

*IBM Data Product Hub as a Service*





---

# Tables of Contents

<b>Welcome</b>	1
What's new	1
Overview	6
Data Product Hub API	8
Known issues and limitations	8
Glossary	9
<b>Getting started with Data Product Hub</b>	10
Logging in to Data Product Hub	11
Managing your settings	13
Video library	13
<b>Setting up and administering</b>	14
Setting up the account and service	17
Setting up the IBM Cloud account	18
Adding users to the IBM Cloud account	19
Provisioning the Data Product Hub service	20
Monitoring your community resource usage	21
Data Product Hub service plans	22
IBM Cloud account security	23
Setting up Data Product Hub	25
Managing storage	25
Managing the community	27
Managing business domains	28
Managing connections	29
Understanding credentials for connections	30
Managing custom properties	33
Roles and permissions	35
Administration	37
Activity Tracker events	37
Managing the Service API key	38
Accessibility	39
Data management	40
Deployment models	41
High availability and disaster recovery	42
Troubleshooting	43
<b>Discovering data products</b>	44
Managing your task inbox as a consumer	45
Requesting a new data product	46
Searching for data products	47
Subscribing to a data product	49
Subscribing to a data product that requires approval	50
Flight client example for accessing a data product	51
<b>Publishing a data product</b>	52
Managing your task inbox as a producer	54
Managing your Insights dashboard	55
Creating a data product	56
Creating a data product from SQL	58
Creating a data product from a query	58
Creating a data product from a custom query	59
Creating a data product from a complex query	61
Creating a data product directly from a source	62
Creating a data product from a catalog	63
Creating a data product from a project	64
Creating a data product from a URL	65

Best practices for creating a data product	65
Managing the lifecycle of data products	66
Creating data source connections	68
Amazon RDS for MySQL connection	69
Amazon RDS for Oracle connection	70
Amazon RDS for PostgreSQL connection	71
Amazon Redshift connection	71
Amazon S3 connection	72
Setting up temporary credentials or a Role ARN for Amazon S3	73
Apache Cassandra connection	74
Apache Derby connection	75
Apache HDFS connection	76
Apache Hive connection	77
Apache Impala connection	77
Cloudant connection	78
Dremio connection	79
Dropbox connection	80
Elasticsearch	80
Google BigQuery connection	81
Google Cloud Storage connection	83
Google Looker connection	86
HTTP connection	86
IBM Cloud Data Engine connection	87
IBM Cloud Databases for MongoDB connection	88
IBM Cloud Databases for PostgreSQL connection	89
IBM Cloud Object Storage connection	89
IBM Cognos Analytics connection	91
IBM Data Virtualization connection	92
IBM Data Virtualization Manager for z/OS connection	92
IBM Db2 Big SQL connection	93
IBM Db2 for i connection	93
IBM Db2 for z/OS connection	95
IBM Db2 on Cloud connection	96
IBM Db2 Warehouse on Cloud connection	96
IBM Db2 connection	97
IBM Informix connection	98
IBM Netezza Performance Server connection	99
IBM Planning Analytics connection	100
IBM watsonx.data Presto connection	100
MariaDB connection	105
Microsoft Azure Blob Storage connection	106
Microsoft Azure Cosmos DB connection	107
Microsoft Azure Data Lake Storage connection	108
Microsoft Azure File Storage connection	110
Microsoft Azure SQL Database connection	111
Microsoft SQL Server connection	112
MongoDB connection	113
MySQL connection	114
OData connection	115
Oracle connection	116
PostgreSQL connection	117
Presto connection	118
Salesforce.com connection	119
SAP OData connection	119
SingleStoreDB connection	120
Snowflake connection	121
Teradata connection	121
Working with delivery methods	122
Delivery methods for connectors	123

---

# Documentation for IBM Data Product Hub as a Service

## Get a data product

[Log in to Data Product Hub](#)

[Find data products for your business](#)

[Subscribe to a data product](#)

[Overview](#)

## Publish a data product

[Publish a data product](#)

[Manage data products](#)

[Create data source connections](#)

## Set up Data Product Hub

[Set up the IBM Cloud account and Data Product Hub service](#)

[Set up Data Product Hub](#)

## Get help

[What's new](#)

[Open and review support cases !\[\]\(41aea2746216b27a6939d696d8e035da\_img.jpg\)](#)

[Learn more about IBM Data Product Hub !\[\]\(7bc43b319a082987e20f7bf78f4bab80\_img.jpg\)](#)

---

## What's new

Check back often to learn about new features and updates for Data Product Hub.

### Week ending 14 Feb 2025

---

#### **Data Product Hub is now available in the Toronto region**

14 Feb 2025

The Toronto data center has been added to the regions for Data Product Hub. Data Product Hub is now available in US (Dallas), Australia (Sydney) and Canada (Toronto). For more information, see [Logging in to Data Product Hub](#).

### Week ending 10 Jan 2025

---

#### **Data Product Hub has achieved SOC2 Certification**

08 Jan 2025

## Week ending 13 Dec 2024

---

### Deprecation of IBM Cloud Object Storage Lite Plan

13 Dec 2024

The Cloud Object Storage Lite plans that you provisioned prior to 1 July 2024 are deprecated as of 15 December 2024. At some point existing Cloud Object Storage buckets might be removed. To retain your Data Product Hub data, you must upgrade your Cloud Object Storage service to the Standard plan. The data stored in Cloud Object Storage includes uploaded data contracts, visualizations for items in a data product, and data associated with projects. If you do not upgrade your Cloud Object Storage plan to Standard, your data might be permanently deleted. The Standard Plan offers a Free Tier allowance that allows free storage for up to 5GB per month. For more information, see [Cloud docs: How do I upgrade a service instance from a Lite Plan to a Standard Plan?](#). You might need to upgrade your IBM Cloud account before you can switch to the Standard plan.

### Data Product Hub available from the Sydney, Australia region

9 Dec 2024

The Sydney, Australia (au-syd) data center has been added to the regions for Data Product Hub. Data Product Hub is now available in US (Dallas) and Australia (Sydney).

## Week ending 8 Nov 2024

---

### Monitor your insights with dashboards

7 Nov 2024

Both administrators and data producers can now use the Insights dashboard to monitor their data products and community activity. The Insights dashboard provides a comprehensive, centralized overview of open tasks, delivered data products, community resources, and more. By delivering real-time metrics, the Insights dashboard provides detailed data insights at scale, increases workflow efficiency, and helps ensure transparency across all data community. For details, see [Managing your Insights dashboard](#).

### Create custom business domains to organize your data products

7 Nov 2024

Improve your data community's organization and optimize your data products' searchability by creating custom business domains. With new, custom business domains, you can easily organize your data products into intuitive categories and curate your community for your business needs. You can also create subdomains that further refine your data community's organization and empower data consumers to easily find the data products they need. For more information, see [Managing business domains](#).

### Add custom properties to data products

7 Nov 2024

You can now create and add custom properties to data products to optimize searchability, classification, and organization. By adding custom properties, you can structure information and curate your data product to meet specific business needs. For details, see [Managing custom properties](#).

### Enhance and update your published data products

7 Nov 2024

To further enhance and improve data products, you can now create new versions of your published data products. When creating a new version, you can add or delete data assets, change the available delivery methods, and manage the access level. By continuously creating new versions of published data products, you can help ensure data accuracy and currency. For more details, see [Creating new versions of data products](#).


### Preapprove data consumers for data products requiring approval

7 Nov 2024

For data consumers who frequently access data products requiring approval, you can now create a list of preapproved users or user groups. By defining a preapproved list, you streamline the subscription process for both data consumers and producers and improve efficiency in delivering data products. For more information, see [Completing your data product](#).

### Send notifying comments for requests for new data products

7 Nov 2024

You can now send notifying comments for requests for new data products from your **Task inbox**. The approver for a data product request can enter a comment or question for the requester. The requester receives a notification under the **Notification bell** icon  or by email. The approver is also notified when the requester responds. Previously comments were recorded in the feed for the task, but there was no notification, so comments could go unnoticed. For more information, see [Managing your task inbox as a producer](#) and [Managing your task inbox as a consumer](#).

## Week ending 4 Oct 2024

---

### Add an embedded preview to data products

3 Oct 2024

To ensure a continuous, focused flow and enhance browser organization, you can now embed a preview of your data contents when creating data products from URLs. This preview allows you to open and view the data product contents directly within Data Product Hub, rather than be redirected to another browser tab. For more details, see [Creating a data product from URL](#).

## Week ending 30 Aug 2024

---

### Use a complex SQL query to create a data product

29 Aug 2024

You can now create a data product by using complex SQL that allows you to query temporary and transient tables. A Snowflake connection is required to use complex SQL to create data products. For more information, see [Creating a data product from a complex query](#).

## Week ending 16 Aug 2024

---

### New service plan added: Lite plan

13 Aug 2024

You can now provision an instance of Data Product Hub using a Lite plan. The Lite plan provides a free trial to Data Product Hub and introduces users to the data community. For more information, see [Data Product Hub service plans](#).

## Week ending 26 July 2024

---

### Download data extract delivery method

26 July 2024

You can now add file extracts from your data source to a data product using the download data extract delivery method. The data extract delivery method requires two connections, one for connecting to the database you are extracting from (the source) and another connection for delivering the file to the consumer (the target). For more information, see [Working with delivery methods](#). To find the connectors that can serve as source and target connections, see [Delivery methods for connectors](#).

## Week ending 12 July 2024

---

## Customizable queries

12 July 2024

You can now curate a data product to contain specific parameters and values by using customizable queries. By using customizable queries, data products can be reused more frequently and consumers receive the specific data that they need. For more information, see [Creating a data product from a customizable query](#).

## Week ending 19 June 2024

---

### Data Product Hub is generally available on IBM Cloud

19 June 2024

You can provision an instance of Data Product Hub from the [IBM Cloud catalog](#). For more information, see [Provisioning the Data Product Hub service](#).

## Week ending 7 June 2024

---

### New data product creation method: Add from catalog

7 June 2024

You can now create data products by adding assets from a catalog in IBM Knowledge Catalog. With IBM Knowledge Catalog, you can run metadata enrichment on your data assets, which adds important information describing the asset's contents and data quality. For details, see [Creating a data product from a catalog](#).

### New data product creation method: Add from project

7 June 2024

You can now create data products by adding data assets from a project. You can use this method to create a data product containing a query or if you do not have access to IBM Knowledge Catalog. For details, see [Creating a data product from a project](#).

## Week ending 17 May 2024

---

### New delivery method added: the Flight service

17 May 2024

You can now create and deliver data products with the Flight service. With the Flight service, you receive real-time read and write access to data sources and you no longer need a connection asset to data sources requiring personal credentials. For more information, see [Working with delivery methods](#).

## Week ending 3 May 2024

---

### Setting up notifications

3 May 2024

You can set up notifications to alert you about overdue tasks, status updates on requested data products, and access to data products requiring approval. These notifications can be delivered on-screen or sent to your email. For more information, see [Managing your notifications](#).

## Week ending 5 April 2024

---

### Sharing connections

5 April 2024



You can share connections with community members who have the **Editor** or **Admin** role. They can use the connection to create and deliver data products. Be cautious when you share connections since you are sharing your credentials. For more information, see [Sharing a connection](#).

## Data products that require approval

5 April 2024

Producers can restrict access to data products by requiring an approval for the subscription. One or more community members are designated as approvers. The consumer provides a business justification for the subscription, which is tracked in the Task inbox. Access to the data product is Approved or Rejected by the approver. See [Completing your data product](#) and [Subscribing to a data product that requires approval](#)

## Requesting a new data product

5 April 2024

Consumers can curate data to their business needs by creating and submitting a data product request. The consumer indicates the data requirements, terms of usage, and a delivery method for the request, which is then submitted to producers for review and approval. If approved, the producer creates the data product and delivers it to the consumer upon completion. See [Requesting a new data product](#).

## New connectors added: Google Cloud Storage, Microsoft Azure Blob Storage, and Microsoft Azure File Storage

5 April 2024

Google Cloud Storage, Microsoft Azure Blob Storage, and Microsoft Azure File Storage connectors are now available for creating and delivering data products. See [Connectors for Data Product Hub](#).

# Week ending 23 February 2024

---

## Edit and delete connections

23 February 2024

You can now **Edit** and **Delete** connections from the list of the data source connections that you own. The **Admin** or **Editor** role is required to manage connections. For more information, see [Managing data source connections](#).


## Rotate the Service API key

23 February 2024

An **Admin** can now rotate the Service API key to provide a more secure deployment of Data Product Hub. The Service API key is stored on IBM Cloud. For more information, see [Managing the Service API key](#).

## Edit the name and business domain for a data product

23 February 2024

You can now edit the name and business domain for a published data product. Click the **Pencil** icon  to edit. Click the checkmark to save your changes.

# Week ending 02 February 2024

---

## Connections list

01 February 2024

You can now view a list of the data source connections that you created. The **Admin** or **Editor** role is required to list connections. For more information, see [Viewing data source connections](#).

## Week ending 01 December 2023

---

### **Closed beta release of Data Product Hub**

1 December 2023

This is a closed beta release available to internal IBM users and to authorized external users.

## Week ending 3 November 2023

---

### **Experimental release of Data Product Hub**

2 November 2023

This is an experimental release of Data Product Hub. It is being distributed internally with limited availability.

---

## Data Product Hub overview

IBM Data Product Hub as a Service is a self-service solution that is used by data-driven enterprises to share data products. On Data Product Hub, data producers can publish curated data products to share with data consumers in their community. Data consumers can easily access data products for their business needs.

Data producers use the full-service Data Product Hub solution to package, productize, and share their data-rich assets. Data producers manage data as a product to provide more value to teams across the organization. When a producer publishes their data product, they assign a short description, business domain, and recommended usage to guide consumers who are searching for a data product. Producers ensure that published data products are accurately represented. Producers ensure that the items in a data product are accurate and available. Producers are also responsible for upholding any relevant clauses in the data contract.

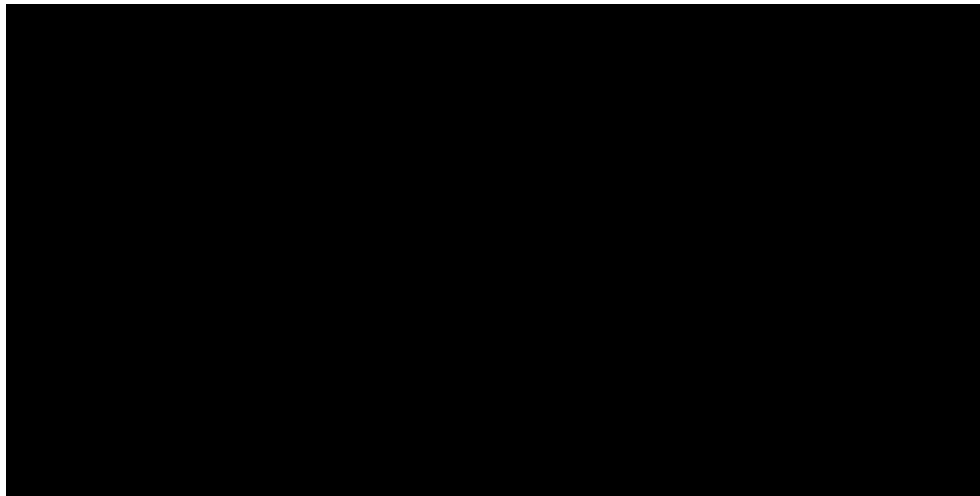
Data consumers use Data Product Hub to discover and access data products for their business needs. Consumers can read descriptions of the data product and decide whether it fits their needs. Consumers are responsible to read and uphold any relevant clauses in the data contract. Data consumers quickly and easily access the right data by using an intuitive interface.

## What is a data product?

---

Data products can contain one or more data or data-related assets. They are curated, packaged, and distributed to be easily accessible and reusable. Unlike data assets in governance catalogs, data products are managed as products with lifecycle management, wide distribution, and multiple purposes to provide maximum business value.

This video provides a visual method to learn the concepts and tasks in this documentation.



## **Packaged and curated**

A data product is a collection of curated data or data-related assets that are packaged for reuse and distribution on Data Product Hub. Data products can contain data, as well as models, dashboards, and other computational asset types. To provide flexibility, you can use several methods to add items to a data product. For example, you can add items using queries, from a direct source, or from a connected catalog.

## **Managed as a product**

Data products are assigned to a designated owner who manages them through the draft, published, and retired stages. Similar to other products, data products have a market demand, value proposition, and defined terms and conditions. The typical workflow for creating and managing data products iterates through use case definition, preparing the data, defining the data contract, testing, and publishing.

Using Data Product Hub, you can apply product management principles to your data products, including:

- Define the use case for the data product: Clearly define the purpose and objectives of the data product, including what problem it solves or what value it provides to users.
- Assign an accountable owner: The owner manages the lifecycle of a data product from initiation to end of life.
- Restrict distribution: Designate approvers for data products to restrict distribution to approved consumers.
- Prepare the data: Ensure that the data assets are high quality, accurate, and secure. Remove any Personal Identifiable Information (PII). Choose an appropriate data source and delivery method.
- Include a data contract: The data contract establishes transparency by describing the Terms and Conditions and Service Level Agreements.
- Testing and Validation: Thoroughly test the data product before publishing.
- Publish new versions as needed.

## **Optimized for large-scale distribution**

Data products are highly reusable across any number of consumers who are community members. They can be delivered through multiple delivery methods for maximum availability.

## **Associated with a data contract**

Data products are associated with a data contract that outlines the terms and conditions of usage. The data contract provides assurance on both ends for data products that are distributed across teams.

## **Learn more**

---

- [Setting up Data Product Hub](#)
- [Introduction to publishing data products](#)
- [Finding a data product](#)

Parent topic: [Documentation for Data Product Hub](#)

---

# Data Product Hub API

Data Product Hub includes a REST API for working with data products and also requires API calls from the Watson Data API.

The Data Product Hub API provides programmatic control for creating and managing data products.

Data Product Hub SDKs for Java, Node, Python, and Go are available to access the API from your code. The client libraries that are provided by the SDKs implement best practices for using the API and reduce the amount of required code.

Use the Data Product Hub API to create, delete, and update data products. See [Data Product Hub API Service](#).

You also use the following APIs from the [Data and AI Common Core API](#) to create or consume data products:

- Assets
- Catalogs
- Asset Lists
- Connections
- Global Search

Parent topic: [Documentation for Data Product Hub](#)

---

## Known issues and limitations

The following known issues and limitations apply to Data Product Hub.

---

### Known issues

#### Regional support

Data Product Hub is currently available in the **US Dallas (us-south)**, **Australia (Sydney) au-syd** and **Canada (Toronto) ca-tor** regions. The region switcher lists other regions, but they are not supported at this time.

#### Visualization availability in the Sydney region

The visualization to preview the items in a data product is not available from the **Australia (Sydney) - au-syd** region.

#### Assign the Viewer role to the Data Product Hub Service ID (for early adopters)

If you provisioned Data Product Hub before May 2024, the IBM Cloud account administrator must explicitly assign the **Data Product Hub > Viewer** role to the Data Product Hub Service ID. This step is required only if the Data Product Hub service was provisioned before May 2024. The access policy is automatically assigned for Data Product Hub instances provisioned starting in May 2024.

Follow these steps to assign a role to the Service ID:

1. As the account administrator or owner, open the **Manage > Access(IAM) > Service IDs** page for your IBM Cloud account.
2. Search for the **data-product-admin-service-id-xxxx** Service ID and click it to open the **Access** page.
3. If the **Data Product Hub > Viewer** role is not listed under **Access policies**, click **Assign access**.
4. Choose the Data Product Hub Service and complete the steps to assign the **Viewer** role.

---

### Limitations

#### Use the API to delete connections that are in use

From the user interface, you cannot delete a connection that is in use by a published data product. Using the user interface, you must retire all the data products that use the connection before you can delete the connection. You can override the user

interface by using the API. Use the following API call to delete a connection that is in use by one or more published data products:

Call the **DELETE Connection** endpoint to delete connections. For example:

```
DELETE /v2/connections/{connection_id}
```

When a connection is deleted by using the API, the items in the data products that use the deleted connection cannot be delivered. If there are items in a data product that use other connections, those items will still be deliverable.

For details, see [Delete connection](#).

**Parent topic:** [Documentation for Data Product Hub](#)

---

## Glossary

This glossary provides terms and definitions for Data Product Hub.

[A](#) | [C](#) | [D](#) | [M](#) | [N](#) | [P](#) | [S](#) | [I](#) |

### A

---

#### **approval processes**

Automated workflows based on templates that provide standardized tasks for approving access to data products.

### C

---

#### **community**

A group of users that are members of an instance of Data Product Hub and may either create or subscribe to data products, or administer and configure the Data Product Hub.

#### **connection**

Connectors to data sources that require credentials to access data assets.

### D

---

#### **data contract**

The required data contract is a document describing the Terms and Conditions for using the data product. It represents a contract between the data producer and subscribers regarding the quality and contents of the data contained in the data product.

#### **data extract**

A method for delivering data products that generates a file, or extract, from a table and stores it in a specified location to make it available for downloading by the subscriber.

#### **data product**

A collection of optimized data or data-related assets that are packaged for reuse and distribution with controlled access. Data products contain data as well as models, dashboards, and other computational asset types. Unlike data assets in governance catalogs, data products are managed as products with multiple purposes to provide business value.

#### **data product item**

Individual assets that are included in a data product. Items in a data product can be a variety of types, including tables, CSV files, results of a SQL query, URLs, etc. Delivery methods and custom properties are assigned to items.

#### **delivery method**

The method for receiving delivery of the items in a data product as determined by the connector type.

#### **domain**

An area of expertise or activity in business that can be used to categorize data products. Domains can be divided into

additional levels called subdomains.

## M

---

### **member**

Users that have been added to the Data Product Hub community and assigned a role that allows a certain level of access.

## N

---

### **notifications**

Automated messages that provide status updates for workflow tasks.

## P

---

### **properties**

Business-specific metadata assigned to data products or items in a data product that help users discover data products efficiently. Properties may be standard properties included with Data Product Hub or custom properties tailored for a company's business needs.

## S

---

### **subscription**

A subscription is an agreement to receive access to a data product. Acceptance of the terms and conditions in the data contract is required before access is granted.

## T

---

### **task**

Automated tasks that track participation in a review process.

### **task inbox**

A dashboard-style screen containing tasks assigned to and initiated by a Data Product Hub member where progress on the tasks is easily tracked.

**Parent topic:** [Documentation for Data Product Hub](#)

---

## Getting started with Data Product Hub

You can sign up with Data Product Hub to start creating and sharing data within your organization. Data Product Hub provides an end-to-end experience in which consumers can request data products and producers can create and deliver those requests.

This video provides a visual method to learn the concepts and tasks in this documentation.

## Setting up the solution as an administrator

---

To set up the Data Product Hub solution for your organization, see [Setting up the account and service](#).

## Start working

---

Once your administrator provisions the Data Product Hub instance and assigns collaborator roles, you can start exploring the solution:

- [Log in](#) to your Data Product Hub account. Ensure you have an IBMid for IBM Cloud.
- [Configure](#) your profile and settings.

## Learn about the solution

---

Once you log in, you can start exploring the solution to discover and share data products within your organization:

- If you are a data product consumer, see [Getting a data product](#).
- If you are a data product producer, see [Publishing a data product](#).
- If you are an administrator, see [Setting up and administering](#) and [Administration](#).

### Other information

- [Data Product Hub API](#)
- [Known issues and limitations](#)
- [Roles and permissions](#)

**Parent topic:** [Documentation for Data Product Hub](#)

---

## Logging in to Data Product Hub

You must have an IBMid and appropriate role assignments to log in to Data Product Hub.

## Regional availability

---

Data Product Hub is available in the following regions:

- **US (Dallas) - us-south**
- **Australia (Sydney) - au-syd**

- **Canada (Toronto) - ca-tor**

## Accessing Data Product Hub

---

When Data Product Hub is provisioned for your organization, you can access it with the following URL:

<https://dataplatfom.cloud.ibm.com/dpx?context=dph>

Follow these steps to log in:

1. Enter your IBMid.
2. Select your organization's IBM Cloud account.
3. Select the region.

If you receive an error when logging in, see [Troubleshooting](#) for possible solutions.

Following are the requirements for accessing Data Product Hub:


- An IBMid for IBM Cloud
- Membership in the IBM Cloud account for Data Product Hub
- Appropriate IAM role assignments on IBM Cloud
- Appropriate collaborator roles for Data Product Hub

Following are the application prerequisites for Data Product Hub:

- A provisioned instance of Data Product Hub on an IBM Cloud account
- The correct account is selected
- Either **US (Dallas)**, **Australia (Sydney)** or **Canada (Toronto)** is selected as the region

## Resource availability

---


In Data Product Hub, you can work with resources that reside in the current account. To work with resources in another account, select the account using the **Account switcher** icon .

## Switching accounts or regions


---

You can switch accounts or regions easily in the header.

To switch to your organization's account:

1. Log in to Data Product Hub with your IBMid.
2. Click the **Account switcher** icon  in the header.
3. Select your organization's account from the list of accounts.
4. Optional: If the **Account switcher** icon isn't visible, you can select the account from the dropdown menu.

To switch regions:

1. Log in to Data Product Hub with your IBMid.
2. Click the **Region switcher** icon  in the header.
3. Select **US (Dallas)**, **Australia (Sydney)** or **Canada (Toronto)**.
4. Optional: If the **Region switcher** icon isn't visible, you can select the region from the dropdown menu.

## Learn more

---

- [Adding users to the account and assigning roles](#)
- [Roles and permissions for Data Product Hub](#)
- [IBM Cloud catalog](#)
- [IBM Cloud docs: IAM access](#)
- [IBM Cloud docs: Inviting users to an account](#)
- [IBM Cloud docs: Managing your account](#)



---

## Managing your settings

You can manage your profile, change the user interface mode, and configure notifications in IBM Data Product Hub as a Service.

### Managing your profile

---

To configure your profile, click your avatar in the banner and then click **Profiles and settings**.

You can make the following changes to customize your profile:

- Add or change your avatar photo.
- Customize your service location filters by resource group and location. Ensure that you select regions where Data Product Hub is available, currently Dallas.
- Set your user interface to dark theme. You can switch between Light and Dark theme by using the toggle.
- View your account details, such as **Account ID** and **Account Type**. If you are using a trial account, you can view how many days remain in your trial.
- Choose to leave Data Product Hub. When you leave the service, all personally identifiable information and data are removed completely from all data sources, including backups, after 30 days.

### Managing your notifications

---

To configure your notifications settings, click the **Notification bell** icon  and then click the **Settings** icon .

For consumers, the notifications provide status updates on requested data products and access to data products requiring approval.

For producers, the notifications inform them of new data product requests, granting access to data products requiring approval, and overdue tasks.

You can make the following changes to your notifications settings:

- Choose how to receive notifications. You can receive notifications that appear briefly on-screen or to your email.
- If you select **Do not disturb**, you continue to see notifications on the home page and the number of notifications on the bell.
- Choose the projects from which you want to receive notifications. A list of all your projects is provided.

Parent topic: [Getting started with Data Product Hub](#)

---

## Video library

Watch short videos for producers and consumers to learn about Data Product Hub. The videos show concepts and common tasks performed in Data Product Hub.

Note:

These videos provide a visual and auditory method to learn the concepts and tasks in this documentation.

# Overview for setting up and administering Data Product Hub

When you set up and administer Data Product Hub, you complete steps on both IBM Cloud and on Data Product Hub. Several important initialization steps are performed upon the initial login to Data Product Hub. The initial login can be performed by either the account administrator or by a delegated user.

## Step 1: Set up the IBM Cloud account

Before you can configure Data Product Hub, the IBM Cloud account owner or administrator must complete the prerequisite configuration steps on IBM Cloud. On IBM Cloud, the account administrator creates an account, adds users or access groups, assigns roles, and provisions the Data Product Hub and IBM Cloud Object Storage services.

### Required roles

IAM Platform role: **IBM Cloud account owner or administrator**

The following image outlines the steps that are performed by the IBM Cloud account administrator:



### IBM Cloud account owner or admin

- Invite users
- Assign IAM roles (including the Data Product Hub Manager role)
- Provision services

IBM Cloud configuration tasks

Task	Where performed?	Frequency
Create an account.	IBM Cloud	Once, unless more than one account is needed.

Task	Where performed?	Frequency
Provision the Data Product Hub service.	IBM Cloud catalog	Once. Only one instance of Data Product Hub is allowed per account.
Add users to the account.	IBM Cloud IAM	Initial setup and ongoing to add users.
Assign IAM roles, including the Data Product Hub Manager role.	IBM Cloud IAM	Initial setup and ongoing if roles change or new users are added.
Provision Cloud Object Storage.	IBM Cloud catalog	Once.

The Data Product Hub Manager role is an IAM service role that must be assigned to a user. That user must log in to initialize Data Product Hub.

Important:

Data Product Hub automatically generates a Service ID in your IBM Cloud account. The Service ID takes the form *data-product-admin-service-id-catalog\_id*. The Service ID is required to process data product subscriptions in your catalog on IBM Cloud. Do not delete, lock, or modify the Data Product Hub Service ID. You can view the Service IDs for your account at [Service IDs](#).

The Service ID requires an API key for authorization. A best practice for secure operations is to periodically rotate the Service API key. See [Managing the Service API key](#).

## Step 2: Initialize Data Product Hub

The initialization process automatically configures Data Product Hub to prepare it for your data community. A catalog is created with associated sample data products, delivery methods, workflows, and domains. The Data Product Hub Service ID is configured to make API calls.

