:20 PM..PM	CUSTOMER	Type: String		
Format:	application/octet-stream	NAME	Type: String		
Size:	312 KB	COUNTRY	Type: String		
		LATITUDE	Type: Decimal		
		LONGITUDE	Type: Decimal		
		STREET_ADD...	Type: String		
		CITY	Type: String		
		STATE	Type: String		
		STAT	Type:		
Business Terms	There are no terms available for this asset.				
Tags	Auto Insurance				
Reviews	★★★★★ 1 review				
Classification	None				

**Data masking in progress**  
This asset is being masked by the data enforcement rule: Protect Credit Card Information. You can wait here to see a preview of the asset or we can notify you when the preview is ready.

Notify Me

You should immediately see the message, “**Data Masking in progress**”, with a spinning progress wheel. It will take a minute to load so be patient and let it finish.

**Note** - The **data citizen** user does not own the data asset so the data protection rules will be enforced and only see the protected version of the data as defined by the data protection rules defined that are based on the data classes of the data. That is why you did the additional work to classify the additional credit card expiration date and validation columns in the data profile of the **Auto Insurance Customers** table.

IBM Cloud Pak for Data All auto insurance

Search Results > Catalogs > Auto Insurance > Auto Insurance Customers

Overview Access Review Profile Lineage

DATA ASSET

### Auto Insurance Customers

Remove Download Add to Project

Description	All U.S. auto insurance customers	Schema:	28 Columns 328 Rows 6 4 Columns masked		
Added:	Jan 09, 2020 6:20 PM..PM	CUSTOMER	Type: String		
Format:	application/octet-stream	NAME	Type: String		
Size:	312 KB	COUNTRY	Type: String		
		LATITUDE	Type: Decimal		
		LONGITUDE	Type: Decimal		
		STREET_ADD...	Type: String		
		CITY	Type: String		
		STATE	Type: String		
		STAT	Type:		
Business Terms	There are no terms available for this asset.				
Tags	Auto Insurance				
Reviews	★★★★★ 1 review				
Classification	None				

An error occurred attempting to preview this asset.  
This file doesn't have any content. It might be corrupted.

If you see the error above, don’t be alarmed. It’s a known timing issue that is being addressed; the page just needs to be refreshed.

6. Click the **Refresh** button under the lock icon in the middle of the page. If that does not work, **refresh** the page using your browser’s refresh method.

Search Results > Catalogs > Auto Insurance > Auto Insurance Customers

Overview Access Review Profile Lineage

DATA ASSET

### Auto Insurance Customers

Description All U.S. auto insurance customers

Added: Jan 09, 2020 6:20 PM..PM Format: application/octet-stream Size: 312 KB

Business Terms There are no terms available for this asset.

Tags Auto Insurance

Reviews ★★★★☆ 1 review

Classification None

Schema: 28 Columns 328 Rows ⑧ 4 Columns masked ⓘ Preview: 328 rows Last refresh: 1 hour ago ⓘ Refresh

⑦ 1 column is obfuscated 3 columns are redacted Learn more

CUSTOMER ⓘ	NAME ⓘ	COUNTRY ⓘ	LATITUDE ⓘ	T_ADD... ⓘ	CITY ⓘ	STATE ⓘ	STAT ⓘ
AB72731	Adolph Skitch	US	41.75113981	-88.0127658	1001 W 75th Street	Woodridge	Illinois IL
AB96670	Wynnie Dunnnett	US	39.2781	-120.1203	100 Northstar Dr	Truckee	California CA
AC58002	Waylan Treleven	US	47.76121	-122.3464	18325 Aurora Ave N	Shoreline	Washington WA
AH99727	Penny Duckhouse	US	33.88081187	-118.0288381	16610 Valley View Av	La Mirada	California CA
AN36757	Malory Gantz	US	34.02091598	-84.31698227	1530 Old Alabama Rc	Roswell	Georgia GA
AP63665	Sianna Kissell	US	41.77025751	-88.20481022	1951 W Jefferson Av	Naperville	Illinois IL
AR47849	Janine McCreath	US	40.9581	-74.0747	177 Route 17 South	Paramus	New Jersey NJ
AS97690	Milicent Caveau	US	40.8591	-73.9694	136-38 Linwood Plaza	Fort Lee	New Jersey NJ

7. Hover over the **i** information icon next to the “4 columns masked” label with a lock icon in front of it.

**Notice** that 3 columns are redacted, and one is obfuscated.

8. Scroll to the right until you see the **NATIONAL\_ID** column.

Search Results > Catalogs > Auto Insurance > Auto Insurance Customers

Overview Access Review Profile Lineage

DATA ASSET

### Auto Insurance Customers

Description All U.S. auto insurance customers

Added: Jan 09, 2020 6:20 PM..PM Format: application/octet-stream Size: 312 KB

Business Terms There are no terms available for this asset.