Data Product Hub is initialized upon the initial login. The initial login must be performed by one of the following types of users:

- The administrator for the IBM Cloud account, or
- A delegated user with the **Manager** role for the Data Product Hub service.

You have a choice of 2 options for initializing Data Product Hub. For option 1, the account administrator performs the initial login step. For option 2, the account administrator delegates another user as the Data Product Hub **Manager** who performs the initial login step.

For instructions on how to log in to Data Product Hub, see [Logging in to Data Product Hub](#).

### Option 1: Account administrator logs in to initialize Data Product Hub

For option 1, the administrator for the IBM Cloud account logs in to Data Product Hub to initialize it. They then assign the IAM **Manager** role for the Data Product Hub service to themselves or another user. They are automatically added as a member of the Data Product Hub community with the **Admin** collaborator role.

#### Required roles for the account administrators

IAM Platform role for Data Product Hub: **Any role**

IAM Service role for Data Product Hub: **Manager**

Data Product Hub collaborator role: **Admin** (automatically assigned upon login)

The following image outlines the login step that is performed by the account administrator:



### Account admin

- Log in to initialize Data Product Hub

#### Initializing Data Product Hub

Task	Where performed?	Frequency
Log in to initialize Data Product Hub.	Data Product Hub login screen	Once.

### Option 2: Data Product Hub Manager logs in to initialize Data Product Hub

For option 2, the account administrator delegates another user to perform the initial login step by assigning them the IAM Service role Data Product Hub **Manager** and other required roles. The **Manager** logs in to initialize Data Product Hub. When the **Manager** logs in to Data Product Hub for the first time, they are automatically assigned the **Admin** collaborator role. They can add more administrators to manage the Data Product Hub community by assigning users with the **Admin** collaborator role.

#### Required roles for the Data Product Hub Manager

IAM Platform role for All Account Management services: **Administrator**

IAM Platform role for Data Product Hub: **Administrator**

IAM Service role for Data Product Hub: **Manager**

Data Product Hub collaborator role: **Admin** (automatically assigned upon login)

The following image outlines the login step that is performed by the Data Product Hub **Manager**:



### Data Product Hub Manager

- Log in to initialize Data Product Hub
- Add an administrator to the Data Product Hub community with the **Admin** collaborator role

#### Initializing Data Product Hub

Task	Where performed?	Frequency
Log in to initialize Data Product Hub.	Data Product Hub login screen located at <a href="https://datapatform.cloud.ibm.com/dpx?context=dph">https://datapatform.cloud.ibm.com/dpx?context=dph</a>	Once
Add an administrator to the Data Product Hub community with the <b>Admin</b> collaborator role.	Manage community page located at <b>Administration&gt;Configurations and settings&gt;Manage community</b>	Once

## Step 3: Set up Data Product Hub

After Data Product Hub is initialized by logging in, the Data Product Hub **Admin** adds members to the data community and assigns collaborator roles. The **Admin** role is automatically assigned to the user that initializes the instance, but can also be assigned to any other users. The **Admin** sets up storage for data contracts and data extracts, and manages data source connections.

#### Required roles

IAM Platform role: **Any role**  
Data Product Hub collaborator role: **Admin**

The following image outlines the steps to set up the Data Product Hub community:



#### Data Product Hub admin

- Add members and assign collaborator roles
- Manage storage
- Manage data source connections

#### Data Product Hub Admin tasks

Task	Where performed?	Frequency
Create a bucket in a Cloud Object Storage instance to store data contracts.	Data Product Hub Configurations and settings	Once, unless a new Cloud Object Storage instance is needed.
Add users and assign roles to manage the data community.	Data Product Hub Configurations and settings	Initial setup and ongoing if roles change or new users are added.
Manage data source connections.	Data Product Hub Configurations and settings	Ongoing.

## Learn more

- [Setting up an IBM Cloud account and Data Product Hub service](#)
- [Setting up Data Product Hub](#)
- [General administration for Data Product Hub](#)
- [Logging in to Data Product Hub](#)

Parent topic: [Documentation for Data Product Hub](#)

## Setting up an IBM Cloud account and Data Product Hub service

Setting up Data Product Hub includes steps to create an IBM Cloud account, provision the service, and add users with appropriate roles.

### Setting up the IBM Cloud account

The IBM Cloud account must be set up before you can set up Data Product Hub.

Complete these steps to set up the IBM Cloud account:

- Create an IBM Cloud account (or use an existing account)
- Provision Data Product Hub service and select a pricing plan
- Invite users to the account (if they are not already members) and assign appropriate IAM roles

Setting up an IBM Cloud account, adding users, and assigning IAM roles requires familiarity with the IBM Cloud IAM services.

**Required roles to complete this task** If you are the Owner of the IBM Cloud account, you have the required permissions to provision Data Product Hub service and add users. Nonowners who add users must be assigned the correct roles, as shown in the following table:

Roles and permissions for account administrator

Service name	IAM Platform roles	IAM Service access roles	Permissions
All Identity and Access enabled services	Administrator	Manager	All permissions in roles
All Account Management services	Editor	Not applicable	All permissions in roles

Refer to the following topics for details:

- [Setting up an IBM Cloud account](#)
- [Provisioning the Data Product Hub service and choosing a plan](#)
- [Adding users to the account and assigning roles](#)
- [Roles and permissions for Data Product Hub](#)

## Next steps

---

- [Setting up Data Product Hub](#)

## Learn more

---

- [IBM Cloud docs: IAM access](#)
- [IBM Cloud docs: Setting up access groups](#)

**Parent topic:** [Setting up and administering Data Product Hub](#)

---

# Setting up an IBM Cloud account

An IBM Cloud account is required to provision an instance of Data Product Hub. As the IBM Cloud account owner or administrator, you set up a payment method in the IBM Cloud account and then provision an instance of the Data Product Hub.

## Creating a IBM Cloud Pay-As-You-Go account

---

IBM Cloud accounts that are paid by credit card are called Pay-As-You-Go accounts. You pay only for billable services that you use, with no long-term contracts or commitments. You must have a billable IBM Cloud account to provision the Data Product Hub service. The billing information is stored but not charged for free plans.

To set up a new Pay-As-You-Go account:

1. Create an IBM Cloud account, if you don't already have one. You can create an account at [Create an IBM cloud account](#).
2. Enter the information for your account, including your payment information.
3. Click **Create account** to submit your information.

After your payment information is processed, you can access and provision the services in the IBM Cloud catalog, including the Data Product Hub service.

For billable services, you receive a monthly invoice for usage beyond any allowances.

## Next steps

---

- [Adding users to the account](#)
- [Provisioning Data Product Hub service](#)
- [Setting up Data Product Hub](#)

## Learn more

---

- [IBM Cloud docs: Account types](#)

Parent topic: [Account and service configuration](#)

---

## Adding users to the account and assigning roles

As an **Administrator**, you add the people in your organization who need access to the Data Product Hub to the IBM Cloud account and then assign them the appropriate IAM roles for their tasks.

Invite users by sending an invitation from the [Users](#) screen of the IBM Cloud account. The new users receive an email invitation to join the account. They must accept the invitation to be added to the account.

---

## Adding users to your IBM Cloud account

You invite users to your IBM Cloud account by sending an email invitation from the **Manage>Access(IAM)>Users** screen in IBM Cloud. The user accepts the invitation to join the account. After the user accepts the invitation, you assign them roles (or access groups) to provide the necessary permissions to work in Data Product Hub.

You have two options for assigning IAM roles. You can assign roles to individual users or you can create access groups to expedite role assignment.

### IBM account membership

Users need an IBMid to be authorized for Data Product Hub. If the invited user does not have an IBMid, it is created for them when they join the account.

---

## Assigning roles

The roles and permissions for Data Product Hub are described in [Roles and permissions for Data Product Hub](#).

The primary user personas are data product consumer, data product producer, and administrator. Following are the minimum role assignments for each persona.

#### Required roles for data product consumers

IAM Platform role: **Viewer**

Data Product Hub collaborator role: **Viewer**

#### Required roles for data product producers

IAM Platform role: **Viewer**

Data Product Hub collaborator role: **Editor**

#### Required roles for administrators

IAM Platform role: **Viewer**

IAM Service role: **Manager**

Data Product Hub collaborator role: **Admin**

### Working with access groups

Access groups expedite role assignments by grouping permissions for large numbers of users. You create a group and assign policies and rules to the group. When you assign a user to an access group, their access rights are determined by the group policies. All members of an access group have the same access permissions, and all members are updated when the group is edited.

Select **Administration>Access(IAM)>Access groups** to set up access groups for Data Product Hub.

You can also indicate the access group when you invite users to the account.

### Assigning roles individually

Roles can be assigned to individual users. Select **Administration>Access(IAM)>Assign access** for each user.

You can also indicate the access policy when you invite users to the account.

## Next step

The next step is to add the users to the Data Product Hub community and assign them collaborator roles to work in Data Product Hub. For more information, see [Managing the Data Product Hub community](#).

## Learn more

- [IBM Cloud docs: IAM access](#)
- [IBM Cloud docs: What is IBM Cloud Identity and Access Management](#)
- [IBM Cloud docs: Setting up access groups](#)

**Parent topic:** [Account and service configuration](#)

## Provisioning the Data Product Hub service

Provision the Data Product Hub service on IBM Cloud and choose a service pricing plan.

## Required roles

The required roles depend on whether you are provisioning or initializing Data Product Hub.

### Required roles for provisioning Data Product Hub

If you are the **Owner** of the IBM Cloud account, you have the required permissions to provision the Data Product Hub service. Other users who provision the service must be assigned the correct roles, as shown in the following table:

Roles and permissions for provisioning a service

Service name	IAM Platform roles	IAM Service access roles	Permissions
All Identity and Access enabled services	Administrator	Manager	All permissions in roles
All Account Management services	Editor	Not applicable	All permissions in roles
Data Product Hub	Any	Manager	•data-product-hub.dashboard.view •data-product-hub.catalog.manage

### Required roles for initializing Data Product Hub

To initialize the Data Product Hub, you must either be the IBM Cloud account **Owner** or **Administrator**, or be assigned the IAM Service role of **Manager** for the Data Product Hub service.

## Provisioning Data Product Hub

Only one service instance of Data Product Hub is allowed for an IBM Cloud account at a time.

Follow these steps to provision the service:

1. Log in to your IBM Cloud account.
2. Open the **Catalog** and search for Data Product Hub.
3. Select Data Product Hub and read about the service on the **About** tab.
4. Select a pricing plan on the **Create** tab.
5. Review and agree to the **Terms**.
6. Click **Create**.

The service is added to your **Resource list**.



# Initializing Data Product Hub

After the service is provisioned, the account owner or administrator, or someone with the **Manager** role for the Data Product Hub service, must log in to Data Product Hub to initialize the catalog for the account.

Upon the initial login, a catalog and the associated sample data products, delivery methods, workflows, and domains are created for the IBM Cloud account. The catalog tracks the data products and other work in Data Product Hub for the account.

## Choosing a service plan

See [Data Product Hub service plans](#) for a description of available plans.

## Deprovisioning Data Product Hub

Data products are associated with the IBM Cloud account, not with the Data Product Hub instance. When a Data Product Hub instance is deprovisioned, the underlying data products are placed in a suspended state. If a new instance is provisioned in the same account within 30 days, the suspended data products are recovered and made available to the user. Any data product that is in a suspended state beyond 30 days is expired and automatically deleted.

For the Essential plan, if the Data Product Hub service instance is inactive for 60 days, the data products are purged from the database.

## Next steps

- [Setting up Data Product Hub](#)

Parent topic: [Setting up an IBM Cloud account and Data Product Hub service](#)

# Monitoring your account resource usage

Data Product Hub offers several service plans with resource usage limits and charge for additional usage. You can monitor the resources usage of your community or your service instance to ensure the limits are not exceeded.

For the Lite plan, you cannot exceed the limits of the plan. You must wait until the start of your next billing month to use resources that are calculated monthly. To provision more resources, you can upgrade to a paid plan.

For the Essentials plans, you received unlimited data products, and pay per unit of CUH and any additional data shares.

## Capacity unit hours (CUH) for compute usage

Many tools consume compute usage that is measured in capacity unit hours (CUH). A capacity unit hour is a specific amount of compute capability with a set cost.

### Calculating CUH consumption

Capacity unit hours (CUH) are consumed by running assets, but not working in tools. As a result, there is no consumption charge for defining an experiment in Projects, but there is a charge for running assets in experiments.

To monitor your service instance usage, log in to your IBM Cloud account, select **Manage > Billing and usage**, and click **Usage**.

You can view a summary of your usage and billing on [IBM Cloud](#). From the home page, choose **Manage > Billing and usage**. Then, click **Usage** to view the usage for each service.

Processes that consume CUH on Data Product Hub

Tool or Process	Workspace	CUH rate
-----------------	-----------	----------

Tool or Process	Workspace	CUH rate
<a href="#">Data Refinery</a>	Project	Multiple rates

The following processes do not consume CUH for Data Product Hub, but consumes CUH for IBM Knowledge Catalog.

Processes that consume CUH on IBM Knowledge Catalog

Tool or Process	Workspace	CUH rate
<a href="#">Profiling</a>	Project and Catalog	One rate
<a href="#">Metadata enrichment</a>	Project	One rate

## Data shares

Each service plan comes with a limited number of data shares. A data share is calculated when a consumer subscribes to a data product and accesses the data product.

## Published data products

For the Lite plan, you can publish 25 data products for free. To publish more data products, you must upgrade to a paid plan.

## Learn more

- [Data Product Hub service plans](#)

Parent topic: [Setting up and administering](#)

## Data Product Hub service plans

Data Product Hub offers various service plans. Your service plan determines the amount of data products, data shares, and capacity-unit hours available to you each month.

## Available plans

IBM Cloud account owners can choose between a Lite or Essentials plan.

### Lite

The Lite plan is suitable for individual users and provides a free trial of Data Product Hub.

### Essentials

The Essentials plan is suitable for organizations who are interested in discovering and publishing data products.

## Plan offerings

Each plan has a monthly limit of available data products, data shares, and capacity-unit hours (CUH).

Table 1. Data Product Hub service plan offerings

Plan	Features and capabilities	Pricing
Lite	<ul style="list-style-type: none"> <li>• 10 published data products</li> <li>• 25 data shares/month</li> <li>• 20 CUH/month</li> </ul>	Free
Essentials	<ul style="list-style-type: none"> <li>• Unlimited published data products</li> <li>• 250 data shares/month</li> <li>• CUH charged per unit used</li> </ul>	<ul style="list-style-type: none"> <li>• \$1.25 USD/CUH</li> <li>• \$12 USD/data share</li> </ul>

You can only provision one instance of the Lite plan per account.

# Billing

You are only charged for the time that the service instance is provisioned. For example, if you provision an instance of Data Product Hub on 1 Jan and then de-provision it on 20 Jan, you are charged for 20 days. For more details about billing and usage, see [Monitoring your account resource usage](#).

You can view a summary of your monthly billing on [IBM Cloud](#). From the home page, choose **Manage > Account and billing**. To view the usage for each service, click **Usage**.

## Learn more

[Provisioning the Data Product Hub service](#)

Parent topic: [Setting up and administering](#)

# IBM Cloud account security

Account security mechanisms for Data Product Hub are provided by IBM Cloud. These security mechanisms, including SSO and role-based, group-based, and service-based access control, protect access to resources and provide user authentication.

Table 1. Account security mechanisms for Data Product Hub

Mechanism	Purpose	Responsibility	Configured on
<a href="#">Access (IAM) roles</a>	Provide role-based access control for services	Customer	IBM Cloud
<a href="#">Access groups</a>	Configure access groups and policies	Customer	IBM Cloud
<a href="#">Resource groups</a>	Organize resources into groups and assign access	Customer	IBM Cloud
<a href="#">Service level roles</a>	Provide role-based access control	Customer	IBM Cloud
<a href="#">Service IDs</a>	Enables an application outside of IBM Cloud access to your IBM Cloud services	Customer	IBM Cloud
<a href="#">Service ID API keys</a>	Authenticates an application to a Service ID	Customer	IBM Cloud
<a href="#">Activity Tracker</a>	Monitor events related to Data Product Hub	Customer	IBM Cloud
<a href="#">Multifactor authentication (MFA)</a>	Require users to authenticate with a method beyond ID and password	Customer	IBM Cloud
<a href="#">Single sign-on authentication</a>	Connect with an identity provider (IdP) for single sign-on (SSO) authentication by using SAML federation	Shared	IBM Cloud

## IAM access roles

You can use IAM access roles to provide users access to all resources that belong to a resource group. You can also give users access to manage resource groups and create new service instances that are assigned to a resource group.

For step-by-step instructions, see [IBM Cloud docs: Assigning access to resources](#)

## Access groups

After you set up and organize resource groups in your account, you can streamline access management by using access groups. Create access groups to organize a set of users and service IDs into a single entity. You can then assign a policy to all group members by assigning it to the access group. Thus you can assign a single policy to the access group instead of assigning the same policy multiple times per individual user or service ID.

By using access groups, you can minimally manage the number of assigned policies by giving the same access to all identities in an access group.

For more information, see [IBM Cloud docs: Setting up access groups](#).

## Resource groups

---

Use resource groups to organize your account's resources into logical groups that help with access control. Rather than assigning access to individual resources, you assign access to the group. Resources are any service that is managed by IAM, such as databases. Whenever you create a service instance from the Cloud catalog, you must assign it to a resource group.

Resource groups work with access group policies to provide a way to manage access to resources by groups of users. By including a user in an access group, and assigning the access group to a resource group, you provide access to the resources contained in the group. Those resources are not available to nonmembers.

The Lite account comes with a single resource group, named "Default", so all resources are placed in the Default resource group. With paid accounts, Administrators can create multiple resource groups to support your business and provide access to resources on an as-needed basis.

For step-by-step instructions, see [IBM Cloud docs: Managing resource groups](#)

For tips on configuring resource groups to provide secure access, see [IBM Cloud docs: Best practices for organizing resources and assigning access](#)

## Service level roles

---

Service level roles control access to Data Product Hub. Predefined or custom roles can be assigned.

See [Roles and permissions for Data Product Hub](#).

## Service IDs

---

You can create service IDs in IBM Cloud to enable an application outside of IBM Cloud access to your IBM Cloud services. Service IDs are not tied to a specific user. If a user leaves an organization and is deleted from the account, the service ID remains intact to ensure that your service continues to work. Access policies that are assigned to each service ID ensure that your application has the appropriate access for authenticating with your IBM Cloud services.

One way in which Service IDs and access policies can be used is to manage access to the Cloud Object Storage buckets.

For more information, see [IBM Cloud docs: Creating and working with service IDs](#).

## Service ID API keys

---

For extra protection, Service IDs can be combined with unique API keys. The API key that is associated with a Service ID can be set for one-time use or unlimited use. For more information, see [IBM Cloud docs: Managing service IDs API keys](#).

## Activity Tracker

---

The Activity Tracker collects and stores audit records for API calls (events) made to resources that run in the IBM Cloud. You can use Activity Tracker to monitor the activity of your IBM Cloud account to investigate abnormal activity and critical actions, and to comply with regulatory audit requirements. The events that are collected comply with the Cloud Auditing Data Federation (CADF) standard. IBM services that generate Activity Tracker events follow the IBM Cloud security policy.

For a list of events that apply to Data Product Hub, see [Auditing events in Activity Tracker](#).

For instructions on configuring Activity Tracker, see [IBM Cloud docs: Getting started with IBM Cloud Activity Tracker](#).

## Multifactor authentication

---

Multifactor authentication (or MFA) adds an extra layer of security by requiring multiple types of authentication methods upon login. After entering a valid username and password, users must also satisfy a second authentication method. For example, a time-sensitive passcode is sent to the user, either through text or email. The correct passcode must be entered to complete the login process.

For more information, see [IBM Cloud docs: Types of multifactor authentication](#).

## Single sign-on authentication

---

Single sign-on (SSO) is an authentication method that enables users to log in to multiple, related applications that use one set of credentials.

Data Product Hub supports SSO using Security Assertion Markup Language (SAML) federated IDs. SAML federation requires coordination with IBM to configure. SAML connects IBMids with the user credentials that are provided by an identity provider (IdP). For companies that have configured SAML federation with IBM, users can log in to Data Product Hub with their company credentials. SAML federation is the recommended method for SSO configuration with Data Product Hub.

The [IBMId Enterprise Federation](#) describes the steps that are required to federate your identity provider (IdP). You need an IBM Sponsor, which is an IBM employee that works as the contact person between you and the IBMid team.

For an overview of SAML federation, see [IBM Cloud docs: Which SAML federation options exist in IBM Cloud?](#).

---

## Setting up Data Product Hub

After your IBM Cloud account and users are configured in IBM Cloud IAM, the next step is to set up Data Product Hub.

### Required roles to complete this task

IAM Platform role: **Viewer**

IAM Service role: **Manager**

Data Product Hub collaborator role: **Admin**

The IAM Service role of **Manager** is reserved for the Data Product Hub administrator. When the administrator logs in, Data Product Hub is initialized with a catalog.

## Steps for the Data Product Hub administrator

---

Complete the following steps as a Data Product Hub administrator:

- Log in to initialize Data Product Hub. For details, see [Logging in to Data Product Hub](#).
- Select a Cloud Object Storage instance. For details, see [Managing storage](#).
- Add community members and assign collaborator roles. For details, see [Managing the Data Product Hub community](#).
- Create data source connections. For details, see [Connectors for Data Product Hub](#).
- Create business domains. For details, see [Creating custom business domains](#).
- Create custom properties. For details, see [Managing custom properties](#).

Parent topic: [Setting up and administering Data Product Hub](#)

---

## Managing storage

You manage the storage location in Cloud Object Storage for the PDF files for the data contracts. You can designate a target storage location for data extracts that are contained in a data product.

### Required roles to complete this task

Collaborator role: **Admin**

IAM Service role for Cloud Object Storage: **Manager**

## Storage for data contracts

---

Data Product Hub requires an instance of IBM Cloud Object Storage with a dedicated bucket for storing data contracts and for establishing a default project. The data contracts are stored as PDF files that describe the terms and conditions for data products. Either an uploaded file or a public URL for the data contract is required, but only the PDFs require storage space. The maximum file size that can be uploaded is 50 MB.

As the IBM Cloud account owner or Administrator, you provision an instance of Cloud Object Storage. As the Data Product Hub Administrator, you select a Cloud Object Storage instance and create a bucket to store the PDF files for the data contracts and other system files. A bucket is a logical abstraction that provides a container for data. When the bucket is created, the Default Data Product Delivery project is automatically created to permit the delivery of data extracts. If your account already has an existing Cloud Object Storage instance, you can create the bucket in that instance. You need the IAM Service role of **Manager** for Cloud Object Storage to create a bucket.

Do not delete the Default Data Product Delivery project, as it is required for Data Product Hub.

Familiarity with IBM Cloud and Cloud Object Storage is required to configure IBM Cloud Object Storage for Data Product Hub.

- [Create a new Cloud Object Storage instance](#)
- [Use an existing Cloud Object Storage instance](#)

## New Cloud Object Storage instance

To provision an instance of IBM Cloud Object Storage:

1. Log in to IBM Cloud.
2. In the homepage, select **Create resource**.
3. Search for the **Object Storage** tile and select it.
4. Select a plan and give the service instance a name. The Standard plan includes a Free Tier with 5GB of free storage for 12 months.
5. In Data Product Hub, from **Administration > Configurations and settings > Storage**, select the **Cloud Object Storage instance** you created.
6. Click **Create bucket** to create the bucket for your PDFs.

For more information on the IBM Cloud Object Storage instance, see [IBM Cloud docs: Getting started with Cloud Object Storage](#)

## Existing Cloud Object Storage instance

To select an existing Cloud Object Storage instance and create a bucket:

1. From **Administration > Configurations and settings > Storage**, select a Cloud Object Storage instance for Data Product Hub.
2. Click **Create bucket** to create the bucket for your PDFs.

## Restoring a deleted instance of Cloud Object Storage

---

If the Cloud Object Storage instance where you are storing your data contracts is deleted, it is not deleted immediately. Instead, it is scheduled for reclamation in 7 days. You can restore the instance and the data before the 7 day reclamation period. After 7 days, the data is irreversibly destroyed.

**Important:** There is no reclamation period for Cloud Object Storage instances under the Lite plan. You cannot restore a deleted instance under the Lite plan. For Lite plan, the data is permanently deleted when the instance is deleted.

For information on how to restore a deleted Cloud Object Storage instance within the 7 day reclamation period, see [Restoring a resource by using the CLI](#).

## Storage for data extracts

---

For the **Download data extract** delivery method, you designate a connection as the target storage location for data extract files contained in data products. Typically, you use the same file storage location for all data extracts. Consumers can download the data extracts from this location.

The connection for the target storage location for data extracts requires read/write credentials.

To find connectors that support data extracts, both source and target, see [Delivery methods for connectors](#).

## Learn more

---

- [IBM Cloud docs: Getting started with Cloud Object Storage](#)
- [Understanding delivery methods](#)
- [Delivery methods for connectors](#)
- [Download data extract delivery method](#)

Parent topic: [Setting up and administering Data Product Hub](#)

---

## Managing the Data Product Hub community

The Data Product Hub administrator adds users or groups to the community and assigns the appropriate roles. The roles control access to the actions that can be taken on Data Product Hub.

### Required roles to complete this task

Collaborator role: **Admin**

## Adding members or groups

---

To add members to Data Product Hub, they must first be added to the IBM Cloud account and assigned appropriate IAM roles. You can add individual users or user groups as community members. You create user groups in IBM Cloud Identity and Access Management (IAM). Roles that are assigned to groups apply to all members of the group.

To add members to the community:


1. From the navigation menu, click **Administration > Configurations and settings**. Then, click **Manage community**.
2. Click **Add members** and choose whether to add members as **Viewer**, **Editor**, or **Admin**.
3. Select the individual users or access groups from your IBM Cloud account. You can use the *Show* drop-down to organize the complete list of members by **Users** and **Groups**. Then, click **Add** to confirm your selection.

## Assigning roles

---

You must assign a collaborator role to each member of your data community. Each collaborator role has different tools and processes available to them to help fulfill their business goals.

To assign a member to a collaborator role:

1. Select the **Actions menu**  next to the user or group.
2. Assign the **Viewer**, **Editor**, or **Admin** role.

To assign roles more quickly, you can select multiple members and update their roles simultaneously:

1. Check mark all the members who need new collaborator roles.
2. Click the **Role** icon and choose the new collaborator role. Then, click **Save** to confirm your update.
3. View the collaborator role and ensure that all members are assigned to the correct new role.

The roles are defined as:

- **Viewer**: For data product consumers who discover and subscribe to data products. The **Viewer** role provides minimal permissions.
- **Editor**: For data product producers who author, publish, and manage data products. The **Editor** role includes the permissions for **Viewer**.
- **Admin**: For administrators who add users and assign roles and other configuration tasks. The **Admin** role includes permissions for **Viewer** and **Editor**. The **Admin** role is usually assigned to one person who is responsible for managing

users for Data Product Hub.

## Searching for a member or group


---

You can search for a member or a group in your community and open their IBM Cloud Access (IAM) page to review their user details and access policies. When you search for a member, the search results include the Access groups of which they are a member. Select a member or group to open the IBM Cloud Access (IAM) page.

## Removing members from Data Product Hub

---

To remove a user or group from Data Product Hub:

1. Select the Actions menu (  ) next to the user or group that you want to remove.
2. Select **Remove**.

When a member is removed, they can no longer access Data Product Hub. If a group is removed, all members are removed.

## Learn more

---

- [Roles and permissions for Data Product Hub](#)
- [Adding users to the IBM Cloud account](#)
- [IBM Cloud docs: IAM access](#)

**Parent topic:** [Setting up and administering Data Product Hub](#)

---

## Managing business domains

You can create custom business domains to better organize your data products and tailor your data community for your business needs. Business domains, together with subdomains, enhance navigation and the search experience by organizing data under clear, intuitive categories. You can create an unlimited number of domains and subdomains.

### Required roles to complete this task

Collaborator role: **Admin**

To create a custom business domain:


1. From the navigation menu, click **Configurations and Settings**. Then, select **Business domains**.
2. Create a new business domain and provide a name and short description. Creating a new domain also adds a new tile on your Data Product Hub home page.
3. Confirm and add your domain.

## Adding subdomains

---

To optimize your data product management, you can also add subdomains under your main business domains. Subdomains make it easy to organize your data products into detailed, focused categories. Subdomains do not appear on your Data Product Hub home page, but can be accessed under each business domain.

To add a subdomain:

1. From the main business domain, click the Action menu  and select **Add subdomain**.
2. Provide a name and short description of the subdomain. Then, confirm and add the subdomain.


## Deleting domains and subdomains:

---



You can delete business domains and subdomains to improve organization and remove retired data products. Keep your domains up to date to enhance security and optimize storage space.

To delete a business domain or subdomain:

1. Click the Action menu  and select **Delete**.

There are limitations to deleting a business domain or subdomain:

- You cannot delete business domains that are attached to a published data product.
- If a data product draft is attached to a deleted domain, you cannot publish the draft until you select an active domain.

**Parent topic:** [Setting up Data Product Hub](#)

---


## Managing data source connections

You can manage the data source connections that you own. You can edit or delete a connection, or share it with community members.

### Required roles to complete this task

Collaborator roles: **Editor** or **Admin**

To edit, share, or delete your connections:

1. From the navigation menu, choose **Administration > Configurations and settings > Connections**.
2. From the row's **Actions menu** , choose **Edit**, **Share**, or **Delete**.

If no connections are listed, you can create one by clicking **New connection**. For more information, see [Connectors for Data Product Hub](#). Note that platform connections are not available on Data Product Hub.

---

## Deleting a connection

When you delete a connection, any future subscriptions that use the connection will fail.

You cannot delete a connection that is currently used by a published data product. To delete a connection that is currently in use, you must first retire the data products that use the connection. See [Managing data products](#) for instructions on how to retire a data product.


---

## Sharing a connection

You can share connections that you own with community members who can then use them to create and deliver data products. You can share connections with any community member who has the **Editor** or **Admin** roles. You cannot share a connection with a user group.

When you share a connection with a community member, they edit the connection with their personal credentials.

To share a connection:


1. From the **Connections** tab, click the row's **Actions menu**  and select **Share**. If the connection was previously shared, those members are listed.
2. Click **Share with members** to list the community members who are eligible to use the connection.
3. Choose one or more members and click **Save** to share your connection.

When a connection is shared, the member can view the connection information. They cannot view the credentials of any other user, including the functional credentials that are entered by the connection owner when the connection is created.

If the member's role changes to **Viewer**, sharing access for connections is removed.

To remove sharing access:

When you stop sharing for a connection, the user whose access was removed cannot create and publish new data products using that connection. However, existing published data products using the connection can be delivered.


1. Select **Share** from the row's **Actions menu** . The members who are currently sharing the connection are listed.
2. Checkmark the member and choose **Stop sharing access** to remove sharing access.

## Creating a data product with a shared connection

If a connection has been shared with you, the connection will appear in your list of connections. You can use that connection to create and deliver data products.

Before using a shared connection to create and deliver a data product, you must enter your personal credentials.

To enter your credentials for a connection:

1. From the navigation menu, choose **Administration > Configurations and settings > Connections**.
2. Select **Edit** from the row's **Actions menu** .
3. Select the authentication method.
4. Enter your service credentials.

## Learn more

- [Connectors for Data Product Hub](#)

Parent topic: [Setting up and administering Data Product Hub](#)

## Understanding credentials for connections

Understand how credentials are handled in Data Product Hub so that you can effectively manage the credentials for data source connections. Credentials are attached to data source connections to allow secure access to producers when they add items from a data source to a data product. Credentials also provide access to a connection for delivering data products to consumers.

## Credential types

Data Product Hub uses functional and personal credentials to provide secure access to data sources through a connection. The connection owner supplies the functional credentials when they create the connection. The producer supplies the personal credentials when they create a data product.

Credential types

Credential type	Entered by	Purpose	Credential characteristics
Functional credentials	The connection owner enters the credentials when they create a connection. The credentials of the connection owner are used as the functional credentials.	Runs a query on a data source for automated delivery of items in a data product.	<ul style="list-style-type: none"><li>• Need stable (long-lived) credentials to avoid failed data product deliveries due to expired credentials.</li><li>• Need read access to the data source. Data extract targets require read/write access.</li></ul>

Credential type	Entered by	Purpose	Credential characteristics
Personal credentials	Producers with whom the connection was shared must enter their personal credentials to create data products.	<ul style="list-style-type: none"> <li>Validates access to a data source during the creation of the data product.</li> <li>Connects to a data source to allow selection of items for a data product.</li> <li>Runs queries to collect technical metadata and data profile visualizations.</li> </ul>	<ul style="list-style-type: none"> <li>No one can view the functional credentials or the credentials of any other user.</li> <li>When the credential owner accesses the connection to create a data product, their credentials serve as both the functional credentials and the personal credentials.</li> </ul>

## How credentials are created for connections

Producers (users with the **Editor** role) or administrators (users with the **Admin** role) can create connections either directly in Data Product Hub or as a connection asset in a project.

- When you create a connection in Data Product Hub, you become the owner of the connection and the connection is visible only to you unless you share it.
  - Sharing a connection: As the owner of a connection, you can share the connection with other producers or administrators so that they can use the connection to create data products. In order to receive a shared connection, the user must have either the **Editor** or **Admin** role. The recipient of a shared connection can view the connection information. They cannot view the credentials of any other user or the functional credentials for the connection. Each recipient of a shared connection enters their personal credentials to access the connection.
  - Designating a new owner for a connection: The connection owner can designate a new owner for the connection to a member with the **Editor** or **Admin** role. The new owner adds their personal credentials and can use the connection to create data products. Assigning a new owner for a connection doesn't change the functional credentials. However, the new owner becomes responsible for maintaining the functional credentials by rotating them as needed.
- When you create a connection as a connection asset in a project or catalog in Watson Studio or watsonx.ai, the connection is visible to all members of the project or catalog. Project members can view the connection information, but cannot view the credentials for any other user. Each project member enters their own personal credentials to access the connection.

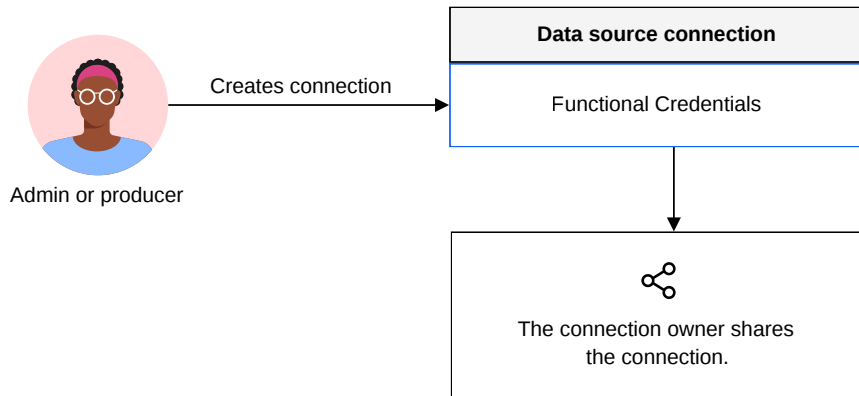
## Flow for credentials when creating data products

The credential flow starts when an administrator or producer creates a connection and their credentials become the functional credentials for the connection. When the connection owner shares the connection, the producer with whom the connection is shared enters their personal credentials and creates a data product. The data product is delivered to subscribers using the functional credentials. Neither the functional or personal credentials are visible to other users.

### Step 1: Administrator or producer creates a connection

The administrator or producer creates a data source connection to create data products and to deliver them to consumers. They become the owner of the connection. The owner's credentials become the functional credentials for the connection. The credential is long-lived and has appropriate permissions to access the items in the data product. The owner shares the connection with other producers or administrators who can use it to create data products.

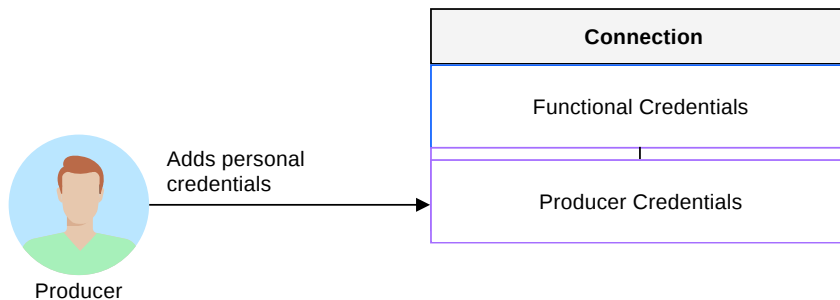
1. Admin or producer creates a connection and shares it



### Step 2: Producer enters personal credentials

The producer adds their personal credentials to the shared connection. Producers cannot view the functional credentials.

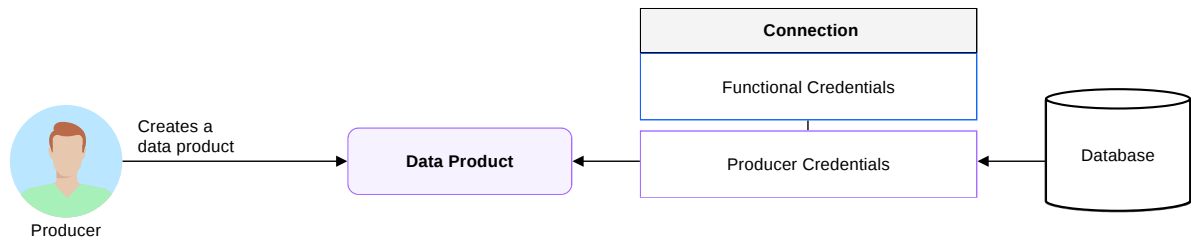
2. Producer adds their personal credential to the shared connection



### Step 3: Producer creates a data product

The producer creates a data product using the shared connection and their personal credentials. The producer's personal credentials are used to access the data source to select items to include in the data product.

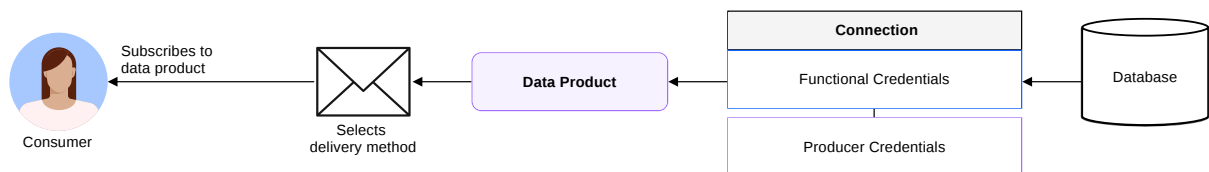
### 3. Producer creates a data product



### Step 4: Consumer subscribes to the data product

The consumer subscribes to the data product and selects a delivery method. Data Product Hub asynchronously delivers the data product using the functional credentials for the connection. The functional credentials are never exposed to the consumer.

### 4. Consumer subscribes to data product



## Best practices

When creating a connection, you supply credentials with broad access to the data source so the tables and other items can be added to a data product. By using broad access and long-lived credentials, you provide the necessary access to producers who create data products. You also limit the number of connections that you need to manage.

However, the permissions provided by the credentials should be narrow, usually read-only. The exception is the credentials for connections to target locations for storing data extracts, which require read/write permissions.

Parent topic: [Managing connections](#)

## Managing custom properties

You can tailor data products to your organization by defining custom properties to extend the default set of properties. You can create custom properties that apply widely to data products or that apply to the items in a data product. The custom properties are listed on the **Additional information** tab for a data product. The custom properties for items in a data product are listed on the **Data product contents** tab when you select an item. Custom properties, together with the default properties, help to clearly define your data products.

You can create up to 1000 custom properties.

### Required roles to complete this task

IAM Platform role: Account owner or **Administrator** for the Data Product Hub service  
IAM Service role: **Manager** for the IBM Cloud Pak for Data service  
Data Product Hub collaborator role: **Admin**

Custom properties provide additional information about data products to further define your data products. Administrators define properties that are relevant to their organization. Producers can edit the values for custom properties when creating a data product. Custom properties are available to all producers in your Data Product Hub community for use in their data products.

Data Product Hub also imports properties defined for data assets when adding data assets from a catalog into your data product. The properties for an imported asset cannot be edited in Data Product Hub.

To create custom properties:

1. Select **Configuration and settings>Custom properties>Customize**.
2. Select the asset type for the custom properties, either **Data product**, for data product wide properties, or **Data product part**, for data product item properties.
3. Add a new group to organize your custom properties. Custom properties must be added to a group.
4. Open the group and select **New property**.
5. Select the basic parameters and advanced settings for each property.

## Define the custom property

Use these parameters to define a custom property:

Parameters for custom properties

Parameter	Description and restrictions
Name	Names must be unique. The character limit is 255.
Unique identifier	Generated identifier which can be edited.
Description	Optional. The character limit is 255.
Property type	Type cannot be changed after a custom property is created. Choose from: <ul style="list-style-type: none"><li>• Text</li><li>• Integer</li><li>• Decimal</li><li>• Date</li><li>• User and user group</li><li>• Predefined values</li></ul>

### Types for custom properties

The property type determines the values that can be entered for the property.

Types for custom properties

Type	Entry options for producers
Text	Text entry
Integer	Whole numbers
Decimal	Decimal numbers
Date	Accepts dates only. Valid date format is mm/dd/yyyy
User and user group	Select from menu. Generates a list of community members.
Predefined values	Select from menu. Generates a dropdown menu containing the specified predefined values.

### Select advanced settings

Use the advanced settings to further define the property:

#### Make searchable and visible for filtering

Enable this option so that producers and users can search for properties using the search field and filters. Search by property is available on the **Home** page, the **Search data products** page, and **My work** page.

Modifying this setting after the custom property is created does not apply the change to existing assets that use the property.


#### Allow multiple values

Enable this option to allow the producer to enter or select multiple values for the same property. If you do not select this option, only one value per property is allowed. This option cannot be edited after the property is saved. For entry fields

such as **Text** and **Integer**, you can add multiple fields. For menus such as **Predefined values** and **User and user groups**, you can select more than one value.

## Editing properties

---

Select **Configuration and settings>Custom properties>Customize>Asset type** to view a list of custom properties by group. Locate the property that you want to edit and select **Edit** from the **Actions menu** .


The following parameters cannot be edited after the property is created:

- **Property type**
- **Allow multiple values**

If you change the setting for **Make searchable**, the property is not automatically updated for existing assets. You must edit the value of the custom property directly on the asset for the change to take effect.

## Deleting properties

---

To delete a property, select **Delete** from the **Actions menu** . If you delete a custom property, the property and its values are removed from all assets that include this property. You cannot undo deletion for a custom property.

## Learn more

---

- [Creating a data product](#)

Parent topic: [Setting up Data Product Hub](#)

---

# Roles and permissions for Data Product Hub

Review the roles and permissions that users need for working with Data Product Hub.

## IAM roles and collaborator roles

---

Data Product Hub users require two types of roles:

- Roles assigned in IBM Cloud, which are called IAM roles
- Roles assigned in Data Product Hub, which are called collaborator roles

As the IBM Cloud account owner or administrator, you assign IAM roles to individual users or to access groups on IBM Cloud using [Manage users and access](#).

The IAM role assignments provide Platform or Service level permissions for IBM Cloud. Any of the IAM Platform roles of **Viewer, Editor, Operator, and Administrator** can be assigned to most users who work with Data Product Hub. The minimum IAM Platform role for working in Data Product Hub is **Viewer** for the users who will be consumers or producers. The exception is the Data Product Hub **Manager**, who must be assigned the IAM Platform **Administrator** role.

The account administrator can delegate a Data Product Hub **Manager** to initialize Data Product Hub by logging in for the first time. The **Manager** also requires other roles, as described in [Delegate a Data Product Hub Manager](#).

## Assigning IAM roles

---

Assign IAM roles in IBM Cloud by navigating to **Manage>Access(IAM)**. You can assign roles to individual users, or create access groups to expedite the assignment of roles to groups of users who require the same permissions.

### Creating access groups

Access groups allow you to assign the same roles and permissions to a group of users, rather than making assignments to individual users. IAM access groups are created and managed entirely on IBM Cloud. You can modify an access group after you create it. You can add and delete members, add and delete policies, and make other modifications as needed. When you modify the policies of an access group, the new policies are immediately applied to all members of the group. When you add a user to an access group, they are assigned the permissions of the group.

Access groups save time when assigning collaborator roles in Data Product Hub. For example, you can create an access group for consumers and one for producers. Then you assign the **Viewer** role to the Consumers group. Assign the **Editor** role to the Producers group. When you add a new user, add them to the appropriate access group.

For instructions on creating access groups in IBM Cloud, see [Setting up access groups](#).

## Delegate a Data Product Hub Manager

Either the IBM Cloud account administrator or their delegate must be the first user to log in to Data Product Hub to initialize it. The IAM Service role of **Manager** can be assigned to the delegated user who is going to log in to initialize Data Product Hub.

The account administrator assigns the **Manager** and other roles to delegate a user who can initialize Data Product Hub. The required roles are described in the following table:

IAM roles for the Data Product Hub Manager

Service	Role level	Role	Action
Data Product Hub	Service	Manager	Initialize Data Product Hub upon initial log in
Data Product Hub	Platform	Administrator	Initialize Data Product Hub upon initial log in
All Account Management services	Platform	Administrator	Initialize Data Product Hub upon initial log in
Cloud Object Storage	Service	Manager	Configure a bucket for storing data contracts
Cloud Object Storage	Platform	Administrator	Configure a bucket for storing data contracts

After logging in to initialize Data Product Hub, the Data Product Hub **Manager** performs the following next steps:

- Creates a Cloud Object Storage bucket for storing data contracts. See [Managing storage](#).
- Adds the account administrator to the community with the **Admin** role. See [Managing the Data Product Hub community](#).
- Adds members to the community with appropriate roles. See [Managing the Data Product Hub community](#).

## Assigning collaborator roles

Data Product Hub requires that all users have a collaborator role. Collaborator roles are assigned by the Data Product Hub Administrator from the **Administration>Configurations and settings>Manage community** page.

Collaborators have one of these roles that provide permissions:

- **Viewer**: Data product consumers who discover and subscribe to data products.
- **Editor**: Data product producers who author, publish, and manage data products. Editor role includes permissions for **Viewer**.
- **Admin**: Administrators who add users and assign roles and other configuration tasks. **Admin** role includes permissions for **Viewer** and **Editor**.

The following table shows the actions that you can complete depending on your collaborator role.

+ indicates that users need to be owners of a subscription or data product to perform the action.

Permissions by role

Action	Viewer	Editor	Admin
Log in to Data Product Hub	✓	✓	✓
View the Data Product Hub home page	✓	✓	✓
Search for published data products	✓	✓	✓
Subscribe to a data product	✓	✓	✓
Send and receive notifying comments	✓	✓	✓
View subscriptions	✓ +	✓ +	✓ +
Publish, edit, and delete data products		✓ +	✓ +



Action	Viewer	Editor	Admin
Manage data products from My work page		✓ +	✓ +
Create data product drafts and versions		✓	✓
Add custom properties to a data product		✓	✓
Accept or reject requests for new data products		✓	✓
Approve access to data products		✓	✓
Create a list of preapproved users		✓	✓
Create connections to data sources		✓	✓
Edit credentials for a shared connection		✓	✓
View the insights dashboard		✓	✓
Add or delete users or groups			✓
Assign and modify roles			✓
Create and delete business domains			✓
Create custom properties			✓

## Learn more

---

- [IBM Cloud docs: IAM access](#)
- [IBM Cloud docs: What is IBM Cloud Identity and Access Management](#)
- [IBM Cloud docs: Setting up access groups](#)
- [Managing the Data Product Hub community](#)

Parent topic: [Setting up and administering Data Product Hub](#)

---

## General administration for Data Product Hub

Understand the general administration information, including activity tracker events, data management, high availability, disaster recovery, and shared responsibilities for Data Product Hub.

## Shared responsibilities

---

The responsibilities for managing the lifecycle and operation of Data Product Hub on IBM Cloud are shared between IBM® and the customer. The management responsibilities are exclusive to IBM, to the customer, or shared between IBM and the customer. The breakdown of responsibilities for operating Data Product Hub follows the shared responsibilities for all IBM Cloud products. For more information, see [Shared responsibilities for using IBM Cloud products](#).

## Administration

---

You can view the general administration topics at the following links:

- [Activity tracker events](#)
- [Managing the Service API key](#)
- [Accessibility](#)
- [Data management](#)
- [Deployment models](#)
- [High availability and disaster recovery](#)
- [Troubleshooting](#)

Parent topic: [Setting up and administering Data Product Hub](#)

---

## Auditing events in Activity Tracker

You can see the events for actions in Data Product Hub in the IBM Cloud Activity Tracker. The Activity Tracker audits events to help identify security incidents, detect unauthorized access, and comply with regulatory and internal auditing requirements.

## Events for Data Product Hub

Activity Tracker events for Data Product Hub

Action	Description
data-product-hub.configuration.create	Initialize the default resources for Data Product Hub.
data-product-hub.configuration.monitor	Get the status of the default resource initialization for Data Product Hub.
data-product-hub.configuration-apikey.rotate	Rotate the API key for Data Product Hub.
data-product-hub.data-product.list	Retrieve a list of all data products.
data-product-hub.data-product.read	Retrieve the details of a data product by a data product ID.
data-product-hub.data-product.create	Create a new data product.
data-product-hub.data-product-draft.list	Retrieve a list of data product drafts.
data-product-hub.data-product-draft.create	Create a new draft for an existing data product.
data-product-hub.data-product-draft.read	Retrieve a draft of an existing data product.
data-product-hub.data-product-draft.update	Update the data product draft identified by the ID.
data-product-hub.data-product-draft.delete	Delete the data product draft identified by the ID.
data-product-hub.data-product-draft-contract-term-document.upload	Upload a contract document to the data product draft.
data-product-hub.data-product-draft-contract-term-document.complete	Complete the upload of a contract document.
data-product-hub.data-product-draft-contract-term-document.read	Retrieve a contract document for a data product draft.
data-product-hub.data-product-draft-contract-term-document.delete	Delete a contract document.
data-product-hub.data-product-draft-contract-term-document.update	Update a contract document.
data-product-hub.data-product-draft.publish	Publish a draft of an existing data product.
data-product-hub.data-product-release.list	Retrieve a list of data product releases.
data-product-hub.data-product-release.read	Retrieve the release of an existing data product.
data-product-hub.data-product-release.update	Update the data product release identified by the ID.
data-product-hub.data-product-release-contract-term-document.read	Retrieve a contract document for a data product release.
data-product-hub.data-product-release.expire	Retire a release of an existing data product.

Parent topic: [General administration for Data Product Hub](#)

## Managing the Service API key

Certain operations in Data Product Hub are performed by a functional admin user and require an API key for authorization. The functional admin user and a Service API key are generated when Data Product Hub is initialized. The Service API key works together with the unique functional admin user to protect your assets. You can rotate the Service API key as needed to help secure your operations.

## About the Service API key

The Service API key provides privileged access to the Data Product Hub instance and assets. It is critical and sensitive and thus requires proper protection. Regular rotation of the Service API key according to your organization's security policies is one way to provide secure operations and to protect your data products from unauthorized use. Security best practices recommend regular rotation of the Service API key.

## About the functional admin user

---

When first initialized, Data Product Hub automatically creates a Service ID with the name *data-product-admin-service-ID-`<catalog_id>`*. This Service ID serves as the functional admin user for Data Product Hub. The functional admin user is a user ID with global rights to read asset information, connect to data sources, and run jobs within Data Product Hub for delivering data products. This user ID is an IBM Cloud Service ID. You can view the Service IDs for your account at [Service IDs](#).

Important: Do not delete, lock, or modify the Service ID for the functional admin user for Data Product Hub.

## Rotating an API key

---

### Required roles

IAM Platform role: **Admin** or Account owner

Data Product Hub collaborator role: **Admin**

When Data Product Hub is initialized, a Service API key is created. This key is stored on IBM Cloud and authorizes runtime operations for Data Product Hub. Periodically, you must generate a new Service API key. The API key might become stale or invalid, or your security policies require that you periodically rotate keys. A typical security policy suggests that you rotate API keys every 30 to 60 days.

To rotate a key:

1. Navigate to **Configurations and settings > Service API key**.
2. Click **Rotate**.

A new key is created to replace the current key. The old key is removed and is not available for use. Account owners and administrators can view the Service API key on IBM Cloud by accessing **Manage > Access(IAM) > API keys**.

## Learn more

---

[Understanding API keys](#)

Parent topic: [General administration for Data Product Hub](#)

---

## Accessibility features in Data Product Hub content and documentation

IBM is committed to accessibility. Accessibility features that follow compliance guidelines are included in Data Product Hub content and documentation to benefit users with disabilities.

Data Product Hub documentation uses the latest W3C Standard, [WAI-ARIA 1.0](#) to ensure compliance with the [United States Access Board Section 508 Standards](#), and the [Web Content Accessibility Guidelines \(WCAG\) 2.0](#).

The Data Product Hub online product documentation is enabled for accessibility. Accessibility features help users who have a disability, such as restricted mobility or limited vision, to use information technology products successfully. Documentation is provided in HTML so that it is easily accessible through assistive technology. With the accessibility features of Data Product Hub, you can do the following tasks:

- Use screen-reader software and digital speech synthesizers to hear what is displayed on the screen. Consult the product documentation of the assistive technology for details on using assistive technologies with HTML-based information.
- Use screen magnifiers to magnify what is displayed on the screen.
- Operate specific or equivalent features by using only the keyboard.

For more information about the commitment that IBM has to accessibility, see [IBM Accessibility](#).

## TTY service

---

In addition to standard IBM help desk and support websites, IBM established a TTY telephone service for use by deaf or hard of hearing customers to access sales and support services:

800-IBM-3383 (800-426-3383) within North America

## More interface information

---

The Data Product Hub user interfaces do not have content that flashes 2 - 55 times per second.

The Data Product Hub web user interfaces rely on cascading stylesheets to render content properly and to provide a usable experience. If you are a low-vision user, you can adjust your operating system display settings, and use settings such as high contrast mode. You can control font size by using the device or web browser settings.

**Parent topic:** [General administration for Data Product Hub](#)

---

## Managing customer data for Data Product Hub

Customer data security is paramount. The following information outlines some of the ways that customer data is protected in Data Product Hub and what you are expected to do to help in these efforts.

- [Customer responsibility](#)
- [IBM's commitment to GDPR](#)
- [GDPR statement that applies to log files](#)
- [Secure deletion from the Data Product Hub service](#)

## Customer responsibility

---

Customers are responsible for ensuring their own compliance with various laws and regulations, including the European Union General Data Protection Regulation (GDPR). Customers are solely responsible for obtaining the advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulations that might affect the customer's business. They are responsible for any actions that are necessary to comply with such laws and regulations. The products, services, and other capabilities that are described here are not suitable for all customer situations and might have restricted availability. IBM does not provide legal, accounting, or auditing advice or represent or warrant that its services or products ensure that customers are in compliance with any law or regulation.

## IBM's commitment to GDPR

---

Learn more about IBM's own [GDPR readiness journey and our GDPR capabilities](#) and offerings to support your compliance journey.

## GDPR statement that applies to log files

---

Pay close attention to data privacy principles when you select a data product. Processing of PI is governed by vigorous legal requirements and is only allowed if it is based on an explicit legal basis. These regulations mandate that PI is processed only for the purpose for which it was collected. No other processing in a manner that is incompatible with this initial purpose is permissible. For these and other constraints these regulations place on your use of PI, we highly recommend that you do not use "real" PI in your data product unless it is allowed or permissible. You may substitute real PI using test data that is available on the public sphere.

## Secure deletion

---

Data Product Hub does not directly store any personally identifiable information and data. All customer data is stored in customer-managed storage.

However, anyone that has personally identifiable information and data (PII) stored as part of using the Data Product Hub service, has the right to obtain erasure of that data from the controller without undue delay. The controller has the obligation to

erase personal data without undue delay where one of the following conditions exists:

- There is PII data that is stored in the Data Product Hub service
- User email address and full name are stored as metadata related to the Data Product Hub repository assets.
- User provided service credentials.

Repository asset content can be securely deleted by performing one of the methods for permanently deleting personal data.

### Options for permanently deleting personal data

To delete personal data permanently, remove the entire Data Product Hub service instance from IBM Cloud. This is possible by sending a deprovisioning request through various channels, such as the IBM Cloud UI, CLI, or REST API.

For the Data Product Hub service, personally identifiable information and data are removed completely from all data sources, including backups, after 30 days.

## Learn more

---

- [How do I know that my data is safe?](#)

Parent topic: [General administration for Data Product Hub](#)

---

## Deployment models for Data Product Hub

Data Product Hub is hosted as a secure multi-tenant solution. Multi-tenant databases are used for all metadata storage in Data Product Hub. For data products, the metadata about the data source is collected and stored in the multi-tenant databases. Customer-owned data is not copied or moved from the source location.

Data Product Hub is deployed in the Dallas (us-south) region.

Uploaded PDF files for data contracts, data profile information, and project system data are stored in the Cloud Object Storage instance that is provisioned and owned by the customer. See [Managing security and compliance with IBM Cloud Object Storage](#) for deployment models for Cloud Object Storage.

## Delivery methods and data isolation

---

- Open URL or Download: When data items are accessed by using the Open URL and Download delivery methods, source data does not flow through Data Product Hub. Customers receive a link to access content or download files directly from the source.
- The Flight service: For the Flight service delivery method, consumers receive a code snippet to connect to a specific data asset using the Flight service. Data moves through the Flight service data plane on the solution to the user's application. Data access through the Flight service depends on both the source access controls and the connection credentials in Data Product Hub. The connection that is used for Flight service delivery must be configured with read-only credentials.
- Download data extract: Download data extract has two stages for moving data. The first stage is extracting the data and generating a CSV file. A Data Refinery job copies data from the source system to the target system. Data flows through the Analytics engine on the data plane. The second stage is downloading the extracted CSV file. Data flows directly from the target system to the consumer and does not move through Data Product Hub.

## Deleting data from Data Product Hub

---

You can remove data products, data sources, and the Data Product Hub instance.

To remove a published data product and its data:

Retire the data product from the **My Work** page.

To remove access to all data for a data source:

1. Retire all published data products with items that use the connection to the data source.
2. Delete the connection from **Configurations and settings > Connections**.

To remove all data from Data Product Hub:

Delete your Data Product Hub instance from your IBM Cloud account. Your metadata will be maintained for 30 days, after which it will be permanently deleted.

## Learn more

---

[What is multi-tenant?](#)

**Parent topic:** [General administration for Data Product Hub](#)

---

## Understanding high availability and disaster recovery

High availability (HA) is a core discipline in an IT infrastructure to keep your applications running, even after a partial or full site failure. The main purpose of high availability is to eliminate potential points of failure in an IT infrastructure. Disaster recovery involves a set of policies, tools, and procedures for returning a system, an application, or an entire data center to full operation after a catastrophic interruption. It includes procedures for copying and storing an installed system's essential data in a secure location, and for recovering that data to restore normalcy of operation.