Tags Auto Insurance

Reviews ★★★★☆ 1 review

Classification None

Schema: 28 Columns 328 Rows ⑧ 4 Columns masked ⓘ Preview: 328 rows Last refresh: 43 seconds ago ⓘ Refresh

⑨ Column masked The values in this column are obfuscated by this data enforcement rule: Protect US Social Security Numbers Learn more

NATIONAL_ID ⓘ	CREDITCARD_NUM... ⓘ	CREDITCARD_T... ⓘ	CREDITCARD... ⓘ	CREDITCARD... ⓘ	EDUCATION... ⓘ	EMPLOYMENT_...
759-11-3090	XXXXXXXXXX	U	Credit Card Expira...	Credit Card Validati...	Text	Code
820-74-9266	XXXXXXXXXX	JCB	XXXXXXXXXXXX	XXXXXXXXXXXX	Bachelor	Employed
778-86-7182	XXXXXXXXXX	Discover	XXXXXXXXXXXX	XXXXXXXXXXXX	College	Unemployed
249-23-7130	XXXXXXXXXX	VISA	XXXXXXXXXXXX	XXXXXXXXXXXX	Doctor	Medical Leave
738-54-3112	XXXXXXXXXX	Diners Club	XXXXXXXXXXXX	XXXXXXXXXXXX	Bachelor	Disabled
		JCB	XXXXXXXXXXXX	XXXXXXXXXXXX	Master	Employed
		Diners Club	XXXXXXXXXXXX	XXXXXXXXXXXX	Master	Medical Leave
		JCB	XXXXXXXXXXXX	XXXXXXXXXXXX	High School or Br	Employed
		Discover	XXXXXXXXXXXX	XXXXXXXXXXXX	College	Employed

Notice that the **NATIONAL\_ID**, **CREDITCARD\_NUMBER**, **CREDITCARD\_EXP** and **CREDITCARD\_CVV** columns have a lock icon next to their name indicating that the data is being protected.

9. Hover over the **lock** icon next to the **NATIONAL\_ID** column.

Hover over the **lock** icon on the other columns being protected as well.

IBM Cloud Pak for Data

All auto insurance

Search Results > Catalogs > Auto Insurance > Auto Insurance Customers

Overview Access Review Profile Lineage

DATA ASSET

Auto Insurance Customers

Description: All U.S. auto insurance customers

Added: Jan 09, 2020 6:20 PM..PM  
Format: application/octet-stream  
Size: 312 KB

Business Terms: There are no terms available for this asset.

Tags: Auto Insurance

Reviews: ★★★★☆ 1 review

Classification: None

**Schema:** 28 Columns 328 Rows 4 Columns masked

Preview: 328 rows Last refresh: 2 hours ago Refresh

NATIONA...	CREDITCARD_NU...	CREDITCARD_T...	CREDITCARD...	CREDITCARD...	Type: String	Type: String	Type: String	Type: String	Type: String
454-23-9752	XXXXXXXXXXXX	JCB	XXXXXXXXXXXX	XXXXXXXXXXXX	Bachelor	Employed			
722-58-1257	XXXXXXXXXXXX	Discover	XXXXXXXXXXXX	XXXXXXXXXXXX	College	Unemployed			
503-81-4643	XXXXXXXXXXXX	VISA	XXXXXXXXXXXX	XXXXXXXXXXXX	Doctor	Medical Leave			
759-11-3090	XXXXXXXXXXXX	Diners Club	XXXXXXXXXXXX	XXXXXXXXXXXX	Bachelor	Disabled			
820-74-9266	XXXXXXXXXXXX	JCB	XXXXXXXXXXXX	XXXXXXXXXXXX	Master	Employed			
778-86-7182	XXXXXXXXXXXX	Diners Club	XXXXXXXXXXXX	XXXXXXXXXXXX	Master	Medical Leave			
249-23-7130	XXXXXXXXXXXX	JCB	XXXXXXXXXXXX	XXXXXXXXXXXX	High School or Br	Employed			
738-54-3112	XXXXXXXXXXXX	Discover	XXXXXXXXXXXX	XXXXXXXXXXXX	College	Employed			

10. Click on the **Profile and settings** icon in the top right corner.

11. Click **Log out**.

## Summary

You completed the IBM Watson Knowledge Catalog tutorial: **Fuel Data Innovation and AI with IBM® Watson Knowledge Catalog**.

You explored: Creating a Governed Knowledge Catalog, Discovering and Cataloging Data Assets, Understanding and Socializing Data Assets, Shopping for Data, Preparing Data for Analytics and AI and Protecting Sensitive Information.

To further your education on Cloud Pak for Data and Watson Knowledge Catalog and many other IBM products and solutions, visit the [IBM Demos](#) website.