### High availability

---

Data Product Hub is highly available within the Dallas region.

#### Responsibilities

To find out more about responsibility ownership for using IBM Cloud products between IBM and the customer, see [Shared responsibilities for IBM Cloud products](#).

#### What level of availability do I need?

You can achieve high availability on different levels in your IT infrastructure and within different components of your cluster. The level of availability that is right for you depends on several factors, such as your business requirements, the service level agreements (SLAs) that you have with your customers, and the resources that you want to expend.

#### What level of availability does IBM Cloud offer?

The level of availability that you set up for your cluster impacts your coverage under the IBM Cloud high availability service level agreement terms.

Service level objectives (SLOs) describe the design points that the IBM Cloud services are engineered to meet. Data Product Hub is designed to achieve an availability target of 99.9%.

The SLO is not a warranty and IBM does not issue credits for failure to meet an objective. Refer to the SLAs for commitments and credits that are issued for failure to meet any committed SLAs. For a summary of all SLOs, see [IBM Cloud service level objectives](#).

#### Locations

For more information about service availability within regions and data centers, see [Service and infrastructure availability by location](#).

### Disaster recovery

---

If a failure occurs, a failover design is established to keep your resources running without action on your part. For more information, see [How IBM Cloud ensures high availability and disaster recovery](#) to learn more about the high availability and disaster recovery standards in IBM Cloud. You can also learn more about [Service Level Agreements](#).

## Responsibilities

To find out more about responsibility ownership for using IBM Cloud products between IBM and the customer, see [Shared responsibilities for IBM Cloud products](#).

## Disaster recovery strategy

Data Product Hub is a highly available, regional service that runs in the Dallas (us-south) region. Data Product Hub exists in multiple availability zones with no single point of failure. The configuration data that is associated with your instance of Data Product Hub is backed up.

IBM Cloud has business continuity plans in place to provide for the recovery of services within hours if a disaster occurs.

Data Product Hub provides mechanisms to protect your data and restore service functions. Business continuity plans are in place to achieve targeted recovery point objective (RPO) and recovery time objective (RTO) for the service.

**Parent topic:** [General administration for Data Product Hub](#)

---

# Troubleshooting

Review troubleshooting tips for Data Product Hub.

---

## Resolving login errors for Data Product Hub

If you receive an error when logging in, you can try one of the following solutions:

### Troubleshooting login errors

Error message	Possible Causes	Solutions
Entitlement is not available	<ul style="list-style-type: none"><li>• Data Product Hub is not provisioned in the IBM Cloud account.</li><li>• You are working in the wrong IBM Cloud account or region.</li><li>• You have insufficient IAM permissions for the IBM Cloud account.</li></ul>	Check the following: <ul style="list-style-type: none"><li>• Data Product Hub is provisioned in the account.</li><li>• You are working in the correct account and region. You can change the account or region by using the switchers located in the header.</li><li>• You have the correct IAM permissions on IBM Cloud. Contact your administrator to update your IAM roles.</li></ul>
Initialization failed	You are not a member of Data Product Hub.	Contact your Data Product Hub administrator to add you as a community member.

---

## IBM Cloud Object Storage tips

You must have access to an instance of Cloud Object Storage to create a data source connection to Cloud Object Storage, to access a project, or to upload a PDF for a data contract.

Follow these tips to ensure you have access to Cloud Object Storage:

- You can either provision an instance of Cloud Object Storage or the account administrator can grant you access to an existing instance using IAM for the IBM Cloud account.
- When creating a data source connection to Cloud Object Storage, create the Cloud Object Storage credentials with the Hash-based Message Authentication Code (HMAC) option. For more information, see [Using HMAC credentials](#).
- For information about storage for data contracts, see [Storage for data contracts](#).
- For information about creating a connection to IBM Cloud Object Storage, see [IBM Cloud Object Storage connection](#).

## Cannot add items from a catalog or project to a data product

---

- **Symptom:** When creating a data product from a project or catalog, you select assets but they are not added to the data product.
- **Cause:** When you access a catalog or project from Data Product Hub, you will see a list of assets. Not all of the listed assets can be added to a data product. Assets must be from a connected data source in order to be added to a data product. Assets that have been uploaded explicitly to a catalog cannot be added to a data product.
- **Solution:** Select connected data assets from a supported data source for your data product.

**Parent topic:** [General administration for Data Product Hub](#)

---

## Discovering data products

With Data Product Hub, you can browse and search for data products for your business needs. From the home page, you can view details for a data product, subscribe to a data product, and track your subscriptions.

This video provides a visual method to learn the concepts and tasks in this documentation.

Only published data products are available for subscriptions. Before you can discover and subscribe to a data product you must set up the necessary storage requirements and a data producer must publish the data product. For details, see [Managing storage](#) and [Publishing data products](#).

When you find a suitable data product, click the tile to view the details for the product and subscribe to the product. Some data products require approval and are delivered only if approved.

If you don't see a data product that suits your needs, you can submit a request for a new data product. Include your business justification in the request.

## Learn more

---

- [Finding a data product](#)
- [Subscribing to a data product](#)
- [Subscribing to a data product that requires approval](#)
- [Requesting a new data product](#)
- [Flight client example for accessing a data product](#)

**Parent topic:** [Documentation for Data Product Hub](#)



---

## Managing your task inbox as a consumer

As a consumer, your data product requests are listed as tasks which are managed in your **Task inbox**. Use the **Task inbox** to monitor approval status and comments for your requests.



### Monitoring a request for access to a data product that requires approval

---

You can monitor the progress of your request for approval in the *Requested by you* list in your **Task inbox**.

After you submit a request for access to a data product that requires approval, it is delivered to all producers for review. Producers can accept or reject the request.

To view your data product request:



1. From the navigation menu , select **Task inbox**. The **Requested by you** tab lists your requests and organizes them as **In progress** or **Completed**.
2. To further organize your requests, click the **Filter** icon  and indicate your advanced filters.
3. Click **See all activities** to track status changes and content revisions.

### Monitoring a request for a new data product


---

After you submit a request for a new data product, it is delivered to all producers for review. An approver can accept or reject the request, or enter a question or comment.


To view your data product request:

1. From the navigation menu , select **Task inbox**. The **Requested by you** tab lists your requests and organizes them as **In progress** or **Completed**.
2. To further organize your requests, click the **Filter** icon  and indicate your advanced filters.
3. Click **See all activities** to track status changes and content revisions.

#### Exchanging comments for new data product requests

You can exchange comments and questions about your requests for new data products. After a data product producer has accepted your request, they can send you a notifying comment to ask for more information. You receive the notification under the **Notification bell** icon  or by email (or both), depending on your preferences.

Notifications generate the following automatic workflow:

1. Creates a new task in your **Task inbox** and marks any previous tasks as completed.
2. Sends a notification either under the **Notification bell** icon  or by email (or both) to you. You can set your preferences for how to receive notifications.
3. Stores the comment in the feed for your **Task inbox**. Click **See all activities** to view the history of the comments for the task.

You can respond to the comment by clicking **Comment and notify** from the task. The producer who is working on your request receives a notification of your response. When the producer responds again, you will receive a link to the new task that contains the response. You must access comments from the most recent notification, as older notifications are marked as completed.

#### Obtaining the URL for the new data product

When the producer marks the new data product request as completed, you can obtain the URL from the task in the *Requested by you* > *Completed* list in your **Task inbox**.

## Learn more

---

- [Requesting a new data product](#)
- [Managing your task inbox as a producer](#)
- [Managing your notifications](#)

Parent topic: [Getting a data product](#)

---

## Requesting a new data product

If consumers do not discover a suitable data product to support their business needs, they can request a new data product. Consumers can specify the data requirements, delivery method, and terms and conditions of usage of their data product. After submitting a request, producers review the request details and create the data product. Completed data products are then delivered to the consumer.

### Required roles to complete this task

Collaborator roles: **Viewer**, **Editor**, or **Admin**

To request a new data product, log in to [Data Product Hub](#) with your credentials and complete the following steps:

1. From the Data Product Hub home page, select **Request a new data product**.
2. Edit and enter a name for your request. If no name is entered, a default name with the request date is populated.
3. Click **Select a business domain** and choose the best industry for your data.
4. Complete the following fields on the **Content requirements** tab. Producers review this information to create the data product to meet the consumer's business needs.
  - **Overview:** Describe what you need the data product to contain.
  - **Business justification:** Describe why you need the data product to support your business needs.
  - **Key features:** Describe what metrics to include and how to format the data.
  - **Delivery method:** Select a preferred delivery method. This is how you will receive your data product, although data products are sometimes delivered by using a different method due to security risks, privacy concerns, or resources available.
  - **Due date:** Specify a date that you need the data product by.
  - **Sample:** Indicate how you want your data to be structured and visually look. If possible, attach a sample for the producer to reference.
5. Complete the security, privacy, and licensing questions on the **Data contract requirements** tab. These questions determine what kind of data is included (sensitive vs public information), who can see the data, and the quality of the data.
  - **Privacy requirement:** Indicate whether you need access to sensitive information and if you have permission to access such information.
  - **Sharing requirement:** Indicate whether you plan to share your data product and if your shared users are internal or external to your organization.
  - **Data quality requirements:** Describe any conditions or standards that you need the data product to meet. For example, you can request the data product to meet a minimum criteria of accuracy, completeness, and uniqueness.
  - **Data refresh requirement:** Indicate whether you want your data product to have recurrent refreshes and schedule the frequency of refreshes.
6. Click **Submit** to confirm your request.

---




## Monitoring a data product request

After a data product request is submitted, it is delivered to all producers for review. Based on the consumer and the request details, producers can accept or reject the request. If the producer needs more information, they can send a notification with a

comment or question. You can track the status of your request in your **Task inbox**.

When the data product that you requested is published, the producer enters the URL in the task and clicks **Complete**. You can access the link from the **Completed** list in your **Task inbox**.

To view your data product request:

1. From the navigation menu , select **Task inbox**. The **Requested by you** tab lists your requests and organizes them as **In progress** or **Completed**.
2. To further organize your requests, click the **Filter** icon  and indicate your advanced filters.
3. Click **See all activities** to track status changes and content revisions.
4. If the producer needs more information, they will send you a notifying comment. Links to comments are sent under the **Notification bell** icon  or by email (or both).

## Learn more

---

- [Finding a data product](#)
- [Subscribing to a data product](#)
- [Subscribing to a data product that requires approval](#)
- [Managing your notifications](#)
- [Managing your task inbox as a consumer](#)

Parent topic: [Documentation for Data Product Hub](#)

---

## Searching for data products

Data Product Hub provides multiple ways to find a data product that is suitable for your needs. You can search by keyword or business domain.

### Required roles to complete this task

Collaborator roles: **Viewer**, **Editor**, or **Admin**

To open Data Product Hub:

1. Enter <https://datapatform.cloud.ibm.com/dpx?context=dph> in your browser.
2. Log in with your IBMid.

The landing page displays sample data products and the most recently published data products.

## Searching for a data product

---

You can enter a keyword or search string to retrieve relevant data products. At least 3 alphanumeric characters are required.

1. Type your search terms or characters in the search field and press Return.
2. Optional: You can select a suggested search term.
3. Click a tile to view the details for a data product.

If you are ready to subscribe, click **Subscribe**.

You can customize your searches with these techniques:

- [Searching for the start of a word](#)
- [Searching for a part of a word](#)
- [Searching for a phrase](#)
- [Searching for multiple alternative words](#)

## Searching for the start of a word

To search for words that start with a letter or letters, enter the first 1-3 letters of the word. If you enter only one letter, words that start with that letter are returned. If you enter two or three letters, words that start with those letters are prioritized over the words that contain those letters. For example, if you search for **i**, you get results like **initial** and **infinite**, but not **definite**. If you search for **in**, you also get results that contain **definite** ranked lower in the results list.

## Searching for a part of a word

To search for partial word matches, include more than 3 letters. For example, if you search for **conn**, you might get results like **connection** and **disconnect**.

Only the first 12 characters in a word are used in the search. Any search terms that you enter that are longer than 12 characters are truncated to the first 12 characters.

Searches for partial words don't work in the description fields.

## Searching for a phrase

To search for a specific phrase, surround the phrase with double quotation marks. For example, if you search for "**customer data**", your results contain exactly that phrase.

You can include a quoted phrase within a longer search string. For example, if you search for **credit card "customer data"**, you might get results that contain **credit card**, **credit**, **card**, and **customer data**.

When you search for a phrase in English, natural language analysis optimizes the search results in the following ways:

- Words that are not important to the search intent are removed from the search query.
- Phrases in the search string that are common in English are automatically ranked higher than results for individual words.

For example, if you search for **find credit card interest in United States**, you might get the following results:

- Matches for **credit card interest** and **United States** are prioritized.
- Matches for **credit**, **card**, **interest**, **United**, and **States** are returned.
- Matches for **in** are not returned.

## Searching for multiple alternative words

To find results that contain *any* of your search terms, enter multiple words. For example, if you search for **machine learning**, the results contain the word **machine**, the word **learning**, or both words.

# Browsing by business domain

---

When a data producer publishes a data product, they assign a business domain and an optional subdomain. You can narrow your search by choosing a business domain.

To search by business domain:

1. Choose a business domain to view the data products for that domain.
2. From the list of data products for a domain, choose a subdomain to further direct your search.
3. Click a tile to view the details for a data product.

If you are ready to subscribe, click **Subscribe**.

# Viewing data product details

---

Click a tile to view the details for a data product. You can view these details:

- The business domain that was assigned by the publisher
- The publisher's name, the version number, and the published date

- The **Data product contents** provides a description of the contents of the data product that was provided by the publisher.
- The **Recommended usage** provides the business domain and examples for how the data can be used.
- The **Data contract** includes the terms and conditions for using the data product.

When you locate the best data product for your needs, subscribe to the data product! For subscribing instructions, see [Subscribing to a data product](#).

## Learn more

---

- [Subscribing to a data product](#)
- [Viewing your subscriptions](#)

**Parent topic:** [Getting a data product on Data Product Hub](#)

---

## Subscribing to a data product

With Data Product Hub, you can browse and subscribe to data products for your business needs. You can view your subscriptions on **My subscriptions**.

## Placing a data product subscription

---

### Required roles to complete this task

Collaborator roles: **Viewer**, **Editor**, or **Admin**

From the Data Product Hub home page, you can discover data products published by your community or search for specific data products. When you are ready to subscribe, you can click **Subscribe** from the data product tile. If you need more information to help you decide, click the product tile and view the product details.

When you subscribe to a data product, you must indicate your preferred delivery method for each item in the data product and agree to the data contract. After your subscription request is completed, you can access each item using your preferred delivery method.

Watch this video to see an overview of how Data Product Hub can enable data consumers to discover, understand, and subscribe to data products.

This video provides a visual method to learn the concepts and tasks in this documentation.

## Reviewing the data product contents

You can review the product details of a data product to determine whether a data product supports your business needs.

Click the data product tile to review the following product details:


- Domain and publisher:
  - The business domain that was assigned by the publisher.
  - The publisher's name, the version number, and the published date.
- Data product contents:
  - Provides a brief description of the item and its origin.
- Recommended usage:
  - Provides a description for an example application of the data product.
- Data contract:
  - Provides the terms and conditions for using the data product, which can be obtained either from a URL or from an uploaded PDF file.

## Subscribing to a data product

After reviewing the product details, complete the following steps to complete your subscription request:

1. From the data product contents page, click **Subscribe** to open the **Subscriptions details** pane.
2. Choose the delivery method for each item in the data product.
3. Agree to the terms and conditions listed by the data contract.
4. Optional: Provide a brief description of your intended use of the data product, the business needs you want to address, and the outcome you hope to achieve.

## Accessing a data product

Data products are accessed differently depending upon the delivery method. You can check a data product's delivery status in your **My subscriptions** page. To ensure that all subscriptions reflect the latest updates, click the **Refresh status** icon .

The download links expire after 7 days. If your download link expires before you access your data product, you can refresh the subscription to receive a new link.

To access the data product:

1. Open the **My Subscriptions** page and locate your data product.
2. Access your data product according to the delivery method used:
  - For Download, download your data product from a connection to your local directory.
  - For Open URL, access your data product by using the provided link. For some data products, you can view the content in the subscription window. Otherwise, you can view the content in a new browser tab.
  - For Data extract, download your data product as a file extract from a target connection.
  - For Live access with Flight service, run the provided code snippet to retrieve your data product.

For more details about the available delivery methods, see [Working with delivery methods](#).

## Viewing and managing your subscriptions

---

You can view your subscriptions on **My subscriptions**. You can view the following details of your subscriptions:

- The subscription number
- The data product name, description, and version number
- The subscription date and subscription status
- A list of items in the data product with delivery links to access each item

**Parent topic:** [Getting a data product on Data Product Hub](#)

---

## Subscribing to a data product that requires approval

Some data products are restricted and require approval from an approver who is designated by the producer. When you subscribe to a restricted data product, you provide a business justification and a use case that are reviewed by the approver. If your subscription is approved, the data product is delivered to you. If the subscription is rejected, you can submit a new subscription with a new business justification.

#### Required roles to complete this task

Collaborator roles: **Viewer**, **Editor**, or **Admin**

Data products that require approval before delivery are identified by **Access level: Requires approval**.

To subscribe to a data product that requires approval:

1. From the **Subscription details** pane, choose a preferred delivery method for each item in a data product. If only one delivery method is available for the item, it is automatically selected for you.
2. Describe the justification for why you need the data product to support your business needs. The business justification is required. The approver reviews the business justification as part of the approval process.
3. Describe an optional use case to further support your request.
4. Read and agree to the data contract. Then, click **Subscribe** to submit your request.

You can monitor the progress of your request for approval in the **Requested by you** list in your **Task inbox**. If the subscription is approved, you can monitor the delivery status in **My subscriptions**. If the subscription is rejected, you can submit a new subscription with a new business justification.

**Parent topic:** [Getting a data product on Data Product Hub](#)

---

## Flight client example for accessing a data product

After you have subscribed to a data product using the Flight service as the delivery method for one or more items, you can programmatically access the data using an Arrow client. Arrow libraries are available for C, C++, C#, Go, Java, JavaScript, Julia, MATLAB, Python, R and Ruby. See [Apache Arrow](#) for instructions on installing the libraries for each language. This topic provides a Python example for accessing the data in a data product from an Arrow client.

When you subscribe to a data product and select delivery by the Flight service, you receive a Flight URL and Descriptor that are used to access the data product. You can download the Flight URL and Descriptor from the subscription tile that is located under **My subscriptions**.

The Flight URL points to the external route of the Flight service for your environment. The Descriptor contains an asset ID and a catalog ID used to connect to a data source to deliver the items in a data product.

---

## Example for accessing a data product with a Python Flight client

Follow these steps to access a data product with a Flight client in Python:

1. Import the required libraries

Import the **Flight Python** libraries together with the **request** and **json** libraries which are used to make REST API requests.

```
from pyarrow import flight
import requests
import json
```

2. Define an authentication handler

```
class TokenClientAuthHandler(flight.ClientAuthHandler):
    def __init__(self, token):
        super().__init__()
        strToken = str(token)
        self.token = strToken.encode('utf-8')
    def authenticate(self, outgoing, incoming):
        outgoing.write(self.token)
        self.token = incoming.read()
```

```
def get_token(self):
    return self.token
```

### 3. Authenticate with Data Product Hub by using the REST API

The following example authenticates with Data Product Hub using the authentication API. See [Authentication](#).

```
readClient = flight.FlightClient(
    'grpc+tls://api.dataplatform.cloud.ibm.com:443',
    override_hostname='api.dataplatform.cloud.ibm.com',
    disable_server_verification=True)

response = requests.post('https://iam.cloud.ibm.com/identity/token',
    {'grant_type': 'urn:ibm:params:oauth:grant-type:apikey', 'apikey': API_KEY}).json()
token = 'Bearer ' + response['access_token']
readClient.authenticate(TokenClientAuthHandler(token),
    options=flight.FlightCallOptions(timeout=5.0))
```

The variable is defined as follows:

**API\_KEY** is an API key generated for you in IBM Cloud in [API keys](#).

### 4. Initialize the Flight client

```
flightDescriptor = flight.FlightDescriptor.for_command(json.dumps(DESCRIPTOR))
flightInfo = readClient.get_flight_info(flightDescriptor)
```

The variable is defined as follows:

**DESCRIPTOR** is the Flight descriptor in Python copied from the subscription tile for the data product.

### 5. Read the data from the table and load into Pandas

```
for endpoint in flightInfo.endpoints:
    reader = readClient.do_get(endpoint.ticket)
    table = reader.read_all()
```

Your data is stored in the **table** variable and you can now work with the data in your application. For example, you can load the table into a Pandas dataframe using **table.to\_pandas()** to work with the data in Pandas.

```
table.to_pandas()
```

Parent topic: [Getting a data product](#)

---

## Publishing data products

With Data Product Hub, you can create data products from your assets and publish them to share with the data community. Published data products are available for access by qualified consumers. You manage your data products on the **My work** dashboard.

### Required roles

Collaborator role: **Editor**

This video provides a visual method to learn the concepts and tasks in this documentation.



## Preparing data assets for a data product

---

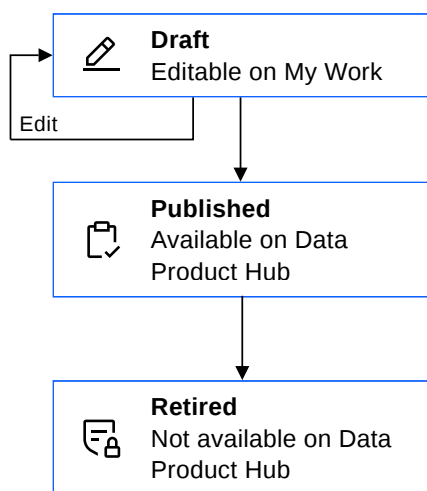
Data products contain curated data assets that preferably completed a Privacy Impact Assessment. Any Personal Identifiable Information (PII) must be removed.

## Data product states

---

A data product moves through several states that indicate whether it is available to consumers on Data Product Hub. The states for a data product are: Draft, Published, and Retired.

The following diagram shows the flow of a data product through each state and the actions that are available to you for each state.



### Draft

While you are authoring a data product, it is in **Draft** state. A draft data product can be edited as needed. You can add items, set the business domain and delivery methods, describe key features and the origin of the data, and add a data contract. You can track versions of the data product by incrementing the version number for each edit. Drafts are listed on the **My work** dashboard and can be edited only by the data product owner. A data product in the draft state is not available to the community.

## Published

Published data products are available to members of your data community who are listed on the **Manage community** page. You can restrict access by indicating that the data product requires an approval before delivery. See [Editing the access level](#). You can always change the access level of published data products, but the contents of the data product cannot be changed after publishing.

Data products remain available on Data Product Hub until you retire them. Data consumers can only find and subscribe to published data products. See [Discovering data products](#).

## Retired

Retired data products are not available for new subscriptions by consumers. They cannot be edited. They are not listed on Data Product Hub and do not appear in search results. Consumers who subscribed to the data product before it was retired can continue to work with it.

# Managing the lifecycle of data products

---

On the **My work** dashboard, you can view all your data products organized by their state. From **My work**, you can edit drafts, publish completed data products, and retire data products that are no longer needed. For details, see [Managing the lifecycle of data products](#).

## Learn more

---

- [Managing the lifecycle of data products](#)
- [Creating a data product from a URL](#)
- [Creating a data product directly from a source](#)
- [Creating a data product from a catalog](#)
- [Creating a data product from a project](#)
- [Creating a data product from SQL](#)
- [Creating a data product that requires approval](#)
- [Creating data source connections](#)

Parent topic: [Documentation for Data Product Hub](#)

---

# Managing your task inbox as a producer

You can manage your assignments and tasks in your **Task inbox**. As a producer, you receive data product requests to complete and access requests for data products that require approval to review.

## Approving requests for access to data products

---

Access requests for data products that require approval can be assigned to specific approvers or claimed by producers. Approvers must have either the **Admin** or **Editor** collaborator role.

To claim a task:

1. From the navigation menu, click **Task inbox** to see your open and completed tasks.
2. You can claim an open task to indicate that you are working on it. If necessary, you can also unclaim a task so that another approver can claim it.

To approve or reject a subscription:

For each task, review the request and choose **Approve** or **Reject**.

If you reject a request, enter a comment to explain to the requester why it was rejected. For rejected requests, the requester can submit a new subscription for the data product with a new business justification.

## Documenting data product activity

---

To document your progress on a requested data product as a producer:

- Select **Assigned to you** to view all data product requests that are assigned to you. Requests are organized as **Open** and **Completed**.
- Click **Claim task** to assign a data product request to yourself. This sends a status update to the requester that a producer viewed and claimed the request.
- Review the request details and **Accept** or **Reject** the request. For further questions or clarifications, you can leave messages to the consumer in the feed by using the *Leave a comment* field.

## Filtering tasks

---

You can narrow down tasks with the following filters:

- **Due date:** Identifies tasks that are overdue or nearing their due date. Tasks become **At risk** when they are not completed 48 hours before their due date.
- **Ownership:** See which tasks are not assigned or claimed.
- **Task type:** Displays access requests and data product requests still awaiting approval. Data Product Hub only uses **Approval**.
- **Request type:** Separates tasks between data product requests and access requests for data products requiring approval.

## Learn more

---

- [Managing your task inbox as a consumer](#)
- [Managing your notifications](#)

Parent topic: [Publishing a data product](#)

---

## Managing your Insights dashboard

### Required roles to complete this task

Collaborator roles: **Editor** or **Admin**

You can use the Insights dashboard to track analytics about data products, monitor important tasks, and manage the data community's resources. This tool helps enhance productivity by offering the most current information on open tasks and delivery statuses, helping ensure that data producers stay notified about upcoming deadlines and data shares. Data administrators can also oversee their community resources and request more resources if necessary.

If you are a data administrator, your dashboard showcases insights for your entire data community.

If you are a data producer, your dashboard showcases insights for your data products. You also receive key insights on how data consumers interact with your data products, your most popular data products, and trends in your data community activity.

## Monitoring your insights

---

To access your Insights dashboard, click **My work** from the home page. Then, select the **Insights** tab.

You can customize your dashboard results by time interval. Choose a preselected time interval or click **Custom** and specify a time interval.

### Understanding your insights

Tile	Description for Admin	Description for Producer
------	-----------------------	--------------------------

Tile	Description for Admin	Description for Producer
<b>Data products</b>	You can view the total number of data products in your community, organized by published and retired status. Use this data to stay up to date on all data products and track the community's growth in publishing new data products. Select a content status to sort the community's data products.	You can view the total number of data products created by you, organized by draft, published, retired status. Use this data to track your progress, and help ensure that data products are complete and accurate before publishing. Select a content status to sort your data products.
<b>Total data shares</b>	You can monitor the total number of data shares that are made to your community's data products. With live updates, you can track the growth of your community's data shares and observe any patterns in data share activity. Make sure that your community has sufficient resources based on your service plan. To verify your available resources, see <a href="#">Data Product Hub service plans</a> and upgrade if your community requires more resources.	You can see the total number of data shares that are made to your data products. A data share is counted when a consumer subscribes to and accesses your data product. You can track your data shares to analyze your content performance and observe your consumers' behavior.
<b>Tasks</b>	You can view and manage all open tasks in your community, based on urgency level. Monitor this data to stay informed on all upcoming deadlines, and ensure that data products are completed and delivered on time. You can select an urgency level to sort your community's tasks.	You can view all the tasks assigned to you, organized by urgency level. You can select an urgency level to sort your tasks, and identify all upcoming and overdue tasks.
<b>Delivery status</b>	You can track the delivery status of your community's subscriptions. Verify that your community's data products are successfully delivered, and make sure to check any unsuccessful or partial deliveries. You can select a delivery status to filter all subscriptions.	You can view the delivery status of your data products. By tracking the delivery status, you can help ensure that data consumers receive their subscriptions successfully. To resolve unsuccessful or partial deliveries, make sure that the data consumer has all the required credentials and connections.

## Learn more

[Managing the lifecycle of data products](#)

Parent topic: [Publishing a data product](#)

## Creating a data product

There are several ways to create and publish a data product. Based on the consumer's request, you can choose how to create the data product. Then you can choose a suitable delivery method, business domain, access level, and other elements for the data product.

To learn about best practices for handling sensitive data and data protection rules, see [Best practices for creating a data product](#).

### Required roles to complete this task

Collaborator role: **Editor**

## Task 1: Selecting a creation method

Select a suitable method to create your data product and complete the unique steps for each method. You can add assets from projects in your data community and from your IBM Cloud platform.

The following table lists the methods to create a data product:

## Methods for creating a data product

Method	Description
<a href="#">From a catalog</a>	Add assets to a data product from a catalog in IBM Knowledge Catalog. A supported data source connection is required. Assets are added by using metadata.
<a href="#">From a customizable query</a>	Consumers can indicate the specific parameters that they need, which act as custom queries that generate the data product. Data products with customizable queries can be reused more frequently.
<a href="#">From a query</a>	Use a SQL query to generate connected data assets for a data product. The query is added as an SQL asset to a project.
<a href="#">From a complex query</a>	Use a complex SQL query to work with temporary and transient tables from a Snowflake connection. The query is added as an SQL asset to a project.
<a href="#">From a project</a>	Add assets to a data product from a project. A supported data source connection is required. Projects are used for data products containing SQL queries.
<a href="#">From a source</a>	Connect to a data source and add stored files to create a data product.
<a href="#">From a URL</a>	Create data products by using one or more public URLs. Use if you do not have a secure connection to the data source.

## Task 2: Completing your data product

After you create a data product draft according to your chosen method, complete the following fields before publishing your data product.

1. Click **Primary business domain** and assign a primary business domain to the data product. Business domains provide a taxonomy to help consumers find the data product that they need.
2. Edit **Access level** and indicate whether the data product is *Available to everyone* or *Requires approval*.

For data products that require approval, consumers provide a business justification when they subscribe. Approvers approve or reject access requests based on this justification. You select the approvers from your community members. A task is sent to each approver's Task inbox. Approvers claim the task to approve or reject the subscription. Delivery starts after the subscription is approved.

For data products that require approval, you have a few options:

- Designate approvers for the subscription or preapprove subscribers:
  - Designate approvers: You designate the approvers from your community members. When a subscription is requested, a task is sent to each approver's Task inbox. Approvers claim the task to approve or reject the subscription based on the business justification provided by the subscriber. Delivery starts after the subscription is approved. If not approved, the consumer can subscribe again with a different justification.
  - Preapprove subscribers: You designate individual users or user groups as preapproved subscribers for the data product. All members of a user group are preapproved. The preapproved list of users doesn't generate a workflow task and doesn't require a designated approver. Users on the preapproved list automatically receive delivery of the data product.
- 3. For **Data product contents**, add the **Key features** and **Origin** of your data product. You can also add more items or preview a visualization. Enter the values for custom properties. Custom properties appear only if the Data Product Hub administrator created them for **Data product part** assets. The system-provided properties are not editable.
- 4. Choose one or more **Delivery methods** for the data product. Available delivery methods are determined by the associated data source connection. For more information on delivery methods and connectors, see [Understanding delivery methods](#) and [Delivery methods for connectors](#).
- 5. The **Additional information** tab appears if custom properties were created by the Data Product Hub administrator for **Data product** assets. Enter the values for the properties for the data product.
- 6. Provide **Recommended usage** information for your data product and describe any previous business needs that your data product addressed. This text appears on the data product tile and helps consumers locate the data product through search.

7. For **Data contract**, enter a URL or upload a PDF file. All data products must have a data contract that outlines the terms and conditions for using that data product. The URL for a data contract must be publicly accessible. For PDFs, a bucket must be configured on IBM Cloud Object Storage and files must be smaller than 50 MB. For instructions on configuring a Cloud Object Storage bucket, see [Managing storage](#).

## Task 3: Publishing your data product

---

Only published data products are available to your Data Product Hub community. Review your draft carefully because you cannot edit the contents of the data product after publishing.

1. Click **Publish**.
2. Return to the Data Product Hub home page to view your published data product.

## Learn more

---

[Managing the lifecycle of a data product](#)

**Parent topic:** [Publishing a data product](#)

---

## Creating a data product from SQL

You can use SQL statements to define a data product by creating an SQL asset in a project. You can create simple or complex queries, or create custom query templates with parameter sets that allow users to generate a custom view of the data.

- [Creating a data product from a query](#)
- [Creating a data product from a complex query](#)
- [Creating a data product from a customizable query](#)

When creating a data product using SQL, you must use a connection that supports SQL assets. The following connections support SQL assets:

- [Amazon Redshift](#)
- [Apache Cassandra](#)
- [Apache Hive](#)
- [Google BigQuery](#)
- [IBM Db2](#)
- [IBM watsonx.data Presto](#)
- [Microsoft SQL Server](#)
- [MySQL](#)
- [Oracle](#)
- [PostgreSQL](#)
- [Snowflake](#)
- [Teradata](#)

---

## Creating a data product from a query

You can customize a view of data and generate data joins by creating a data product with a query. To create a data product with a query, you must create and define an SQL query asset in a project, and then add it to a data product. By using queries, data products can be reused more frequently as producers can edit the SQL query on the same data products to generate different views.

#### Required roles

Collaborator roles: **Editor**

## Creating an SQL query

---

Create and define an SQL query as a data asset in a project. The query determines the contents of your data product.

1. From the Data Product Hub navigation menu, click **Projects > View all projects**.
2. Open or [create a project](#). Then, click **SQL: Create a dynamic view of data** to add a query asset to the project.
3. Complete the indicated fields to define the query. Enter the Preparation, SQL, and Cleanup statements.
4. Click **Create** to connect your query with the indicated data connection. You can verify that your query was validated successfully by locating it in your project's assets page.

## Adding a query to a data product

---

1. From the Data Product Hub navigation menu, click **New data product**.
2. Enter a name for the data product and click the **Add from project** tile.
3. Click **Select items** and add your query. You can use the advanced filters to locate and add your query quickly from your list of data assets.
4. Verify the connections of all assets. To verify a connection, click the connection status and complete the fields by providing credentials. Make sure that you use an appropriate set of credentials as these credentials are used to deliver the data product to consumers following a subscription.
5. Click **Create draft** to confirm your data asset selection. When your draft is successfully created, a static visualization of your data product is generated and viewable by the consumer.

## Next steps

---

After you create your data product draft, see [Completing a data product](#) to finish preparing your data product for publication.

## Learn more

---

[Creating a data product from a customizable query](#)

Parent topic: [Creating a data product](#)

---

## Creating a data product from a customizable query

You can create data products with a dynamic view of data by using customizable parameters and queries. Consumers can specify the parameters, which act as custom queries that return results used to generate the data product. By using customizable queries, data products can be reused more frequently and consumers receive the specific data that they need.

To create a data product that supports customizable queries, you must add and define the necessary parameters and then create an SQL query, which uses the parameters as variables. The consumer subscribes to the data product and indicates the parameters that they need in the subscription details.

#### Required roles

Collaborator roles: **Editor**

## Adding a data source connection

---

Connect to your remote data by adding a data source connection to your project.

1. From the Data Product Hub navigation menu, click **Projects > All projects**.
2. Open or [create a project](#). Then, from the assets page, click **New Asset > Connect to a data source**.
3. Select a data source connection and enter the necessary credentials. Make sure that you use an appropriate set of credentials as these credentials are used to deliver the data product to consumers following a subscription.
4. Click **Create** to validate your data source connection.

## Creating a parameter set

---

Add and define the necessary parameters in a parameter set. You can view the data product request details to create parameters.

1. From the assets page, click **New Asset > Define reusable sets of parameters**.
2. Define the following properties for each of the necessary parameters:
  - **Name:** Enter a name for your parameter.
  - **Type:** Indicate what data type the input should be submitted in. Parameters of type **Email** and **Encrypted** are not currently supported in customizable queries.
  - **Prompt:** Define a prompt that asks users for the parameter.
  - **Default value:** You can pre-populate a property if you are expecting similar responses.
3. Click **Save** to confirm your parameter.

## Creating an SQL query with customizable parameters

---

Define your SQL query that generates your data product by using specified parameters.

1. From the projects page, click **New asset > Create a dynamic view of data**.
2. Complete the following fields to define the data asset:
  - **Name:** Enter a name for your data asset.
  - **Description:** Describe the purpose of the query.
  - **Connection:** Select a connection that supports SQL queries. For supported connections, see [Delivery methods for connectors](#).
  - **Parameter set:** Select a parameter set to create queries for.
  - **Tags (optional):** Add tags to make your asset easier to find.
3. In the SQL query, connect each parameter with a data value. Ensure that all parameter names in the query match their corresponding names in the parameter set. Designate the parameters by using the following syntax: `'#{parameter_1_name}#'`.

Use the following code sample and graphic as an example for writing an SQL query with parameters:

```
SELECT * FROM {DATABASE.NAME}
WHERE
  {DATABASE.NAME.COLUMN1} = '#{parameter_1_name}#' AND
  {DATABASE.NAME.COLUMN2} = '#{parameter_2_name}#';
```



1. Enter sample values for each parameter to generate a data preview. All parameters defined in the query field must contain a sample value. To leave the sample value blank, remove the parameter from the query field.
2. Click **Create** to connect your parameter set with your SQL query. You can verify that your query was validated successfully by locating it in your project's assets page.

## Adding a customizable query to a data product

Create your data product by using the projects option.

1. From the Data Product Hub navigation menu, click **New data product**.
2. Enter a name for the data product and click the **Add from project** tile.
3. Click **Select items** and add your query. You can use the advanced filters to locate and add your query quickly and easily from your list of data assets.
4. Ensure that your **Connection status** is verified and create the draft. When your draft is successfully created, a static visualization of your data product is generated and viewable by the consumer.

## Next steps

After you create your data product draft, see [Completing a data product](#) to finish preparing your data product for publication.

## Learn more

- [Creating data source connections](#)

Parent topic: [Creating a data product](#)

## Creating a data product from a complex query

You can customize a view of your data by creating a data product with a complex query. To create a data product with a complex query, you must create and define an SQL query asset in a project and then add the asset to the data product. By using a complex query, you can create and then drop temporary or transient tables. Complex queries require a Snowflake connection.

### Required roles

Collaborator roles: **Editor** or **Admin**

## Creating a complex SQL query

---

With complex queries, you can create data products that include Snowflake tables or contain SQL assets that query a Snowflake database. The SQL assets can optionally include a custom parameter set. You can create temporary and transient tables for your query, join them if needed, and then drop them after the query executes.

The first step is to add a SQL asset to a project. Complex queries execute in stages and each stage requires certain SQL commands.

1. From the Data Product Hub navigation menu, click **Projects > View all projects**.
2. Add a connection asset for Snowflake. A Snowflake connection is required to create a complex query.
3. Create an optional parameter set.
4. Open or [create a project](#). Then, click **SQL: Create a dynamic view of data** to add a query asset to the project.
5. Select your Snowflake connection. Select the parameter set if you defined one.
6. Complete the indicated fields to define the complex query.
  - For the first stage, **Define tables**, enter the code in the tab to create the tables that you need for the query. Create all tables: persistent, temporary, and transient.
  - For the next stage, **Setup SQL**, re-enter the code to create the temporary and transient tables. Insert the data into the tables and join the tables if needed.
  - For the next stage, **Query**, enter the SELECT statements for your query. This stage is required.
  - For the final stage, **Cleanup SQL**, drop the transient tables and clean up as needed.
7. Click **Create** to connect your query to the Snowflake connection. You can verify that your query was validated successfully by locating it in your project's assets page.

## Adding a query to a data product

---

1. From the Data Product Hub navigation menu, click **New data product**.
2. Enter a name for the data product and click the **Add from project** tile.
3. Click **Select items** and add your query. You can use the advanced filters to locate and add your query quickly from your list of data assets.
4. Verify the connections of all assets. To verify a connection, click the connection status and complete the fields by providing credentials. Make sure that you use an appropriate set of credentials as these credentials are used to deliver the data product to consumers following a subscription.
5. Click **Create draft** to confirm your data asset selection. When your draft is successfully created, a static visualization of your data product is generated and viewable by the consumer.

## Next steps

---

After you create your data product draft, see [Completing a data product](#) to finish preparing your data product for publication.

## Learn more

---

- [Working with Temporary and Transient Tables](#)

Parent topic: [Creating a data product](#)

---

## Creating a data product directly from a source

With Data Product Hub, you can create a data product from files that are stored in a data source. You connect to the data source and then browse for and select your items.

## Selecting items for your data product

---

You are assigned ownership of all data products that you create. Only the owner of a data product can edit, publish, and retire their data products.

### Required roles to complete this task

Collaborator role: **Editor** or **Admin**

One of the methods for creating data products is to browse for items on a connected data source. You can add up to 20 items from one or more data sources. You can add CSV, TXT, and XLST files. A secure connection to a data source is required.

To browse and select items:

1. From the Data Product Hub homepage, select **New data product** and provide a name for your data product. Then, choose the **Add directly from source** tile.
2. Add or choose a data connection. For instructions on how to add a connection, see [Creating data source connections](#).
3. Select the assets to include in your data product. You can expand the drop-down to view more details about each data asset.
4. Click **Create draft** to confirm your selection.

## Next steps

---

After you create your data product draft, see [Completing a data product](#) to finish preparing your data product for publication.

## Learn more

---

- [Creating data source connections](#)
- [Creating a data product with the URL method](#)
- [Managing storage](#)

Parent topic: [Creating a data product](#)

---

## Creating a data product from a catalog

You can create a data product by adding assets from a catalog in IBM Knowledge Catalog. The assets must be added to the catalog from a connected data source. By using IBM Knowledge Catalog, you can run metadata enrichment on your data assets before you add them to a data product. Metadata enrichment helps add important information and context to data assets, such as assigning business terms, checking data quality, and profiling the data. Such information provides data consumers valuable insight in to the data product contents and quality.

### Required roles to complete this task

Collaborator role: **Editor** or **Admin**

To add assets to a data product from a catalog, you must have the following prerequisites:

- Access to one or more catalogs that contain data assets that you own.
- Data assets that were added to a catalog by using a supported data source connector. For a list of the supported connectors for Data Product Hub, see [Connectors for Data Product Hub](#).
- Valid credentials for the data source connection.

## Adding assets from a catalog to a data product

---

1. From the Data Product Hub homepage, select **New data product** and provide a name for your data product. Then, choose the **Add from catalog** tile.
2. Select the items that you want to add from one or more catalogs. You can expand the drop-down to view more details about each data asset. For an item to be included in a data product, it must be a connected data asset from a supported data source connector.
3. Verify the connections of all assets. To verify a connection, click the connection status and complete the fields by providing credentials. Make sure that you use an appropriate set of credentials as these credentials are used to deliver the data product to consumers following a subscription.

4. Click **Create draft** to confirm your data asset selection.

## Next steps

---

After you create your data product draft, see [Completing a data product](#) to finish preparing your data product for publication.

## Learn more

---

- [Creating data source connections](#)

Parent topic: [Creating a data product](#)

---

## Creating a data product from a project

You can create a data product by adding assets from a project. Projects are used for data products containing SQL queries or when you do not have access to IBM Knowledge Catalog. With an SQL query, you can customize consumers' view of a data and generate data joins. By customizing the consumer's view on data, you can hide personal and sensitive information or create filters based on the consumer's request details. This means that data products can be reused more frequently as producers can edit the SQL query on the same data products to generate different views.

### Required roles to complete this task

Collaborator role: **Editor** or **Admin**

Before you begin, you must have the following:

- Valid credentials that verify a data source connection. Ensure that this connection is supported by Data Product Hub. For details, see [Delivery methods for connectors](#).

## Adding data assets to a project

---

1. From the Data Product Hub navigation menu, click **Projects > All projects**.
2. Open or [create a project](#). Then, add new assets to populate your project. You can add assets from projects in your data community and from your IBM Cloud platform.
  - Adding data assets from a connection: Select the **Connect to a data source** tile and complete the fields with your credentials. Then, click **Import assets** and select the data assets from your connection that you want to add.
  - Creating an SQL query: Select the **Create a dynamic view of data** tile and complete the indicated fields to define the query. Do not complete the **Parameter set** field.

## Adding assets from a project to a data product

---

1. From the Data Product Hub homepage, select **New data product** and provide a name for your data product. Then, choose the **Add from project** tile.
2. Select the items from the project assets list that you want to add to your data product. You can expand the drop-down to view more details about each data asset. For an item to be included in a data product, it must be a connected data asset from a supported data source connector.
3. Verify the connections of all assets. To verify a connection, click the connection status and complete the fields by providing credentials. Make sure that you use an appropriate set of credentials as these credentials are used to deliver the data product to consumers following a subscription.
4. Click **Create draft** to confirm your data asset selection. When your draft is successfully created, a static visualization of your data product is generated and viewable by the consumer.

## Next steps

---

After you create your data product draft, see [Completing a data product](#) to finish preparing your data product for publication.

**Parent topic:** [Creating a data product](#)

---

## Creating a data product from a URL

With Data Product Hub, you can create a data product from one or more public URLs.

### Required roles

Collaborator role: **Editor**

The URL method can be used to create a data product if your assets are available from a public URL. It's a good method to use if you do not have a secure connection to the data source.

You are the owner of all the data products that you create. Only the owner of a data product can edit, publish, and retire their data products. A data product has only one owner.

---

## Entering the URL and assigning a name

The URL is the location where your assets are hosted. You can combine files and URLs in a data product. The limit for the total number of items in a data product is 20.

To enter the URL for your data product:

1. From the Data Product Hub home page, click **New data product** and select the **Add from URL** tile.
2. Enter a valid public URL where data assets and other files are hosted. The URL requires the form: **https://url.domain**. A secure, TLS encrypted URL (https) is required.
3. Enter a unique name for the URL to help identify it in the data product subscriptions.
4. Optional: You can embed a preview of your URL that allows consumers to view the data product without leaving their data community. Click **Verify preview** to check that your content can be embedded. Then, enable **Allow preview** and **Save and Confirm** your changes.
5. Click **Create draft** to confirm your URL selection. When your draft is successfully created, a static visualization of your data product is generated and viewable by the consumer.

---

## Next steps

After you create your data product draft, see [Completing a data product](#) to finish preparing your data product for publication. Note that for URLs, the delivery method is automatically set to **Open URL**, which allows consumers to open the URL to access the assets.

**Parent topic:** [Creating a data product](#)

---

## Best practices for creating a data product

The best practices to consider when creating data products include how to protect sensitive data and how to preserve data protection rules.

---

## Protecting sensitive data

You can protect sensitive data by following these practices:

- Follow the principle of least privilege when creating connections and setting delivery credentials.

- Set the access level to **Requires approval** if the data product contains sensitive data, independent of whether data protection rules are in use. For details, see [Completing your data product](#).
- Review the data visualization preview to verify the contents of each item in a data product before publishing. Masked columns are reflected in the data preview. If you are expecting masked columns as determined by data protection rules, but do not see masking applied in the preview, check the connection's credentials. The credentials on the asset's connection must be the delivery credentials and not the asset owner's credentials. Data protection rules are not enforced when the data source is accessed using the credentials of the asset owner.

## Preserving data protection rules

---

Data protection rules that were defined in IBM Knowledge Catalog are preserved in a delivered data product under the following conditions:

- The data source is either IBM Data Virtualization or IBM watsonx.data.
- Access to the data source for the Data extract and Flight service delivery methods is granted using the delivery credentials, rather than the credentials of the asset owner. Deep enforcement engines do not apply data protection rules when the data source is accessed using the credentials of the asset owner.

Parent topic: [Creating a data product](#)

---

## Managing the lifecycle of data products with the My work dashboard

Data producers manage the lifecycle of their data products on the **My work** dashboard. From **My work**, you can publish, edit, or delete drafts. You can view your published data products and retire them as needed.

You can manage the stages of a data product's lifecycle. The stages include authoring drafts, publishing for the data community, and retiring them when no longer needed.

### Required roles to complete this task

Collaborator role: **Editor** or **Admin**

- [Working with draft data products](#)
- [Working with published data products](#)
- [Working with retired data products](#)

## Working with draft data products


---

Draft data products are not available to the community on Data Product Hub. All elements of a draft data product can be edited until it is published. When you are finished editing a draft, you publish it to make it available on Data Product Hub.

From the **Drafts** page, you can view your in-progress data products and select one to edit.


To create a new draft, click **New data product**.

### Editing a draft data product

Select **Edit** from the **Actions menu**  on the tile to view and edit the details for a draft. On the draft details page, you can edit the following elements:

- Title
- Primary business domain
- Version number
- Data product contents
- Recommended usage
- Data contract

### Publishing a draft

From the tile for a draft, select **Publish** from the **Actions menu**  to publish the draft and make it available on Data Product Hub.

When a data producer publishes their data product, they describe typical use cases for the data product. See [Introduction to publishing data products](#).

### Deleting a draft

From the tile for a draft, choose **Delete** from the **Actions menu**  to remove the draft entirely from Data Product Hub.

## Working with published data products

---

A published data product remains available on Data Product Hub until you create a new version or retire it. You cannot add or remove items in a published data product, or remove or update the data contract. You can edit the following fields for a published data product:

- Title
- Description
- Owner
- Primary business domain
- Use case
- Recommended usage
- Key features
- Origin
- Names of the items in a data product


## Creating new versions of data products

---

To make extensive updates to your data product, you can create a new version. You can add or delete data assets, change the available delivery methods, manage the access level, and update the contract information. By regularly creating new versions to update your data products, you ensure that subscribers receive accurate and current data. You can also incorporate consumer feedback to enhance the data product's usability.

When you publish a new version, the current version is automatically retired. Consumers subscribe to the most recent version. Version numbers identify the iterations of a data product to help both data consumers and producers track a data product over time.

To create a new version:

1. From your **Published** tab, click the Actions menu  for a published data product and select **Create a new version**.
2. Follow the provided format and give your new version a unique version number. Then, create a new version of the data product.
3. Update fields for the data product such as primary business domain, access level, and data contract as needed. You can also add or delete data assets and delivery methods.
4. After you complete your edits, publish the new version of your data product.
5. Verify your new data product version in the **Published** tab.

## Working with retired data products


---

You can retire any available data product. Retired data products are not available for new subscriptions by consumers. They cannot be edited and are not listed on Data Product Hub. Retired data products cannot be published again.

When you retire a data product, it is removed from the Home page and no new subscriptions are allowed. However, the impact on existing subscribers depends on the delivery method, as follows:

- Existing subscribers using the Flight delivery method continue to have Flight access to the data product and receive refreshed data.
- Existing subscribers using the data extract or direct download delivery method must subscribe to a new version of the data product to receive refreshed data. They continue to have access to the data they initially subscribed to by using the download link.

To retire a data product:

1. View your published data products on the **My work>Published** tab.
2. Select **Retire** from the **Actions menu** .

## Next steps

---

- [Creating a data product](#)
- [Viewing your subscriptions](#)

Parent topic: [Publishing a data product](#)

---

## Connectors for Data Product Hub

Most methods for creating a data product require a connection to a data source. You can add connections to access a broad selection of data sources to create and deliver data products. Connections are used as either a source for accessing the items in a data product, or as a target for storing data extracts. Source connections are used to create data products and require Read or Read-Only permissions. Target connections are used as a location to save data extracts to deliver data products to consumers. When you create a target connection for your data extracts, assign Read/Write permissions to provide access to the data extracts.

When a connection is created, the credentials entered by the connection owner are automatically saved for delivering a data product. Read-Only access to the data source is recommended.

With the **Download** delivery method, the consumer receives a URL to download the data product directly from the source. The connection requires Read permissions.

With the **Live access with the Flight service** delivery method, the consumer receives a code snippet to include in their client or notebook.

With the **Download data extract** method, you need two connections: a source connection where the files are extracted and a target connection where the files are saved. Consumers access the target connection to download data extracts. Thus the credentials for a target connection require Read/Write permissions, so that the producer can save the file for the data extract and the consumer can download the extracted files.

Each type of connector supports one or more delivery methods. For more information, see [Delivery methods for connectors](#).

## Supported connectors for Data Product Hub

---

- [Amazon RDS for MySQL](#)
- [Amazon RDS for Oracle](#)
- [Amazon RDS for PostgreSQL](#)
- [Amazon Redshift](#) *Supports SQL assets.*
- [Amazon S3](#)
- [Apache Cassandra](#) *Supports SQL assets.*
- [Apache Derby](#)
- [Apache HDFS](#)
- [Apache Hive](#) *Supports source connections only and SQL assets.*
- [Apache Impala](#) *Supports source connections only.*
- [Cloudant](#)
- [Dremio](#) *Supports source connections only.*
- [Dropbox](#)
- [Elasticsearch](#)
- [Google BigQuery](#) *Supports SQL assets.*
- [Google Cloud Storage](#)
- [Google Looker](#) *Supports source connections only.*
- [HTTP](#) *Supports source connections only.*
- [IBM Cloud Data Engine](#)



- [IBM Cloud Databases for MongoDB](#)
- [IBM Cloud Databases for PostgreSQL](#)
- [IBM Cloud Object Storage](#)
- [IBM Cognos Analytics](#) *Supports source connections only.*
- [IBM Data Virtualization](#) *Supports SQL assets.*
- [IBM Data Virtualization Manager for z/OS](#)
- [IBM Db2 Big SQL](#)
- [IBM Db2 for i](#)
- [IBM Db2 for z/OS](#)
- [IBM Db2 on Cloud](#)
- [IBM Db2 Warehouse on Cloud](#)
- [IBM Db2](#) *Supports SQL assets.*
- [IBM Informix](#)
- [IBM Netezza Performance Server](#)
- [IBM Planning Analytics](#)
- [IBM watsonx.data Presto](#) *Supports SQL assets.*
- [MariaDB](#)
- [Microsoft Azure Blob Storage](#)
- [Microsoft Azure Cosmos DB](#)
- [Microsoft Azure Data Lake Storage](#)
- [Microsoft Azure File Storage](#)
- [Microsoft SQL Server](#) *Supports SQL assets.*
- [MongoDB](#) *Supports source connections only.*
- [MySQL](#) *Supports SQL assets.*
- [OData](#)
- [Oracle](#) *Supports SQL assets.*
- [PostgreSQL](#) *Supports SQL assets.*
- [Presto](#) *Supports source connections only.*
- [Salesforce.com](#) *Supports source connections only.*
- [SAP OData](#)
- [SingleStoreDB](#)
- [Snowflake](#) *Supports SQL assets for 5.0.1 and later.*
- [Teradata](#) *Supports SQL assets.*

---

## Amazon RDS for MySQL connection

To access your data in Amazon RDS for MySQL, create a connection asset for it.

Amazon RDS for MySQL is a MySQL relational database that runs on the Amazon Relational Database Service (RDS).

### Supported versions

---

MySQL database versions 5.6 through 8.0

## Create a connection to Amazon RDS for MySQL

---

To create the connection asset, you need these connection details:

- Database (optional): If you do not enter a database name, you must enter the catalog name, schema name, and the table name in the properties for SQL queries.
- Hostname or IP address
- Port number
- Username and password
- SSL certificate (if required by the database server)

## Amazon RDS for MySQL setup

---

For setup instructions, see these topics:

- [Creating an Amazon RDS DB Instance](#)
- [Connecting to a DB Instance Running the MySQL Database Engine](#)

## Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Amazon RDS for MySQL documentation](#) for the correct syntax.

## Learn more

---

[Amazon RDS for MySQL](#)

Parent topic: [Supported connections](#)

---

## Amazon RDS for Oracle connection

To access your data in Amazon RDS for Oracle, create a connection asset for it.

Amazon RDS for Oracle is an Oracle relational database that runs on the Amazon Relational Database Service (RDS).

## Supported Oracle versions and editions

---

- Oracle Database 19c (19.0.0.0)
- Oracle Database 12c Release 2 (12.2.0.1)
- Oracle Database 12c Release 1 (12.1.0.2)

## Create a connection to Amazon RDS for Oracle

---

To create the connection asset, you'll need these connection details:

- Either the Oracle Service name or the Oracle System ID (SID) for the database.
- Hostname or IP address of the database
- Port number of the database. (Default is **1521**)
- SSL certificate (if required by the database server)

Select **Server proxy** to access the Amazon RDS for Oracle data source through a server proxy. Depending on its setup, a server proxy can provide load balancing, increased security, and privacy. The server proxy settings are independent of the authentication credentials and the personal or shared credentials selection. The server proxy settings cannot be stored in a vault.

- **Proxy hostname or IP address:** The proxy URL. For example, <https://proxy.example.com>.
- **Server proxy port:** The port number to connect to the proxy server. For example, 8080 or 8443.
- The **Proxy username** and **Proxy password** fields are optional.

## Amazon RDS for Oracle setup

---

To set up the Oracle database on Amazon, see these topics:

- [Creating an Amazon RDS DB Instance](#)
- [Creating an Oracle DB instance and connecting to a database on an Oracle DB instance](#)
- [Connecting to your Oracle DB instance](#)

## Learn more

---

[Amazon RDS for Oracle](#)

---

## Amazon RDS for PostgreSQL connection

To access your data in Amazon RDS for PostgreSQL, create a connection asset for it.

Amazon RDS for PostgreSQL is a PostgreSQL relational database that runs on the Amazon Relational Database Service (RDS).

### Supported versions

---

PostgreSQL database versions 9.4, 9.5, 9.6, 10, 11 and 12

## Create a connection to Amazon RDS for PostgreSQL

---

To create the connection asset, you need these connection details:

- Database name
- Hostname or IP address
- Port number
- Username and password
- SSL certificate (if required by the database server)

Select **Server proxy** to access the Amazon RDS for PostgreSQL data source through a server proxy. Depending on its setup, a server proxy can provide load balancing, increased security, and privacy. The server proxy settings are independent of the authentication credentials and the personal or shared credentials selection. The server proxy settings cannot be stored in a vault.

- **Proxy hostname or IP address:** The proxy URL. For example, <https://proxy.example.com>.
- **Server proxy port:** The port number to connect to the proxy server. For example, 8080 or 8443.
- The **Proxy username** and **Proxy password** fields are optional.

## Amazon RDS for PostgreSQL setup

---

For setup instructions, see these topics:

- [Creating an Amazon RDS DB Instance](#)
- [Connecting to a DB Instance Running the PostgreSQL Database Engine](#)

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Amazon RDS for PostgreSQL documentation](#) for the correct syntax.

## Learn more

---

[Amazon RDS for PostgreSQL](#)

Parent topic: [Supported connections](#)

---

## Amazon Redshift connection

To access your data in Amazon Redshift, create a connection asset for it.

Amazon Redshift is a data warehouse product that forms part of the larger cloud-computing platform Amazon Web Services (AWS).

## Create a connection to Amazon Redshift

---

To create the connection asset, you need these connection details:

- Database name
- Hostname or IP address
- Port number
- Username and password
- SSL certificate (if required by the database server)

## Amazon Redshift setup

---

See [Amazon Redshift setup prerequisites](#) for setup information.

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Amazon Redshift documentation](#) for the correct syntax.

## Learn more

---

[Amazon Redshift documentation](#)

Parent topic: [Supported connections](#)

---

## Amazon S3 connection

To access your data in Amazon S3, create a connection asset for it.

Amazon S3 (Amazon Simple Storage Service) is a service that is offered by Amazon Web Services (AWS) that provides object storage through a web service interface.

## Create a connection to Amazon S3

---

To create the connection asset, you need these connection details:

- **Bucket:** Bucket name that contains the files. If your AWS credentials have permissions to list buckets and access all buckets, then you only need to supply the credentials. If your credentials don't have the privilege to list buckets and can only access a particular bucket, then you need to specify the bucket.
- **Endpoint URL:** Use for an AWS GovCloud instance. Include the region code. For example, `https://s3.<region-code>.amazonaws.com`. For the list of region codes, see [AWS service endpoints](#).
- **Region:** Amazon Web Services (AWS) region. If you specify an Endpoint URL that is not for the AWS default region (us-west-2), then you should also enter a value for Region.

Select **Server proxy** to access the Amazon S3 data source through a proxy server. Depending on its setup, a proxy server can provide load balancing, increased security, and privacy. The proxy server settings are independent of the authentication credentials and the personal or shared credentials selection.

- **Proxy host:** The proxy URL. For example, `https://proxy.example.com`.
- **Proxy port number:** The port number to connect to the proxy server. For example, 8080 or 8443.
- The **Proxy username** and **Proxy password** fields are optional.

### Credentials

The combination of **Access key** and **Secret key** is the minimum credentials.

If the Amazon S3 account owner has set up temporary credentials or a Role ARN (Amazon Resource Name), enter the values provided by the Amazon S3 account owner for the applicable authentication combination:

- **Access key, Secret key, and Session token**
- **Access key, Secret key, Role ARN, Role session name,** and optional **Duration seconds**
- **Access key, Secret key, Role ARN, Role session name, External ID,** and optional **Duration seconds**

For setup instructions for the Amazon S3 account owner, see [Setting up temporary credentials or a Role ARN for Amazon S3](#).

## Amazon S3 setup

---

See the [Amazon Simple Storage Service User Guide](#) for the setup steps.

## Restriction

---

You can only add files to a data product. You cannot add directories. Filenames must have an extension, for example, *filename.csv*.

## Supported file types

---

The Amazon S3 connection supports these file types: Avro, CSV, Delimited text, Excel, JSON, ORC, Parquet, SAS, SAV, SHP, and XML.

## Table formats

---

In addition to Flat file, the Amazon S3 connection supports these Data Lake table formats: Delta Lake and Iceberg.

## Learn more

---

[Amazon S3 documentation](#)

Parent topic: [Supported connections](#)

---

# Setting up temporary credentials or a Role ARN for Amazon S3

Instead of adding another IAM user to your Amazon S3 account, you can grant them access with temporary security credentials and a Session token. Or, you can create a Role ARN (Amazon Resource Name) and then grant permission to that role to access the account. The trusted user can then use the role.

You can assign role policies to the temporary credentials to limit the permissions. For example, you can assign read-only access or access to a particular S3 bucket.

You can set up one of the following authentication combinations:

- **Access key, Secret key, and Session token**
- **Access key, Secret key, Role ARN, Role session name,** and optional **Duration seconds**
- **Access key, Secret key, Role ARN, Role session name, External ID,** and optional **Duration seconds**

## Access key, Secret key, and Session token

---

Use the AWS Security Token Service (AWS STS) operations in the AWS API to obtain temporary security credentials. These credentials consist of an Access key, a Secret key, and a Session token that expires within a configurable amount of time. For instructions, see the AWS documentation: [Requesting temporary security credentials](#).

## Access key, Secret key, Role ARN, Role session name, and optional Duration seconds

---

If someone else has their own S3 account, you can create a temporary role for that person to access your S3 account. Create the role either with the AWS Management Console or the AWS CLI. See [Creating a role to delegate permissions to an IAM user](#).

The **Role ARN** is the Amazon Resource Name for connection's role.

The **Role session name** identifies the session to S3 administrators. For example, your IAM username.

The **Duration seconds** parameter is optional. The minimum is 15 minutes. The maximum is 36 hours, the default is 1 hour. The duration seconds timer starts every time that the connection is established.

You then provide values for the **Access key**, **Secret key**, **Role ARN**, **Role session name**, and optional **Duration seconds** to the user who will create the connection.

## Access key, Secret key, Role ARN, Role session name, External ID, and optional Duration seconds

---

If someone else has their own S3 account, you can create a temporary role for that person to access your S3 account. With this combination, the **External ID** is a unique string that you specify and that the user must enter for extra security. First, create the role either with the AWS Management Console or the AWS CLI. See [Creating a role to delegate permissions to an IAM user](#). To create the External ID, see [How to use an external ID when granting access to your AWS resources to a third party](#).

You then provide the values for the **Access key**, **Secret key**, **Role ARN**, **Role session name**, **External ID**, and optional **Duration seconds** to the user who will create the connection.

## Learn more

---

[Amazon Resource Names \(ARNs\)](#)

Parent topic: [Amazon S3 connection](#)

---

## Apache Cassandra connection

To access your data in Apache Cassandra, create a connection asset for it.

Apache Cassandra is an open source, distributed, NoSQL database.

## Supported versions

---

Apache Cassandra 2.0 or later

## Create a connection to Apache Cassandra

---

To create the connection asset, you need these connection details:

- **Hostname or IP address**
- **Port number**
- **Keyspace** (optional)
- **Username and password**
- **Read consistency** (optional): Specifies the number of replicas that must respond to a read request before the data is returned to the client application.
  - **all**: Data is returned to the application after all replicas have responded. This setting provides the highest consistency and lowest availability.
  - **local\_one**: Data is returned from the closest replica in the local data center.
  - **local\_quorum**: Data is returned after a quorum of replicas in the same data center as the coordinator node has responded. This setting voids latency of inter -data center communication.
  - **local\_serial**: Data within a data center is read without proposing a new addition or update. Uncommitted transactions within the data center are committed as part of the read.

- **one**: Data is returned from the closest replica. This setting provides the highest availability, but increases the likelihood of stale data being read.
- **quorum**: (Default). Data is returned after a quorum of replicas has responded from any data center.
- **serial**: Data is read without proposing a new addition or update. Uncommitted transactions are committed as part of the read.
- **three**: Data is returned from three of the closest replicas.
- **two**: Data is returned from two of the closest replicas.
- **Write consistency** (optional): Specifies the number of replicas for which the write request must succeed before an acknowledgment is returned to the client application.
  - **all**: A write must succeed on all replica nodes in the cluster for that partition key. This setting provides the highest consistency and lowest availability.
  - **any**: A write must succeed on at least one node. Even if all replica nodes for the given partition key are down, the write can succeed after a hinted handoff has been written. This setting provides the lowest consistency and highest availability.
  - **each\_quorum**: A write must succeed on a quorum of replica nodes across a data center.
  - **local\_one**: A write must succeed on at least one replica node in the local data center.
  - **local\_quorum**: A write must succeed on a quorum of replica nodes in the same data center as the coordinator node. This setting voids latency of inter -data center communication.
  - **local\_serial**: The driver prevents unconditional updates to achieve linearizable consistency for lightweight transactions within the data center.
  - **one**: A write must succeed on at least one replica node.
  - **quorum**: (Default). A write must succeed on a quorum of replica nodes.
  - **serial**: The driver prevents unconditional updates to achieve linearizable consistency for lightweight transactions.
  - **three**: A write must succeed on at least three replica nodes.
  - **two**: A write must succeed on at least two replica nodes.
- **SSL certificate** (if required by the database server)

## Primary keys in SQL statements

---

If you create a target table with an SQL statement and you do not specify a key column, the first column is designated as the primary key.

## Apache Cassandra setup

---

- [Installing Cassandra](#)
- [Configuring Cassandra](#)
- [CREATE KEYSPACE](#)

## Learn more

---

- [cassandra.apache.org](https://cassandra.apache.org)
- [Cassandra Documentation](#)

Parent topic: [Supported connections](#)

## Apache Derby connection

---

To access your data in Apache Derby, create a connection asset for it.

Apache Derby is a relational database management system developed by the Apache Software Foundation.

## Create a connection to Apache Derby

---

To create the connection asset, you need these connection details:

- Database name
- Hostname or IP address
- Port number
- Username and password
- SSL certificate (if required by the database server)

## Apache Derby setup

---

[Apache Derby installation](#)

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Apache Derby documentation](#) for the correct syntax.

## Learn more

---

[Apache Derby documentation](#)

Parent topic: [Supported connections](#)

---

## Apache HDFS connection

To access your data in Apache HDFS, create a connection asset for it.

Apache Hadoop Distributed File System (HDFS) is a distributed file system that is designed to run on commodity hardware. Apache HDFS was formerly Hortonworks HDFS.

## Create a connection to Apache HDFS

---

To create the connection asset, you need these connection details. The WebHDFS URL is required.

The available properties in the connection form depend on whether you select **Connect to Apache Hive** so that you can write tables to the Hive data source.

- WebHDFS URL to access HDFS.
- Hive host: Hostname or IP address of the Apache Hive server.
- Hive database: The database in Apache Hive.
- Hive port number: The port number of the Apache Hive server. The default value is **10000**.
- Hive HTTP path: The path of the endpoint such as `gateway/default/hive` when the server is configured for HTTP transport mode.
- SSL certificate (if required by the Apache Hive server).

## Apache HDFS setup

---

[Install and set up a Hadoop cluster](#)

## Supported file types

---

The Apache HDFS connection supports these file types: Avro, CSV, Delimited text, Excel, JSON, ORC, Parquet, SAS, SAV, SHP, and XML.

## Table formats

---

In addition to Flat file, the Apache HDFS connection supports these Data Lake table formats: Delta Lake and Iceberg.



## Learn more

---

[Apache HDFS Users Guide](#)

**Parent topic:** [Supported connections](#)

---

## Apache Hive connection

To access your data in Apache Hive, create a connection asset for it.

Apache Hive is a data warehouse software project that provides data query and analysis and is built on top of Apache Hadoop.

## Supported versions

---

Apache Hive 1.0.x, 1.1.x, 1.2.x, 2.0.x, 2.1.x, 3.0.x, 3.1.x.

## Create a connection to Apache Hive

---

To create the connection asset, you need the following connection details:

- Database name (optional): If you do not enter a database name, you must enter the catalog name, schema name, and the table name in the properties for SQL queries.
- Hostname or IP address
- Port number
- HTTP path (optional): The path of the endpoint such as the gateway, default, or hive if the server is configured for the HTTP transport mode.
- If required by the database server, the SSL certificate

## Apache Hive setup

---

[Apache Hive installation and configuration](#)

## Restriction

---

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the [SQL Operations](#) in the Apache Hive documentation for the correct syntax.

## Learn more

---

[Apache Hive documentation](#)

**Parent topic:** [Supported connections](#)

---

## Apache Impala connection

To access your data in Apache Impala, create a connection asset for it.

Apache Impala provides high-performance, low-latency SQL queries on data that is stored in popular Apache Hadoop file formats.

## Supported versions

---

Apache Impala 1.3+

## Create a connection to Apache Impala

---

To create the connection asset, you need these connection details:

- Database (optional): If you do not enter a database name, you must enter the catalog name, schema name, and the table name in the properties for SQL queries.
- Hostname or IP address
- Port number
- Username and password
- SSL certificate (if required by the database server)

## Apache Impala setup

---

[Apache Impala installation](#)

## Restriction

---

You can use this connection only for source data. You cannot write to data or export data with this connection.

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Impala SQL Language Reference](#) for the correct syntax.

## Learn more

---

[Apache Impala documentation](#)

Parent topic: [Supported connections](#)

---

## IBM Cloudant connection

To access your data in IBM Cloudant, create a connection asset for it.

Cloudant is a JSON document database available in IBM Cloud.

## Create a connection to Cloudant

---

To create the connection asset, you need these connection details:

- URL to the Cloudant database
- Database name
- Username and password

## Cloudant setup

---

To set up the Cloudant database on IBM Cloud, see [Getting started with IBM Cloudant](#).

When you create your Cloudant service, for **Authentication method**, select **IAM and legacy credentials**.

## Restriction

---

IBM Cloud Query (CQ) is not supported.

## Learn more

---

[IBM Cloudant docs](#)

Parent topic: [Supported connections](#)

---

## Dremio connection

To access your data in Dremio, create a connection asset for it.

Dremio is an open data lake platform. It supports all the major third-party data sources. You can connect to an instance on Dremio Cloud or Dremio Software (on-prem).

## Create a connection to Dremio

---

To create the connection asset, you need these connection details:

- Hostname or IP address: You can create a Dremio Cloud instance only in the European Union (EU) or the United States (US). Use `data.eu.dremio.cloud` for the EU and use `data.dremio.cloud` for the US. Dremio Software can be hosted anywhere.
- Port number: The default port for Dremio Cloud instances is **443** and for Dremio Software it is **32010**.
- Authentication method:
  - To connect to Dremio Cloud, you must use a Personal Access Token for authentication. Select **Port is SSL-enabled** and enter the Personal Access Token. To generate a Personal Access Token, see the instructions [Personal Access Tokens](#) for Dremio Cloud.
  - To connect to Dremio Software, you can use a username and password or you can use a Personal Access Token for authentication. To use a Personal Access Token, select **Port is SSL-enabled** and enter the Personal Access Token. To generate a Personal Access Token, see the instructions [Personal Access Tokens](#) for Dremio Software.
- SSL certificate:
  - Select **Port is SSL-enabled** to connect to Dremio Cloud.
  - Select **Port is SSL-enabled** and enter an SSL certificate if you want to connect to Dremio Software with SSL.

## Dremio setup

---

Dremio can be set up in various deployments, see [Dremio Cluster Deployment](#). To set up Dremio Cloud, see [Dremio Cloud](#).

## Restriction

---

You can use this connection only for reading data. You cannot write data or export data with this connection.

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Dremio SQL Reference](#) for the correct syntax.

## Learn more

---

- [Dremio Software documentation](#)
- [Dremio Cloud documentation](#)

Parent topic: [Supported connections](#)

---

## Dropbox connection

To access your data in Dropbox, create a connection asset for it.

Dropbox is a cloud storage service where you can host and synchronize files on your devices.

---

### Create a connection to Dropbox

To create the connection asset, you need an access token or a refresh token.

Authentication method: You can use an access token for short-term access or a refresh token for long-term access to Dropbox.

- **Access token (short-lived):** The OAuth2 access token that you obtained from the Dropbox [App Console](#) or by following the instructions at the [OAuth Guide](#) in the Dropbox documentation.
- **Refresh token (long-lived):** For setup instructions for the token, see the **Implement refresh tokens** information at [Migrating App Permissions and Access Tokens](#) in the Dropbox documentation. You can find the values for the **App key (Client ID)** and **App secret (Client secret)** in the Dropbox [App Console](#) after you select your app.

---

### Dropbox setup

[Dropbox plans](#)

---

### Supported file types

The Dropbox connection supports these file types: Avro, CSV, Delimited text, Excel, JSON, ORC, Parquet, SAS, SAV, SHP, and XML.

---

### Learn more

- [Dropbox quick start guides](#)
- [Getting started with the Dropbox API](#)

Parent topic: [Supported connections](#)

---

## Elasticsearch connection

To access your data in Elasticsearch, create a connection asset for it.

Elasticsearch is a distributed, open source search and analytics engine. Use the Elasticsearch connection to access JSON documents in Elasticsearch indexes.

---

### Supported versions

Elasticsearch version 6.0 or later

Note:

Elasticsearch version 8.15.0 is not supported. If you are on version 8.15.0, upgrade to version 8.15.2. For more information, see [Elasticsearch release notes](#).

---

### Create a connection to Elasticsearch

To create the connection asset, you need these connection details:

- URL: the URL to access Elasticsearch
- SSL certificate (if required by the database server)

For credentials, choose one of the following methods:

- Username and password  
(Optional) Anonymous access
- API key  
(Optional) Anonymous access

## Elasticsearch setup

---

[Set up Elasticsearch](#)

## Restrictions

---

- For Elasticsearch versions earlier than version 7, read is limited to 10,000 rows.
- For Data Refinery, the only supported action on the target file is to append all the rows of the Data Refinery flow output to the existing data set.

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Elasticsearch Guide for SQL](#) for the correct syntax.

## Learn more

---

- [Elasticsearch](#)
- [Elastic Docs](#)

Parent topic: [Supported connections](#)

---

## Google BigQuery connection

To access your data in Google BigQuery, create a connection asset for it.

Google BigQuery is a fully managed, serverless data warehouse that enables scalable analysis over petabytes of data.

## Create a connection to Google BigQuery

---

To create the connection asset, choose an authentication method. Choices include an authentication with or without workload identity federation.

### Without workload identity federation

- **Account key (full JSON snippet):** The contents of the Google service account key JSON file
- **Client ID, Client secret, Access token, and Refresh token**

### With workload identity federation

You use an external identity provider (IdP) for authentication. An external identity provider uses Identity and Access Management (IAM) instead of service account keys. IAM provides increased security and centralized management. You can use workload identity federation authentication with an access token or with a token URL.

You can configure a Google BigQuery connection for workload identity federation with any identity provider that complies with the OpenID Connect (OIDC) specification and that satisfies the Google Cloud requirements that are described in [Prepare your external IdP](#). The requirements include:

- The identity provider must support OpenID Connect 1.0.

- The identity provider's OIDC metadata and JWKS endpoints must be publicly accessible over the internet. Google Cloud uses these endpoints to download your identity provider's key set and uses that key set to validate tokens.
- The identity provider is configured so that your workload can obtain ID tokens that meet these criteria:
  - Tokens are signed with the RS256 or ES256 algorithm.
  - Tokens contain an aud claim.

For examples of the workload identity federation configuration steps for Amazon Web Services (AWS) and Microsoft Azure, see [Workload Identity Federation with AWS and Azure](#).

## Workload Identity Federation with access token connection details

- **Access token:** An access token from the identity provider to connect to BigQuery.
- **Security Token Service audience:** The security token service audience that contains the project ID, pool ID, and provider ID. Use this format:  
  

```
//iam.googleapis.com/projects/PROJECT_NUMBER/locations/global/workloadIdentityPools/POOL_ID/providers/PROVIDER_ID
```

 For more information, see [Authenticate a workload by using the REST API](#).
- **Service account email:** The email address of the Google service account to be impersonated. For more information, see [Create a service account for the external workload](#).
- **Service account token lifetime** (optional): The lifetime in seconds of the service account access token. The default lifetime of a service account access token is one hour. For more information, see [URL-sourced credentials](#).
- **Token format:** Text or JSON with the Token field name for the name of the field in the JSON response that contains the token.
- **Token field name:** The name of the field in the JSON response that contains the token. This field appears only when the **Token format** is JSON.
- **Token type:** AWS Signature Version 4 request, Google OAuth 2.0 access token, ID token, JSON Web Token (JWT), or SAML 2.0.

## Workload Identity Federation with token URL connection details

- **Security Token Service audience:** The security token service audience that contains the project ID, pool ID, and provider ID. Use this format:  
  

```
//iam.googleapis.com/projects/PROJECT_NUMBER/locations/global/workloadIdentityPools/POOL_ID/providers/PROVIDER_ID
```

 For more information, see [Authenticate a workload using the REST API](#).
- **Service account email:** The email address of the Google service account to be impersonated. For more information, see [Create a service account for the external workload](#).
- **Service account token lifetime** (optional): The lifetime in seconds of the service account access token. The default lifetime of a service account access token is one hour. For more information, see [URL-sourced credentials](#).
- **Token URL:** The URL to retrieve a token.
- **HTTP method:** HTTP method to use for the token URL request: GET, POST, or PUT.
- **Request body** (for POST or PUT methods): The body of the HTTP request to retrieve a token.
- **HTTP headers:** HTTP headers for the token URL request in JSON or as a JSON body. Use format:  

```
"Key1"="Value1", "Key2"="Value2".
```
- **Token format:** Text or JSON with the Token field name for the name of the field in the JSON response that contains the token.
- **Token field name:** The name of the field in the JSON response that contains the token. This field appears only when the **Token format** is JSON.

- **Token type:** AWS Signature Version 4 request, Google OAuth 2.0 access token, ID token, JSON Web Token (JWT), or SAML 2.0.

## Server proxy (optional)

Select **Server proxy** to access the Google BigQuery data source through an HTTPS proxy server. Depending on its setup, a proxy server can provide load balancing, increased security, and privacy. The proxy server settings are independent of the authentication credentials and the personal or shared credentials selection.

- **Proxy host:** The hostname or IP address of the HTTPS proxy server. For example, `proxy.example.com` or `192.0.2.0`.
- **Proxy port:** The port number to connect to the HTTPS proxy server. For example, `8080` or `8443`.
- **Proxy username** and **Proxy password**.

## Other properties

**Project ID** (optional) The ID of the Google project.

**Output JSON string format:** JSON string format for output values that are complex data types (for example, nested or repeated).

- **Pretty:** Values are formatted before sending them to output. Use this option to visually read a few rows.
- **Raw:** (Default) No formatting. Use this option for the best performance.

**Metadata discovery:** The setting determines whether comments on columns (remarks) and aliases for schema objects such as tables or views (synonyms) are retrieved when assets are added by using this connection.

## Permissions

The connection to Google BigQuery requires the following BigQuery permissions:

- `bigquery.job.create`
- `bigquery.tables.get`
- `bigquery.tables.getData`

Use one of three ways to gain these permissions:

- Use the predefined BigQuery Cloud IAM role `bigquery.admin`, which includes these permissions;
- Use a combination of two roles, one from each column in the following table; or
- Create a custom role. See [Create and manage custom roles](#).

First role	Second role
<code>bigquery.dataEditor</code>	<code>bigquery.jobUser</code>
<code>bigquery.dataOwner</code>	<code>bigquery.user</code>
<code>bigquery.dataViewer</code>	

For more information about permissions and roles in Google BigQuery, see [Predefined roles and permissions](#).

## Google BigQuery setup

[Quickstart by using the Cloud Console](#)

## Learn more

- [Google BigQuery documentation](#)

Parent topic: [Supported connections](#)

## Google Cloud Storage connection

To access your data in Google Cloud Storage, create a connection asset for it.

Google Cloud Storage is an online file storage web service for storing and accessing data on Google Cloud Platform Infrastructure.

Connections in Data Product Hub are used by the functional admin user to deliver data products. The personal credentials entered by the connection owner are automatically saved for use by both the connection owner and the functional admin user.

## Create a connection to Google Cloud Storage

---

To create the connection asset, choose an authentication method. Choices include an authentication with or without workload identity federation.

### Without workload identity federation

- **Account key (full JSON snippet):** The contents of the Google service account key JSON file
- **Client ID, Client secret, Access token, and Refresh token**

### With workload identity federation

You use an external identity provider (IdP) for authentication. An external identity provider uses Identity and Access Management (IAM) instead of service account keys. IAM provides increased security and centralized management. You can use workload identity federation authentication with an access token or with a token URL.

You can configure a Google BigQuery connection for workload identity federation with any identity provider that complies with the OpenID Connect (OIDC) specification and that satisfies the Google Cloud requirements that are described in [Prepare your external IdP](#). The requirements include:

- The identity provider must support OpenID Connect 1.0.
- The identity provider's OIDC metadata and JWKS endpoints must be publicly accessible over the internet. Google Cloud uses these endpoints to download your identity provider's key set and uses that key set to validate tokens.
- The identity provider is configured so that your workload can obtain ID tokens that meet these criteria:
  - Tokens are signed with the RS256 or ES256 algorithm.
  - Tokens contain an aud claim.

For examples of the workload identity federation configuration steps for Amazon Web Services (AWS) and Microsoft Azure, see [Workload Identity Federation with AWS and Azure](#).

### Workload Identity Federation with access token connection details

- **Access token:** An access token from the identity provider to connect to BigQuery.
- **Security Token Service audience:** The security token service audience that contains the project ID, pool ID, and provider ID. Use this format:

```
//iam.googleapis.com/projects/PROJECT_NUMBER/locations/global/workloadIdentityPools/POOL_ID/providers/PROVIDER_ID
```

For more information, see [Authenticate a workload by using the REST API](#).

- **Service account email:** The email address of the Google service account to be impersonated. For more information, see [Create a service account for the external workload](#).
- **Service account token lifetime** (optional): The lifetime in seconds of the service account access token. The default lifetime of a service account access token is one hour. For more information, see [URL-sourced credentials](#).
- **Token format:** Text or JSON with the Token field name for the name of the field in the JSON response that contains the token.
- **Token field name:** The name of the field in the JSON response that contains the token. This field appears only when the **Token format** is JSON.
- **Token type:** AWS Signature Version 4 request, Google OAuth 2.0 access token, ID token, JSON Web Token (JWT), or SAML 2.0.

### Workload Identity Federation with token URL connection details



- **Security Token Service audience:** The security token service audience that contains the project ID, pool ID, and provider ID. Use this format:

```
//iam.googleapis.com/projects/PROJECT_NUMBER/locations/global/workloadIdentityPools/POOL_ID/providers/PROVIDER_ID
```

For more information, see [Authenticate a workload using the REST API](#).

- **Service account email:** The email address of the Google service account to be impersonated. For more information, see [Create a service account for the external workload](#).
- **Service account token lifetime** (optional): The lifetime in seconds of the service account access token. The default lifetime of a service account access token is one hour. For more information, see [URL-sourced credentials](#).
- **Token URL:** The URL to retrieve a token.
- **HTTP method:** HTTP method to use for the token URL request: GET, POST, or PUT.
- **Request body** (for POST or PUT methods): The body of the HTTP request to retrieve a token.
- **HTTP headers:** HTTP headers for the token URL request in JSON or as a JSON body. Use format: "Key1"="Value1", "Key2"="Value2".
- **Token format:** Text or JSON with the Token field name for the name of the field in the JSON response that contains the token.
- **Token field name:** The name of the field in the JSON response that contains the token. This field appears only when the **Token format** is JSON.
- **Token type:** AWS Signature Version 4 request, Google OAuth 2.0 access token, ID token, JSON Web Token (JWT), or SAML 2.0.

### Server proxy (optional)

Select **Server proxy** to access the Google Cloud Storage data source through an HTTPS proxy server. Depending on its setup, a proxy server can provide load balancing, increased security, and privacy. The proxy server settings are independent of the authentication credentials and the personal or shared credentials selection. A SSL certificate can be provided for added security.

- **Proxy host:** The hostname or IP address of the HTTPS proxy server. For example, `proxy.example.com` or `192.0.2.0`.
- **Proxy port:** The port number to connect to the HTTPS proxy server. For example, `8080` or `8443`.
- **Proxy username** and **Proxy password**.

### Other properties

**Project ID** (optional) The ID of the Google project.

## Supported file types

---

The Google Cloud Storage connection supports these file types: Avro, CSV, Delimited text, Excel, JSON, ORC, Parquet, SAS, SAV, SHP, and XML.

## Table formats

---

The Google Cloud Storage connection supports these Data Lake table formats: Delta Lake and Iceberg.

## Learn more

---

- [Google Cloud Storage documentation](#)

Parent topic: [Supported connections](#)

---

## Google Looker connection

To access your data in Google Looker, create a connection asset for it.

Google Looker is a business intelligence software and big data analytics platform that helps you explore, analyze and share real-time business analytics.

---

## Create a connection to Google Looker

To create the connection asset, you need these connection details:

- Hostname or IP address
- Port number of the Google Looker server
- Client ID and Client secret

Before you configure the connection, set up API3 credentials for your Google Looker instance. For details, see [Google Looker API Authentication](#).

---

## Google Looker setup

[Set up and administer Looker](#)

---

## Restriction

You can use this connection only for source data. You cannot write to data or export data with this connection.

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the Looker documentation, [Using SQL Runner](#), for the correct syntax.

---

## Supported file types

The Google Looker connection supports these file types: CSV, Delimited text, Excel, JSON.

---

## Learn more

[Google Looker documentation](#)

**Parent topic:** [Supported connections](#)

---

## HTTP connection

To access your data from a URL, create an HTTP connection asset for it.

---

## Supported file

Use the full path in the URL to the file that you want to read. You cannot browse for files.

---

## Certificates

Enter the SSL certificate of the host to be trusted. The SSL certificate is needed only when the host certificate is not signed by a known certificate authority.

## Restriction

---

You can use this connection only for source data. You cannot write to data or export data with this connection.

## Supported file types

---

The HTTP connection supports these file types: Avro, CSV, Delimited text, Excel, JSON, ORC, Parquet, SAS, SAV, SHP, and XML.

**Parent topic:** [Supported connections](#)

---

## IBM Cloud Data Engine connection

To access your data in IBM Cloud Data Engine, create a connection asset for it.

Important:

The IBM Cloud Data Engine connector is deprecated and will be discontinued in a future release. For more information, see [Deprecation of Data Engine](#).

IBM Cloud Data Engine is a service on IBM Cloud that you use to build, manage, and consume data lakes and their table assets in IBM Cloud Object Storage (COS). IBM Cloud Data Engine provides functions to load, prepare, and query big data that is stored in various formats. It also includes a metastore with table definitions. IBM Cloud Data Engine was formerly named "IBM Cloud SQL Query."

## Prerequisites

---

### Create a connection to IBM Cloud Data Engine

---

To create the connection asset, you need these connection details:

- The Cloud Resource Name (CRN) of the IBM Cloud Data Engine instance. Go to the IBM Cloud Data Engine service instance in your resources list in your IBM Cloud dashboard and copy the value of the CRN from the deployment details.
- Target Cloud Object Storage: A default location where IBM Cloud Data Engine stores query results. You can specify any Cloud Object Storage bucket that you have access to. You can also select the default Cloud Object Storage bucket that is created when you open the IBM Cloud Data Engine web console for the first time from IBM Cloud dashboard. See the **Target location** field in the IBM Cloud Data Engine web console.
- IBM Cloud API key: An API key for a user or service ID that has access to your IBM Cloud Data Engine and Cloud Object Storage services (for both the Cloud Object Storage data that you want to query and the default target Cloud Object Storage location).

You can create a new API key for your own user:

1. In the IBM Cloud console, go to **Manage > Access (IAM)**.
2. In the left navigation, select **API keys**.
3. Select **Create an IBM Cloud API Key**.

### Credentials

IBM Cloud Data Engine uses the SSO credentials that are specified as a single API key, which authenticates a user or service ID.

The API key must have the following properties:

- Manage permission for the IBM Cloud Data Engine instance
- Read access to all Cloud Object Storage locations that you want to read from
- Write access to the default Cloud Object Storage target location
- Write access to the IBM Cloud Data Engine instance

## Restrictions

---

You can only use this connection for source data. You cannot write to data or export data with this connection.

## IBM Cloud Data Engine setup

---

To set up IBM Cloud Data Engine on IBM Cloud Object Storage, see [Getting started with IBM Cloud Data Engine](#).

## Supported encryption

---

By default, all objects that are stored in IBM Cloud Object Storage are encrypted by using randomly generated keys and an all-or-nothing-transform (AONT). For details, see [Encrypting your data](#). Additionally, you can use managed keys to encrypt the SQL query texts and error messages that are stored in the job information. See [Encrypting SQL queries with Key Protect](#).

### Running SQL statements

[Video to learn how you can get started to run a basic query](#)

## Learn more

---

- [IBM Cloud Data Engine](#)
- [Connecting to a Cloud Data Lake with IBM Cloud Pak for Data](#)

Parent topic: [Supported connections](#)

---

## IBM Cloud Databases for MongoDB connection

To access your data in IBM Cloud Databases for MongoDB, create a connection asset for it.

IBM Cloud Databases for MongoDB is a MongoDB database that is managed by IBM Cloud. It uses a JSON document store with a rich query and aggregation framework.

## Supported editions

---

- MongoDB Community Edition
- MongoDB Enterprise Edition

## Create a connection to IBM Cloud Databases for MongoDB

---

To create the connection asset, you need these connection details:

- Database (optional): If you do not enter a database name, you must enter the catalog name, schema name, and the table name in the properties for SQL queries.
- Hostname or IP address
- Port number
- Authentication database: The name of the database in which the user was created.
- Username and password
- SSL certificate (if required by the database server)

## IBM Cloud Databases for MongoDB setup

---

[Getting Started Tutorial](#)

## Restrictions

---

- You can only use this connection for source data. You cannot write to data or export data with this connection.
- MongoDB Query Language (MQL) is not supported.

**Related connection:** [MongoDB connection](#)

**Parent topic:** [Supported connections](#)

---

## IBM Cloud Databases for PostgreSQL connection

To access your data in IBM Cloud Databases for PostgreSQL, create a connection asset for it.

IBM Cloud Databases for PostgreSQL is an open source object-relational database that is highly customizable. It's a feature-rich enterprise database with JSON support.

---

## Create a connection to IBM Cloud Databases for PostgreSQL

To create the connection asset, you need these connection details:

- Database name
- Hostname or IP address of the database
- Port number
- Username and password
- SSL certificate (if required by the database server)

---

## IBM Cloud Databases for PostgreSQL setup

[IBM Cloud Databases for PostgreSQL setup](#)

---

## Restriction

For SPSS Modeler, you can use this connection only to import data. You cannot export data to this connection or to an IBM Cloud Databases for PostgreSQL connected data asset.

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the [IBM Cloud Databases for PostgreSQL documentation](#) for the correct syntax.

---

## Learn more

[IBM Cloud Databases for PostgreSQL documentation](#)

**Parent topic:** [Supported connections](#)

---

## IBM Cloud Object Storage connection

To access your data in IBM Cloud Object Storage (COS), create a connection asset for it.

IBM Cloud Object Storage on IBM Cloud provides unstructured data storage for cloud applications. Cloud Object Storage offers S3 API and application binding with regional and cross-regional resiliency.

---

## Create a connection to IBM Cloud Object Storage

To create the connection asset, you need these connection details:

- **Bucket name.** (Optional. If you do not enter the bucket name, then the credentials must have permission to list all the buckets.)
- **Login URL.** To find the **Login URL**:
  1. Go to the Cloud Object Storage **Resource list** at <https://cloud.ibm.com/resources>.
  2. Expand the **Storage** resource.
  3. Click the Cloud Object Storage service. From the menu, select **Endpoints**.
  4. Optional: Use the **Select resiliency** and **Select location** menus to filter the choices.
  5. Copy the value of the public endpoint that is in the same region as the bucket that you want to use.
- **SSL certificate:** (Optional). A self-signed certificate that was created by a tool such as OpenSSL.

## Credentials

Use one of the following combination of values for authentication:

- **Service credentials** The Service credentials must be created with the HMAC option selected.
- **Resource instance ID, API key, Access key, and Secret key** (In this combination, the **Resource instance ID** and **API key** are used for authentication. The **Access key** and **Secret key** are stored.)
- **Access key and Secret key**

To find the value for **Service credentials**:

1. Go to the Cloud Object Storage **Resource list** at <https://cloud.ibm.com/resources>.
2. Expand the **Storage** resource.
3. Click the Cloud Object Storage service, and then click the **Service credentials** tab.
4. Expand the Key name that you want to use.
5. Copy the entire JSON file. Include the opening and closing braces { } symbols.

To find the values for the **API key, Access key, Secret key, and the Resource instance ID**:

1. Go to the Cloud Object Storage **Resource list** at <https://cloud.ibm.com/resources>.
  2. Expand the **Storage** resource.
  3. Click the Cloud Object Storage service, and then click the **Service credentials** tab.
  4. Expand the Key name that you want to use. Copy the values without the quotation marks:
- **API key:** `apikey`
  - **Access key:** `access_key_id`
  - **Secret key:** `secret_access_key`
  - **Resource instance ID:** `resource_instance_id`

## IBM Cloud Object Storage setup

[Getting started with IBM Cloud Object Storage](#)

## Restrictions

The following restrictions apply:

- You must create the Cloud Object Storage credentials with the Hash-based Message Authentication Code (HMAC) option. See [Using HMAC credentials](#).
- You can only add files to a data product. You cannot add directories. Filenames must have an extension, for example, *filename.csv*.

## Supported file types

The IBM Cloud Object Storage connection supports these file types: Avro, CSV, Delimited text, Excel, JSON, ORC, Parquet, SAS, SAV, SHP, and XML.

## Table formats

---

The IBM Cloud Object Storage connection supports these Data Lake table formats: Delta Lake and Iceberg.

**Parent topic:** [Supported connections](#)

---

## IBM Cognos Analytics connection

To access your data in Cognos Analytics, create a connection asset for it.

Cognos Analytics is an AI-fueled business intelligence platform that supports the entire analytics cycle, from discovery to operationalization.

## Supported versions

---

IBM Cognos Analytics 11

## Supported content types

---

- Report (except Reports that require prompts)
- Query

## Create a connection to Cognos Analytics

---

To create the connection asset, you need the following connection details:

- Gateway URL
- SSL certificate (if required by the database server)

### Credentials

## Cognos Analytics setup

---

Instructions for setting up Cognos Analytics: [Getting started in Cognos Analytics](#).

## Restrictions

---

- You can use this connection only for source data. You cannot write to data or export data with this connection.
- Notebooks: Self-signed certificates are not supported for notebooks. The SSL certificate that is imported into the Cognos Analytics server must be signed by a trusted root authority. To confirm that the certificate is signed by a trusted root authority, enter the Cognos Analytics URL into a browser and verify that there is a padlock to the left of the URL. If the certificate is self-signed, the Cognos Analytics server administrator must replace it with a trusted TLS certificate.

### Running SQL statements

To ensure that your SQL statements run correctly, refer to [Working with Queries in SQL](#) in the Cognos Analytics documentation for the correct syntax.

## Learn more

---

[Cognos Analytics documentation](#)

**Parent topic:** [Supported connections](#)

---

# IBM Data Virtualization connection

## Create a Data Virtualization connection

---

To create the connection asset, you need these connection details:

- **Database name**
- **Hostname or IP address** of the database
- **Port number**
- **Instance ID**
- [Credentials information](#)
- **Application name** (optional): The name of the application that is currently using the connection. For information, see [Client info properties support by the IBM Data Server Driver for JDBC and SQLJ](#).
- **Client accounting information** (optional): The value of the accounting string from the client information that is specified for the connection. For information, see [Client info properties support by the IBM Data Server Driver for JDBC and SQLJ](#).
- **Client hostname** (optional): The hostname of the machine on which the application that is using the connection is running. For information, see [Client info properties support by the IBM Data Server Driver for JDBC and SQLJ](#).
- **Client user** (optional): The name of the user on whose behalf the application that is using the connection is running. For information, see [Client info properties support by the IBM Data Server Driver for JDBC and SQLJ](#).
- **SSL certificate** (if required by the database server)

### Credentials

## Restriction

---

You can use this connection only for source data. You cannot write to data with this connection.

## Learn more

---

Parent topic: [Supported connections](#)

---

# IBM Data Virtualization Manager for z/OS connection

To access your data in Data Virtualization Manager for z/OS, create a connection asset for it.

Use the Data Virtualization Manager for z/OS connection to access data in your z/OS mainframe environment.

## Supported versions

---

IBM Data Virtualization Manager for z/OS 1.1.0

## Create a connection to Data Virtualization Manager for z/OS

---

To create the connection asset, you need these connection details:

- Hostname or IP address
- Port number
- Username and password
- SSL certificate (if required by the database server)

### Learn more

[IBM Data Virtualization Manager for z/OS](#)



Parent topic: [Supported connections](#)

---

## IBM Db2 Big SQL connection

To access your data in IBM Db2 Big SQL, create a connection asset for it.

IBM Db2 Big SQL is a high performance massively parallel processing (MPP) SQL engine for Hadoop that makes querying enterprise data from across the organization an easy and secure experience.

---

### Supported versions

Db2 Big SQL for Version 4.1+

---

## Create a connection to IBM Db2 Big SQL

To create the connection asset, you need these connection details:

- Database name
- Hostname or IP address
- Port number
- Username and password
- SSL certificate (if required by the database server)

### Credentials

If you select **Shared**, enter your username and password for the server.

---

## IBM Db2 Big SQL setup

[Installing IBM Db2 Big SQL](#)

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the [IBM Db2 Big SQL documentation](#) for the correct syntax.

---

## Learn more

[Db2 Big SQL documentation](#)

Parent topic: [Supported connections](#)

---

## IBM Db2 for i connection

To access your data in IBM Db2 for i, create a connection asset for it.

Db2 for i is the relational database manager that is fully integrated on your system. Because it is integrated on the system, Db2 for i is easy to use and manage.

---

### Supported versions

IBM DB2 for i 7.2+

---

## Prerequisites

## Obtain the certificate file

A certificate file on the Db2 for i server is required to use this connection. To obtain an IBM Db2 Connect Unlimited Edition license certificate file, go to [IBM Db2 Connect: Pricing](#) and [Installing the IBM Data Server Driver for JDBC and SQLJ](#). For installation instructions, see [Activating the license certificate file for Db2 Connect Unlimited Edition](#).

## Run the bind command

Run the following commands from the Db2 client that is configured to access the Db2 for i server. You need to run the bind command only once per remote database per Db2 client version.

```
db2 connect to DBALIAS user USERID using PASSWORD
db2 bind path@ddcs400.lst blocking all sqlerror continue messages ddcs400.msg grant public
db2 connect reset
```

For information about bind commands, see [Binding applications and utilities](#).

## Run catalog commands

Run the following catalog commands from the Db2 client that is configured to access the Db2 for i server:

1. `db2 catalog tcpip node node_name remote hostname_or_address server port_no_or_service_name`

Example:

```
db2 catalog tcpip node db2i123 remote 192.0.2.0 server 446
```

2. `db2 catalog dcs database local_name as real_db_name`

Example:

```
db2 catalog dcs database db2i123 as db2i123
```

3. `db2 catalog database local_name as alias at node node_name authentication server`

Example:

```
db2 catalog database db2i123 as db2i123 at node db2i123 authentication server
```

For information about catalog commands, see [CATALOG TCPIP NODE](#) and [CATALOG DCS DATABASE](#).

## JT400 drivers

This connection has the option to use the JT400 toolkit. JT400 is a free toolkit that provides JDBC/ODBC drivers for accessing IBM i (formerly known as AS/400) systems. It provides a set of Java classes that can be used to access various resources and services on an IBM i system, such as databases, data queues, program calls, and more.

## Create a connection to Db2 for i

---

To create the connection asset, you need these connection details:

- Driver: For more information, see [Prerequisites](#).
- Hostname or IP address
- Port number
- Location: The unique name of the Db2 location you want to access
- Username and password
- SSL certificate (if required by the database server)

## Restriction

---

For SPSS Modeler, you can use this connection only to import data. You cannot export data to this connection or to a Db2 for i connection connected data asset.

## Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Db2 for i SQL reference](#) for the correct syntax.

## Learn more

---

[IBM Db2 for i documentation](#)

Parent topic: [Supported connections](#)

---

## IBM Db2 for z/OS connection

To access your data in IBM Db2 for z/OS, create a connection asset for it.

Db2 for z/OS is an enterprise data server for IBM Z. It manages core business data across an enterprise and supports key business applications.

## Supported versions

---

IBM Db2 for z/OS version 11 and later.

## Create a connection to Db2 for z/OS

---

To create the connection asset, you need these connection details:

- **Hostname or IP address**
- **Port number**
- **Collection ID:** The ID of the collections of packages to use
- **Location:** The unique name of the Db2 location you want to access
- **Username and password**
- **Application name** (optional): The name of the application that is currently using the connection. For information, see [Client info properties support by the IBM Data Server Driver for JDBC and SQLJ](#).
- **Client accounting information** (optional): The value of the accounting string from the client information that is specified for the connection. For information, see [Client info properties support by the IBM Data Server Driver for JDBC and SQLJ](#).
- **Client hostname** (optional): The hostname of the machine on which the application that is using the connection is running. For information, see [Client info properties support by the IBM Data Server Driver for JDBC and SQLJ](#).
- **Client user** (optional): The name of the user on whose behalf the application that is using the connection is running. For information, see [Client info properties support by the IBM Data Server Driver for JDBC and SQLJ](#).
- **SSL certificate** (if required by the database server)

## Restriction

---

For SPSS Modeler, you can use this connection only to import data. You cannot export data to this connection or to a Db2 for z/OS connected data asset.

## Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Db2 for z/OS and SQL concepts](#) for the correct syntax.

## Learn more

---

[IBM Db2 for z/OS documentation](#)

Parent topic: [Supported connections](#)

---

## IBM Db2 on Cloud connection

To access your data in IBM Db2 on Cloud, create a connection asset for it.

Db2 on Cloud is an SQL database that is managed by IBM Cloud and is provisioned for you in the cloud.

### Create a connection to Db2 on Cloud

---

To create the connection asset, you need the following connection details:

- Database name
- Hostname or IP address
- Port number
- Username and password

### Db2 on Cloud setup

---

[Getting started with Db2 on Cloud](#)

#### Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Structured Query Language \(SQL\)](#) topic in the Db2 on Cloud documentation for the correct syntax.

### Learn more

---

- [Db2 on Cloud documentation](#)
- [SSL connectivity](#)

Parent topic: [Supported connections](#)

---

## IBM Db2 Warehouse connection

To access your data in IBM Db2 Warehouse, create a connection asset for it.

IBM Db2 Warehouse is an analytics data warehouse that gives you a high level of control over your data and applications. You can use the IBM Db2 Warehouse connection to connect to a database in these products:

- IBM Db2 Warehouse in IBM Cloud
- IBM Db2 Warehouse on-prem

### Create a connection to Db2 Warehouse

---

To create the connection asset, you need these connection details:

- **Database name**
- **Hostname or IP address** of the database server
- **Port number**
- **API key** or **Username** and **password**
- **Application name** (optional): The name of the application that is currently using the connection. For information, see [Client info properties support by the IBM Data Server Driver for JDBC and SQLJ](#).
- **Client accounting information** (optional): The value of the accounting string from the client information that is specified for the connection. For information, see [Client info properties support by the IBM Data Server Driver for JDBC and SQLJ](#).

- **Client hostname** (optional): The hostname of the machine on which the application that is using the connection is running. For information, see [Client info properties support by the IBM Data Server Driver for JDBC and SQLJ](#).
- **Client user** (optional): The name of the user on whose behalf the application that is using the connection is running. For information, see [Client info properties support by the IBM Data Server Driver for JDBC and SQLJ](#).
- **SSL certificate** (if required by the database server)

## Authenticating with an API Key

### Db2 Warehouse in IBM Cloud

First add the user ID as an IAM user or as a service ID. For instructions, see **the Console user experience** section of the [Identity and access management \(IAM\) on IBM Cloud](#) topic.

If users want to authenticate with Db2 Warehouse with an IAM API key, the administrator of the Db2 Warehouse instance can add the IAM users by using the User management console, and then the users can each create an API key for themselves by using the IAM access management console.

## Db2 Warehouse setup

---

- IBM Db2 Warehouse on Cloud: [Getting started with Db2 Warehouse on Cloud](#)
- IBM Db2 Warehouse on-prem: [Setting up Db2 Warehouse](#)

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the product documentation in [Learn more](#) for the correct syntax.

## Known issue

---

On Data Refinery, system-level schemas aren't filtered out.

## Learn more

---

- IBM Db2 Warehouse on Cloud [product documentation](#) (IBM Cloud)
- IBM Db2 Warehouse on-prem [product documentation](#)

Parent topic: [Supported connections](#)

---

## IBM Db2 connection

To access your data in an IBM Db2 database, create a connection asset for it.

IBM Db2 is a database that contains relational data.

## Supported versions

---

IBM Db2 10.1 and later

## Create a connection to Db2

---

To create the connection asset, you need the following connection details:

- **Database**
- **Hostname or IP address**
- **Username and password**

- **Port**
- **Application name** (optional): The name of the application that is currently using the connection. For information, see [Client info properties support by the IBM Data Server Driver for JDBC and SQLJ](#).
- **Client accounting information** (optional): The value of the accounting string from the client information that is specified for the connection. For information, see [Client info properties support by the IBM Data Server Driver for JDBC and SQLJ](#).
- **Client hostname** (optional): The hostname of the machine on which the application that is using the connection is running. For information, see [Client info properties support by the IBM Data Server Driver for JDBC and SQLJ](#).
- **Client user** (optional): The name of the user on whose behalf the application that is using the connection is running. For information, see [Client info properties support by the IBM Data Server Driver for JDBC and SQLJ](#).
- **SSL certificate** (if required by your database server)

## Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Structured Query Language \(SQL\)](#) topic in the IBM Db2 product documentation for the correct syntax.

## Learn more

---

[IBM Db2 product documentation](#)

Parent topic: [Supported connections](#)

---

## IBM Informix connection

To access your data in an IBM Informix database, create a connection asset for it.

IBM Informix is a database that contains relational, object-relational, or dimensional data. You can use the Informix connection to access data from an on-prem Informix database server or from IBM Informix on Cloud.

## Supported Informix versions (on-prem)

---

- Informix 14.10 and later. This version does not support the Progress DataDirect JDBC driver, which is used by the Informix connection. The Informix connection supports Informix 14.10 features that are comparable to previous Informix versions, but not the new features. Issues related to DataDirect's JDBC driver are not supported.
- Informix 12.10 and later
- Informix 11.0 and later
- Informix 10.0 and later
- Informix 9.2 and later

## Create a connection to Informix

---

To create the connection asset, specify these connection details:

- Database server name
- Database name You don't have to specify the database. With no database specified, you can import metadata from every database that is available for that connection.
- Hostname or IP address
- Port number The default is **1526**.
- Username and password

## Informix setup

---

To set up Informix, see these topics:

- Informix on-prem: [Creating a database server after installation](#)
- Informix on Cloud: [Getting started with Informix on Cloud](#)

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Guide to SQL: Syntax](#) in the product documentation for the correct syntax.

## Learn more

---

- [Informix product documentation](#) (on-prem)
- [IBM Informix on Cloud](#)
- [IBM Informix on Cloud FAQ](#)

Parent topic: [Supported connections](#)

---

## IBM Netezza Performance Server connection

To access your data in IBM Netezza Performance Server, create a connection asset for it.

Netezza Performance Server is a platform for high-performance data warehousing and analytics.

## Supported versions

---

- IBM Netezza Performance Server 11.x
- IBM Netezza appliance software 7.0.x, 7.1.x, 7.2.x

## Create a connection to Netezza Performance Server

---

To create the connection asset, you need these connection details:

- Database name
- Hostname or IP address
- Port number
- Username and password
- SSL certificate (if required by the database server)

## Netezza Performance Server setup

---

- [Netezza Performance Server Getting started](#)
- [PureData System for Analytics Initial system setup](#)

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the product documentation:

- [Netezza Performance Server SQL command reference](#)
- [PureData System for Analytics IBM Netezza SQL Extensions toolkit](#)

## Learn more

---

- [IBM Netezza Performance Server documentation](#)
- [IBM PureData System for Analytics documentation](#)

Parent topic: [Supported connections](#)

---

# IBM Planning Analytics connection

To access your data in Planning Analytics, create a connection asset for it.

Planning Analytics (formerly known as "TM1") is an enterprise performance management database that stores data in in-memory multidimensional OLAP cubes.

## Supported versions

---

IBM Planning Analytics, version 2.0.5 or later

## Create a connection to Planning Analytics

---

To create the connection asset, you need these connection details:

- TM1 server API root URL
- Authentication type (Basic or CAM Credentials)
- Username and password
- SSL certificate (if required by the database server)

For authentication setup information, see [Authenticating and managing sessions](#).

## Planning Analytics setup

---

Enable TM1 REST APIs on the TM1 Server. See TMI REST API [Installation and configuration](#).

### Cube dimension order

#### Versions earlier than TM1 11.4

For best performance, do not combine string and numeric data in a single cube. However, if the cube does include both string and numeric data, the string elements must be in the last dimension when the cube is created. Reordering dimensions later is ignored.

#### Version TM1 11.4 or later

The default setting in Planning Analytics for cube creation is **current**. This setting might cause errors or unexpected results when you use the Planning Analytics connection. Instead, set the interaction property **use\_creation\_order** value to **true**.

### Planning Analytics view

You cannot alter the schema that is specified for a view. If you want use a different schema, you need to change the view.

## Restriction

---

For Data Refinery, you can use this connection only as a source. You cannot use this connection as a target connection or as a target connected data asset.

## Learn more

---

[Planning Analytics product documentation](#)

Parent topic: [Supported connections](#)

---

# IBM watsonx.data Presto connection



To access your data in IBM watsonx.data, create a connection asset for it. The connection asset includes information for connecting to a watsonx.data instance and to the Presto query engine that is running on that instance.

IBM watsonx.data is an open, hybrid, and governed data lakehouse that is optimized by a query engine for all data and AI workloads.

## Before you begin

---

### Differences between the watsonx.data Presto and the Presto connections

IBM watsonx.data incorporates the Presto SQL Query Engine. Both the watsonx.data Presto and Presto connections can create connection assets to interact with the Presto SQL Query Engine in IBM watsonx.data.

### watsonx.data Presto connection

The watsonx.data Presto connection supports reading from IBM watsonx.data using the Presto SQL Query Engine and supports writing tables in the Iceberg table format to Amazon S3, Apache Ozone, IBM Ceph and IBM Cloud Object Storage buckets in IBM watsonx.data. The connection is also required if you want the integration with IBM Knowledge Catalog to take advantage of the service with watsonx.data.

IBM recommends using the watsonx.data Presto connection when connecting from Cloud Pak for Data to IBM watsonx.data.

For more information about the watsonx.data Presto connection, see the rest of the topic.

### Presto

The Presto connection can create a read-only connection to any Presto engines including the implementation in IBM watsonx.data.

For more information about the Presto connection, see [Presto connection](#).

## Prerequisite

---

Set up an instance of watsonx.data.

- watsonx.data as a Service on IBM Cloud: See [Getting started with watsonx.data on IBM Cloud](#)
- watsonx.data stand-alone software: See [Installing stand-alone watsonx.data](#)

## Create a connection to watsonx.data

---

Your connection details vary between the deployment type chosen. To create the connection asset, in the **Connection details** section of the **Connect to a data source** page, select the deployment type:

- **IBM watsonx.data Developer edition**
- **IBM watsonx.data on IBM Cloud**
- **IBM watsonx.data on Red Hat OpenShift**

You can also leave the deployment type in the default value where you will see the legacy connection details.

The details you need to provide will change based on the deployment type you pick:

### IBM watsonx.data Developer edition

You can import a JSON file to fill in these fields using the **Import connection values**. To get the JSON file needed for this connection, you will need to go to your watsonx.data instance's console page and navigate to the **Connect information** field and you can copy the JSON file.

- **Hostname or IP address:** Find this information in the console under **Configurations > Connection information > Instance details**.
- **Port:** The default port number is **443**. You can find this information in the console under **Configurations > Connection information > Instance details**.

- **Instance ID:** Find this value in the watsonx.data console. Click **Instance details** from the navigation menu. You can also find this information in the console under **Configurations > Connection information > Instance details**.

## IBM watsonx.data on IBM Cloud

You can import a JSON file to fill in these fields using the **Import connection values**. To get the JSON file needed for this connection, you will need to go to your watsonx.data instance's console page and navigate to the **Connect information** field and you can copy the JSON file.

- **Hostname or IP address:** Find this information in the console under **Configurations > Connection information > Instance details**.
- **Port:** The default port number is **443**. You can find this information in the console under **Configurations > Connection information > Instance details**.
- **CRN:** Cloud resource name: Find this value in the watsonx.data console. Click **Instance details** from the navigation menu. You can also find this information in the console under **Configurations > Connection information > Instance details**.

## IBM watsonx.data on Red Hat OpenShift

You can import a JSON file to fill in these fields using the **Import connection values**. To get the JSON file needed for this connection, you will need to go to your watsonx.data instance's console page and navigate to the **Connect information** field and you can copy the JSON file.

- **Hostname or IP address:** Find this information in the console under **Configurations > Connection information > Instance details**.
- **Port:** The default port number is **443**. You can find this information in the console under **Configurations > Connection information > Instance details**.
- **Instance ID:** Find this value in the watsonx.data console. Click **Instance details** from the navigation menu. You can also find this information in the console under **Configurations > Connection information > Instance details**.

## Legacy connection details

### watsonx.data software

To create the connection asset, in the **Connection details** section of the **Connect to a data source** page, select **Connect to watsonx.data on Cloud Pak for Data** and provide these details:

- **Hostname or IP address:** Find this information in the console under **Configurations > Connection information > Instance details**.
- **Port:** The default port number is **443**. You can find this information in the console under **Configurations > Connection information > Instance details**.
- **Instance ID:** Find this value in the watsonx.data console. Click **Instance details** from the navigation menu. You can also find this information in the console under **Configurations > Connection information > Instance details**.
- **Instance name:** Find the instance name in the Cloud Pak for Data web client home page. Click **Services > Instances** from the navigation menu.

### watsonx.data as a Service

- **Hostname or IP address:** Find this information in the console under **Configurations > Connection information > Instance details**.
- **Port:** The default port number is **443**. You can find this information in the console under **Configurations > Connection information > Instance details**.
- **Instance ID:** Find this value in the watsonx.data console. Click **Instance details** from the navigation menu. You can also find this information in the console under **Configurations > Connection information > Instance details**.
- **Instance name:** Find the instance name on the **Resource list** page in IBM Cloud.
- **CRN:** Cloud resource name: Find this value in the watsonx.data console. Click **Instance details** from the navigation menu. You can also find this information in the console under **Configurations > Connection information > Instance details**.

## Credentials

Your credentials vary between the deployment type chosen:

- **IBM watsonx.data Developer edition**
- **IBM watsonx.data on IBM Cloud**
- **IBM watsonx.data on Red Hat OpenShift**

You can also leave the deployment type in the default value where you will see the legacy connection details.

### IBM watsonx.data Developer edition

- **Username and password:** The username and password that is used to log in to the watsonx.data standalone console.

### IBM watsonx.data on IBM Cloud

- **API key:** The API key of the account that has access to the watsonx.data instance on IBM Cloud.

The API key can be generated in the IBM Cloud console.

### IBM watsonx.data on Red Hat OpenShift

You must select an authentication method:

- **Username and password:** The username and password that is used to access Cloud Pak for Data where the watsonx.data instance is located.
- **Username and API key:** The username and API key that is used to access Cloud Pak for Data where the watsonx.data instance is located.

This authentication method is recommended if Cloud Pak for Data uses an Identity Management Service (IAM), for example, LDAP or SSO. The API key is located in the **Profile and settings** of the target Cloud Pak for Data cluster. For information on API keys, see .

## Legacy connection details

### watsonx.data software

The username and password or usernames and API key for the watsonx.data instance. The same credentials are also used for the engine.

You must select the authentication method:

- **Username and password:** The username and password that is used to access Cloud Pak for Data where the watsonx.data instance is located, or the username and password for watsonx.data standalone.
- **Username and API key:** The username and API key that is used to access Cloud Pak for Data where the watsonx.data instance is located, or the username and password for watsonx.data standalone. This authentication method is recommended if Cloud Pak for Data uses an Identity Management Service (IAM), for example, LDAP or SSO. The API key is located in the **Profile and settings** of the target Cloud Pak for Data cluster. For information on API keys, see .

### watsonx.data as a Service

The username and password for the watsonx.data instance. The same credentials are also used for the engine.

- **Username:** The default username is `ibmlhapikey_<cloud-account-email-address>`. For example, `ibmlhapikey_username@example.com`.
- **Password:** The password is the user's API key. To create an API key, see [IBM Cloud docs: Creating an API key in the console](#).

## Certificates

By default, **SSL is enabled** is selected. This setting is recommended for increased security. If you do not use SSL, the data might be subject to vulnerabilities such as data leakage. Although the database that is hosted in watsonx.data can also have an SSL certificate, the connection goes through the engine.

The SSL certificate must be in PEM format.

The SSL certificates information vary between the deployment type chosen:

- **IBM watsonx.data Developer edition**
- **IBM watsonx.data on IBM Cloud**
- **IBM watsonx.data on Red Hat OpenShift**

### IBM watsonx.data Developer edition

The SSL certificate is optional.

If SSL is enabled on a watsonx.data instance on Cloud Pak for Data and the certificate is a self-signed certificate, you must enter the certificate in the **SSL certificate** field.

Ask your watsonx.data administrator if SSL is set up. You can find the SSL certificate in the watsonx.data console under **Configurations > Connection information > Instance details**.

### IBM watsonx.data on IBM Cloud

The SSL certificate is optional.

### IBM watsonx.data on Red Hat OpenShift

The SSL certificate is optional.

If SSL is enabled on a watsonx.data instance on Cloud Pak for Data and the certificate is a self-signed certificate, you must enter the certificate in the **SSL certificate** field.

Ask your watsonx.data administrator if SSL is set up. You can find the SSL certificate in the watsonx.data console under **Configurations > Connection information > Instance details**.

### Engine connection details

Enter the engine connection details

### Supported engine versions

For watsonx.data on Cloud Pak for Data version 5.0.3 and later:

- Presto (Java)
- Presto (C++)

For watsonx.data on Cloud Pak for Data version 5.0.2 and before:

- Presto (Java)

For watsonx.data as a Service:

- Presto (Java)
- Presto (C++)

Provide these engine connection details. Find this information in the watsonx.data web console under **Configurations > Connection information > Engine and service connection details**.

- **Engine's hostname or IP address:** The hostname or IP address is the value of the **Internal host** field.
- **Engine ID:** This value is in the **Engine ID** field.
- **Engine's port:** The port number is the value in the **Internal host** field after the colon (:). The default port number is **8443**.

## Table actions

---

You can use the IBM watsonx.data Presto connector to ingest data into IBM watsonx.data. The data is written in the Iceberg table format.

Using table actions, you can specify what operation you want to do with the Iceberg table.

The following table actions are supported:

**Append**

The connector will create a new Iceberg table or append data to the existing table.

**Create**

The connector will create a new Iceberg table. When using this action on the existing table, you will get an error.

**Truncate**

The connector will remove logically data from the existing table and add new rows of data. This action cannot change the table's column definition.

## watsonx.data web console

---

- [Using the web console for watsonx.data on Cloud Pak for Data](#)
- [Using the web console for watsonx.data as a Service on IBM Cloud](#)
- [Using the web console for watsonx.data stand-alone](#)

## Limitation

---

- The IBM watsonx.data Presto connection does not support **TIME** and **TIMESTAMP** data types when the Presto C++ engine is used

## Learn more

---

- [watsonx.data as a Service on IBM Cloud](#)
- [Stand-alone watsonx.data](#)

**Related connections**

- [Presto connection](#)

Parent topic: [Supported connections](#)

---

## MariaDB connection

To access your data in MariaDB, create a connection asset for it.

MariaDB is an open source relational database. You can use the MariaDB connection to connect to either a MariaDB server or to a Microsoft Azure Database for MariaDB service in the cloud.

## Supported versions

---

- MariaDB server: 10.5.5
- Microsoft Azure Database for MariaDB: 10.3

## Create a connection to MariaDB

---

To create the connection asset, you need these connection details:

- Database (optional): If you do not enter a database name, you must enter the catalog name, schema name, and the table name in the properties for SQL queries.
- Hostname or IP address
- Port number

- Username and password
- SSL certificate (if required by the database server)

## MariaDB setup

---

Setup depends on whether you are connecting from a local MariaDB server or a Microsoft Azure Database for MariaDB database service in the cloud.

- MariaDB server: [MariaDB Administration](#)
- Microsoft Azure Database for MariaDB: [Quickstart: Create an Azure Database for MariaDB server by using the Azure portal](#)

## Learn more

---

- [MariaDB Foundation](#)
- [Microsoft Azure Database for MariaDB](#)

Parent topic: [Supported connections](#)

---

## Microsoft Azure Blob Storage connection

To access your data in Microsoft Azure Blob Storage, create a connection asset for it.

Azure Blob Storage is used for storing large amounts of data in the cloud.

## Create a connection to Microsoft Azure Blob Storage

---

To create the connection asset, choose an authentication method.

Data Product Hub uses the credentials of the connection owner to create and deliver data products. The personal credentials that are entered by the connection owner are automatically saved and used to create a System ID.

**Connection string:** Connection string from the storage account's **Access keys** page on the [Microsoft Azure portal](#).

Example connection string, which you can find in the **ApiKeys** section of the container:

```
DefaultEndpointsProtocol=https;AccountName=sampleaccount;AccountKey=samplekey;EndpointSuffix=core.windows.net
```

Note: Prerequisite for Entra ID authentication:

Microsoft Entra ID is a cloud-based identity and access management service. To obtain connection values for the Entra ID authentication method, sign in to the [Microsoft Azure portal](#) and go to your storage account. For information about Microsoft Entra ID, see [What is Microsoft Entra ID?](#).

### Entra ID client secret credential

- **Tenant ID:** The Microsoft Entra tenant ID. To find the Tenant ID, go to **Microsoft Entra ID > Properties**. Scroll down to the **Tenant ID** field. For more information, see [How to find your Microsoft Entra tenant ID](#).
- **Client ID:** The client ID for authorizing access to Microsoft Azure Blob Storage. To find the Client ID for your application, select **Microsoft Entra ID**. From **App registrations**, select your application. Click **Copy** to copy the Client ID of your application. For more information, see [Register a Microsoft Entra app and create a service principal](#).
- **Client secret:** The authentication key that is associated with the client ID for authorizing access to Microsoft Azure Blob Storage. To find the Client secret for your application, select **Microsoft Entra ID**. From **App registrations**, select your application. Go to **Certificates & secrets > Client secrets**. Click **Copy** to copy the existing Client secret or click **New client secret** to create a new Client secret and copy it. For more information, see [Register a Microsoft Entra app and create a service principal](#).
- **Storage account URL:** Storage account URL.

### Entra ID username password credential

- **Client ID:** The client ID for authorizing access to Microsoft Azure Blob Storage. To find the Client ID for your application, select **Microsoft Entra ID**. From **App registrations**, select your application. Click **Copy** to copy the Client ID of your application. For more information, see [Register a Microsoft Entra app and create a service principal](#).
- **Username and Password:** Username and password for the Microsoft Azure Blob Storage account. You need permission to access the blob without multi-factor authentication.
- **Storage account URL:** Storage account URL.

## Other properties

---

**Container:** The name of the container that contains the files to access.

## Supported file types

---

The Azure Blob Storage connection supports these file types: Avro, CSV, Delimited text, Excel, JSON, ORC, Parquet, SAS, SAV, SHP, and XML.

## Table formats

---

The Azure Blob Storage connection supports these Data Lake table formats: Delta Lake and Iceberg.

## Learn more

---

- [Microsoft Azure Storage account overview](#)
- [Quickstart: Upload, download, and list blobs with the Azure portal](#)
- [Manage storage account access keys](#)

**Parent topic:** [Supported connections](#)

---

## Microsoft Azure Cosmos DB connection

To access your data in Microsoft Azure Cosmos DB, create a connection asset for it.

Azure Cosmos DB is a fully managed NoSQL database service.

## Create a connection to Microsoft Azure Cosmos DB

---

To create the connection asset, you need these connection details:

- Hostname
- Port number
- Master key: The Azure Cosmos Database primary read-write key

**Note:** Prerequisite for Entra ID authentication:

Microsoft Entra ID is a cloud-based identity and access management service. To obtain connection values for the Entra ID authentication method, sign in to the [Microsoft Azure portal](#) and go to your storage account. For information about Microsoft Entra ID, see [What is Microsoft Entra ID?](#).

### Entra ID client secret credential

- **Tenant ID:** The Microsoft Entra tenant ID. To find the Tenant ID, go to **Microsoft Entra ID > Properties**. Scroll down to the **Tenant ID** field. For more information, see [How to find your Microsoft Entra tenant ID](#).
- **Client ID:** The client ID for authorizing access to Microsoft Azure Cosmos DB. To find the Client ID for your application, select **Microsoft Entra ID**. From **App registrations**, select your application. Click **Copy** to copy the Client ID of your

application. For more information, see [Register a Microsoft Entra app and create a service principal](#).

- **Client secret:** The authentication key that is associated with the client ID for authorizing access to Microsoft Azure Cosmos DB. To find the Client secret for your application, select **Microsoft Entra ID**. From **App registrations**, select your application. Go to **Certificates & secrets > Client secrets**. Click **Copy** to copy the existing Client secret or click **New client secret** to create a new Client secret and copy it. For more information, see [Register a Microsoft Entra app and create a service principal](#).
- **Storage account URL:** Storage account URL.

#### Entra ID username password credential

- **Client ID:** The client ID for authorizing access to Microsoft Azure Cosmos DB. To find the Client ID for your application, select **Microsoft Entra ID**. From **App registrations**, select your application. Click **Copy** to copy the Client ID of your application. For more information, see [Register a Microsoft Entra app and create a service principal](#).
- **Username and Password:** Username and password for the Microsoft Azure Cosmos DB account. You need permission to access without multi-factor authentication.
- **Storage account URL:** Storage account URL.

## Azure Cosmos DB setup

---

- Set up Azure Cosmos DB: [Azure portal](#)
- Secure access to data in Azure Cosmos DB: [Master keys](#)

## Restrictions

---

- Only the Core (SQL) API is supported.
- When using Entra ID credentials for authentication with Azure Cosmos DB connection, creating or deleting a collection/container is not supported. Only listing existing collections and making changes to them is supported. For more information, refer to [Azure Cosmos DB documentation](#).

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Azure Cosmos DB documentation](#) for the correct syntax.

## Learn more

---

[Azure Cosmos DB](#)

Parent topic: [Supported connections](#)

---

## Microsoft Azure Data Lake Storage connection

To access your data in Microsoft Azure Data Lake Storage, create a connection asset for it.

Azure Data Lake Storage (ADLS) is a scalable data storage and analytics service that is hosted in Azure, Microsoft's public cloud. The Microsoft Azure Data Lake Storage connection supports access to both Gen1 and Gen2 Azure Data Lake Storage repositories.

## Create a connection to Microsoft Azure Data Lake Storage

---

To create the connection asset, you need these connection details:

Note: Prerequisite for Entra ID authentication:

Microsoft Entra ID is a cloud-based identity and access management service. To obtain connection values for the Entra ID authentication method, sign in to the [Microsoft Azure portal](#) and go to your storage account. For information about Microsoft Entra ID, see [What is Microsoft Entra ID?](#).



## Entra ID client secret credential

- **Tenant ID:** The Microsoft Entra tenant ID. To find the Tenant ID, go to **Microsoft Entra ID > Properties**. Scroll down to the **Tenant ID** field. For more information, see [How to find your Microsoft Entra tenant ID](#).
- **Client ID:** The client ID for authorizing access to Microsoft Azure Data Lake Storage. To find the Client ID for your application, select **Microsoft Entra ID**. From **App registrations**, select your application. Click **Copy** to copy the Client ID of your application. For more information, see [Register a Microsoft Entra app and create a service principal](#).
- **Client secret:** The authentication key that is associated with the client ID for authorizing access to Microsoft Azure Data Lake Storage. To find the Client secret for your application, select **Microsoft Entra ID**. From **App registrations**, select your application. Go to **Certificates & secrets > Client secrets**. Click **Copy** to copy the existing Client secret or click **New client secret** to create a new Client secret and copy it. For more information, see [Register a Microsoft Entra app and create a service principal](#).
- **Storage account URL:** Storage account URL.

## Entra ID username password credential

- **Client ID:** The client ID for authorizing access to Microsoft Azure Data Lake Storage. To find the Client ID for your application, select **Microsoft Entra ID**. From **App registrations**, select your application. Click **Copy** to copy the Client ID of your application. For more information, see [Register a Microsoft Entra app and create a service principal](#).
- **Username and Password:** Username and password for the Microsoft Azure Data Lake Storage account. You need permission to access the file without multi-factor authentication.
- **Storage account URL:** Storage account URL.
- **WebHDFS URL:** The WebHDFS URL for accessing HDFS.  
To connect to a Gen 2 ADLS, use the format, `https://<account-name>.dfs.core.windows.net/<file-system>`  
Where `<account-name>` is the name you used when you created the ADLS instance.  
For `<file-system>`, use the name of the container you created. For more information, see the [Microsoft Data Lake Storage Gen2 documentation](#).
- **Tenant ID:** The Azure Active Directory tenant ID
- **Client ID:** The client ID for authorizing access to Microsoft Azure Data Lake Storage
- **Client secret:** The authentication key that is associated with the client ID for authorizing access to Microsoft Azure Data Lake Storage

Select **Server proxy** to access the Azure Data Lake Storage data source through a proxy server. Depending on its setup, a proxy server can provide load balancing, increased security, and privacy. The proxy server settings are independent of the authentication credentials and the personal or shared credentials selection.

- **Proxy host:** The proxy URL. For example, `https://proxy.example.com`.
- **Proxy port number:** The port number to connect to the proxy server. For example, **8080** or **8443**.
- The **Proxy protocol** selection for HTTP or HTTPS is optional.
- **No proxy:** A comma-separated list of hosts to bypass the proxy configured in the connection.

## Azure Data Lake Storage authentication setup

---

To set up authentication, you need a tenant ID, client (or application) ID, and client secret.

- **Gen1:**
  1. Create an Azure Active Directory (Azure AD) web application, get an application ID, authentication key, and a tenant ID.
  2. Then, you must assign the Azure AD application to the Azure Data Lake Storage account file or folder. Follow Steps 1, 2, and 3 at [Service-to-service authentication with Azure Data Lake Storage using Azure Active Directory](#).
- **Gen2:**
  1. Follow instructions in [Acquire a token from Azure AD for authorizing requests from a client application](#). These steps create a new identity. After you create the identity, set permissions to grant the application access to your ADLS. The Microsoft Azure Data Lake Storage connection will use the associated Client ID, Client secret, and Tenant ID for the application.
  2. Give the Azure App access to the storage container using Storage Explorer. For instructions, see [Use Azure Storage Explorer to manage directories and files in Azure Data Lake Storage Gen2](#).

## Supported file types

---

The Microsoft Azure Data Lake Storage connection supports these file types: Avro, CSV, Delimited text, Excel, JSON, ORC, Parquet, SAS, SAV, SHP, and XML.

## Table formats

---

In addition to Flat file, the Microsoft Azure Data Lake Storage connection supports these Data Lake table formats: Delta Lake and Iceberg.

## Learn more

---

[Azure Data Lake](#)

Parent topic: [Supported connections](#)

---

## Microsoft Azure File Storage connection

To access your data in Microsoft Azure File Storage, create a connection asset for it.

Azure Files are Microsoft's cloud file system. They are managed file shares that are accessible via the Server Message Block (SMB) protocol or the Network File System (NFS) protocol.

## Create a connection to Microsoft Azure File Storage

---

To create the connection asset, you need these connection details:

Connection string: Authentication is managed by the Azure portal access keys.

Note: Prerequisite for Entra ID authentication:

Microsoft Entra ID is a cloud-based identity and access management service. To obtain connection values for the Entra ID authentication method, sign in to the [Microsoft Azure portal](#) and go to your storage account. For information about Microsoft Entra ID, see [What is Microsoft Entra ID?](#).

### Entra ID client secret credential

- **Tenant ID:** The Microsoft Entra tenant ID. To find the Tenant ID, go to **Microsoft Entra ID > Properties**. Scroll down to the **Tenant ID** field. For more information, see [How to find your Microsoft Entra tenant ID](#).
- **Client ID:** The client ID for authorizing access to Microsoft Azure File Storage. To find the Client ID for your application, select **Microsoft Entra ID**. From **App registrations**, select your application. Click **Copy** to copy the Client ID of your application. For more information, see [Register a Microsoft Entra app and create a service principal](#).
- **Client secret:** The authentication key that is associated with the client ID for authorizing access to Microsoft Azure File Storage. To find the Client secret for your application, select **Microsoft Entra ID**. From **App registrations**, select your application. Go to **Certificates & secrets > Client secrets**. Click **Copy** to copy the existing Client secret or click **New client secret** to create a new Client secret and copy it. For more information, see [Register a Microsoft Entra app and create a service principal](#).
- **Storage account URL:** Storage account URL.

### Entra ID username password credential

- **Client ID:** The client ID for authorizing access to Microsoft Azure File Storage. To find the Client ID for your application, select **Microsoft Entra ID**. From **App registrations**, select your application. Click **Copy** to copy the Client ID of your application. For more information, see [Register a Microsoft Entra app and create a service principal](#).
- **Username and Password:** Username and password for the Microsoft Azure File Storage account. You need permission to access the file without multi-factor authentication.
- **Storage account URL:** Storage account URL.

Connections in Data Product Hub are used by the functional admin user to deliver data products. The personal credentials entered by the connection owner are automatically saved for use by both the connection owner and the functional admin user.

## Azure File Storage setup

---

Set up storage and access keys on the Microsoft Azure portal. For instructions see [Manage storage account access keys](#). Example connection string, which you can find in the **ApiKeys** section of the container:

```
DefaultEndpointsProtocol=https;AccountName=sampleaccount;AccountKey=samplekey;EndpointSuffix=core.windows.net
```

Choose the method to create and manage your Azure Files:

- [Quickstart: Create and manage Azure Files share with Windows virtual machines](#)
- [Quickstart: Create and manage Azure file shares with the Azure portal](#)
- [Quickstart: Create and manage an Azure file share with Azure PowerShell](#)
- [Quickstart: Create and manage Azure file shares using Azure CLI](#)
- [Quickstart: Create and manage Azure file shares with Azure Storage Explorer](#)

## Restriction

---

- Microsoft Azure's maximum file size is 1 TB.
- When using Entra ID credentials for authentication with Azure File Storage connection, creating or deleting a share/container is not supported. Only listing an existing shares and making changes to them is supported.

## Supported file types

---

The Azure File Storage connection supports these file types: Avro, CSV, Delimited text, Excel, JSON, ORC, Parquet, SAS, SAV, SHP, and XML.

## Table formats

---

The Azure File Storage connection supports these Data Lake table formats: Delta Lake and Iceberg.

## Known issues

---

- During the upload, the data is appended in portions to a temporary file and then converted into the file. Depending on the size of the streamed content, there might be a delay in creating the file. Wait until all the data is uploaded.

## Learn more

---

[Azure Files](#)

Parent topic: [Supported connections](#)

---

## Microsoft Azure SQL Database connection

To access your data in a Microsoft Azure SQL Database, create a connection asset for it.

Microsoft Azure SQL Database is a managed cloud database that is provided as part of Microsoft Azure.

## Create a connection to Microsoft Azure SQL Database

---

To create the connection asset, you need these connection details:

- Database name
- Hostname or IP address
- Port number
- SSL certificate (if required by the database server)

## Credentials

Choose an authentication method:

### Username and password

Username and Password to access the database in Microsoft Azure SQL Database.

Note: Prerequisite for Entra ID authentication:

Microsoft Entra ID is a cloud-based identity and access management service. To obtain connection values for the Entra ID authentication method, sign in to the [Microsoft Azure portal](#). For information about Entra ID, refer to the following Microsoft documentation:

- [What is Microsoft Entra ID?](#)
- [Use Microsoft Entra authentication](#)
- [Microsoft Entra authentication with SQL documentation](#)

### Entra ID client secret credential

- **Client ID:** The client ID for authorizing access to Microsoft Azure. To find the Client ID for your application, select **Microsoft Entra ID**. From **App registrations**, select your application. Click **Copy** to copy the Client ID of your application. For more information, see [Register a Microsoft Entra app and create a service principal](#).
- **Client secret:** The authentication key that is associated with the client ID for authorizing access to Microsoft Azure. To find the Client secret for your application, select **Microsoft Entra ID**. From **App registrations**, select your application. Go to **Certificates & secrets > Client secrets**. Click **Copy** to copy the existing Client secret or click **New client secret** to create a new Client secret and copy it. For more information, see [Register a Microsoft Entra app and create a service principal](#).

### Entra ID username password credential

Username and Password for the Microsoft Azure account.

## Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Azure SQL Database documentation](#) for the correct syntax.

# Microsoft Azure SQL Database setup

[Getting started with single databases in Azure SQL Database](#)

## Learn more

Parent topic: [Supported connections](#)

# Microsoft SQL Server connection

Create a connection asset for Microsoft SQL Server.

Microsoft SQL Server is a relational database management system.

## Supported versions

- Microsoft SQL Server 2000+
- Microsoft SQL Server 2000 Desktop Engine (MSDE 2000)

- Microsoft SQL Server 7.0

## Create a connection to Microsoft SQL Server

---

To create the connection asset, specify these connection details:

- Database name You don't have to specify the database. With no database specified, you can import metadata from every database that is available for that connection.
- Hostname or IP address
- Port number or Instance name If the server is configured for dynamic ports, use the Instance name.
- Username and password
- Domain name If the Microsoft SQL Server has been set up in a domain that uses NTLM (New Technology LAN Manager) authentication, select **Use Active Directory** and enter the name of the domain that is associated with the username and password.
- SSL certificate If required by the database server.

## Microsoft SQL Server setup

---

[Microsoft SQL Server installation](#)

## Restriction

---

Except for NTLM authentication, Windows Authentication is not supported.

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Transact-SQL Reference](#) for the correct syntax.

## Learn more

---

Parent topic: [Supported connections](#)

---

## MongoDB connection

To access your data in MongoDB, create a connection asset for it.

MongoDB is a distributed database that stores data in JSON-like documents.

## Supported editions and versions

---

### MongoDB editions

- MongoDB Community
- IBM Cloud Databases for MongoDB. See [IBM Cloud Databases for MongoDB connection](#) for this data source.
- MongoDB Atlas
- WiredTiger Storage Engine

### MongoDB versions

- MongoDB 3.6 and later, 4.x, 5.x, and 6.x
- Microsoft Azure Cosmos DB for MongoDB 3.6 and later, 4.x

## Create a connection to MongoDB

---

To create the connection asset, you need these connection details:

- Database (optional): If you do not enter a database name, you must enter the catalog name, schema name, and the table name in the properties for SQL queries.
- Hostname or IP address
- Port number
- Authentication database: The name of the database in which the user was created.
- Username and password
- SSL certificate (if required by the database server)

## MongoDB setup

---

[MongoDB installation](#)

## Restrictions

---

- You can only use this connection for source data. You cannot write to data or export data with this connection.
- MongoDB Query Language (MQL) is not supported.

## Learn more

---

- [MongoDB tutorials](#)
- [mongodb.com](#)

**Related connection:** [IBM Cloud Databases for MongoDB connection](#)

**Parent topic:** [Supported connections](#)

---

## MySQL connection

To access your data in MySQL, create a connection asset for it.

MySQL is an open-source relational database management system.

## Supported versions

---

- MySQL Enterprise Edition 5.0+
- MySQL Community Edition 4.1, 5.0, 5.1, 5.5, 5.6, 5.7

## Create a connection to MySQL

---

To create the connection asset, you need these connection details:

- Database (optional): If you do not enter a database name, you must enter the catalog name, schema name, and the table name in the properties for SQL queries.
- Hostname or IP Address
- Port number
- Encoding (optional): The character encoding for your data. If not specified, the default character set of the database server is used. If you change the value, enter a valid character encoding, for example, UTF-8.
- Username and password
- SSL certificate (if required by the database server)

Select **Server proxy** to access the MySQL data source through a server proxy. Depending on its setup, a server proxy can provide load balancing, increased security, and privacy. The server proxy settings are independent of the authentication credentials and the personal or shared credentials selection. The server proxy settings cannot be stored in a vault.

- **Proxy hostname or IP address:** The proxy URL. For example, <https://proxy.example.com>.
- **Server proxy port:** The port number to connect to the proxy server. For example, 8080 or 8443.
- The **Proxy username** and **Proxy password** fields are optional.

## Running SQL statements

To ensure that your SQL statements run correctly, refer to the [MySQL documentation](#) for the correct syntax.

## MySQL setup

---

[MySQL Installation](#)

## Learn more

---

[MySQL documentation](#)

Parent topic: [Supported connections](#)

---

## OData connection

To access your data in OData, create a connection asset for it.

The OData (Open Data) protocol is a REST-based data access protocol. The OData connection reads data from a data source that uses the OData protocol.

## Supported versions

---

The OData connection is supported on OData protocol version 2 or version 4.

## Create a connection to OData

---

To create the connection asset, you need these connection details:

- **Service root URL:** The URL to access the service root of a site that has implemented the OData protocol. Consult the client product's documentation to confirm the Service root URL.
- **Timeout seconds:** Timeout value for the HTTP calls.

Credentials type:

- API Key
- Basic
- None

Encryption:  
SSL certificate (if required by the database server)

## OData setup

---

To set up the OData service, see [How to Use Web API OData to Build an OData V4 Service without Entity Framework](#).

## Restrictions

---

- For Data Refinery, you can use this connection only as a source. You cannot use this connection as a target connection or as a target connected data asset.
- For SPSS Modeler, you cannot create new entity sets.

## Learn more

---

[www.odata.org](http://www.odata.org)

Parent topic: [Supported connections](#)

---

## Oracle connection

To access your data in Oracle, create a connection asset for it.

Oracle is a multi-model database management system.

## Supported versions

---

- Oracle Database 19c and 21c

## Create a connection to Oracle

---

To create the connection asset, you need the following connection details:

- Service name or Database (SID)
- Hostname or IP address
- Port number
- SSL certificate (if required by the database server)
- Alternate servers: A list of alternate database servers to use for failover for new or lost connections.  
Syntax: (**servername1**[:**port1**] [**;****property=value** [**;**...]] [**,****servername2**[:**port2**] [**;****property=value** [**;**...]] ...)

The server name (**servername1**, **servername2**, and so on) is required for each alternate server entry. The port number (**port1**, **port2**, and so on) and the connection properties (**property=value**) are optional for each alternate server entry. If the port is unspecified, the port number of the primary server is used.

If the port number of the primary server is not specified, the default port number **1521** is used.

The optional connection properties are the **ServiceName** and **SID**.

- Metadata discovery: The setting determines whether comments on columns (remarks) and aliases for schema objects such as tables or views (synonyms) are retrieved when assets are added by using this connection.

Select **Server proxy** to access the Oracle data source through a server proxy. Depending on its setup, a server proxy can provide load balancing, increased security, and privacy. The server proxy settings are independent of the authentication credentials and the personal or shared credentials selection. The server proxy settings cannot be stored in a vault.

- **Proxy hostname or IP address:** The proxy URL. For example, <https://proxy.example.com>.
- **Server proxy port:** The port number to connect to the proxy server. For example, 8080 or 8443.
- The **Proxy username** and **Proxy password** fields are optional.

## Oracle setup

---

[Oracle installation](#)



## Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Oracle Supported SQL Syntax and Functions](#) for the correct syntax.

## Learn more

---

[Oracle product documentation](#)

Parent topic: [Supported connections](#)

---

## PostgreSQL connection

To access your data in PostgreSQL, create a connection asset for it.

PostgreSQL is an open source and customizable object-relational database.

## Supported versions

---

- PostgreSQL 15.0 and later
- PostgreSQL 14.0 and later
- PostgreSQL 13.0 and later
- PostgreSQL 12.0 and later
- PostgreSQL 11.0 and later
- PostgreSQL 10.1 and later
- PostgreSQL 9.6 and later

## Create a connection to PostgreSQL

---

To create the connection asset, you need the following connection details:

- Database name
- Hostname or IP address
- Port number
- Username and password
- SSL certificate (if required by the database server)

Select **Server proxy** to access the PostgreSQL data source through a server proxy. Depending on its setup, a server proxy can provide load balancing, increased security, and privacy. The server proxy settings are independent of the authentication credentials and the personal or shared credentials selection. The server proxy settings cannot be stored in a vault.

- **Proxy hostname or IP address:** The proxy URL. For example, <https://proxy.example.com>.
- **Server proxy port:** The port number to connect to the proxy server. For example, 8080 or 8443.
- The **Proxy username** and **Proxy password** fields are optional.

## PostgreSQL setup

---

[PostgreSQL installation](#)

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the [SQL Syntax](#) in the PostgreSQL documentation.

## Learn more

---

[PostgreSQL documentation](#)

---

## Presto connection

To access your data in Presto, create a connection asset for it.

Presto is a fast and reliable SQL engine for Data Analytics and the Open Lakehouse.

---

## Supported versions

- Version 0.279 and earlier

---

## Create a connection to Presto

To create the connection asset, you need these connection details:

- Hostname or IP address
- Port
- Username
- Password (required if you connect to Presto with SSL enabled)
- SSL certificate (if required by the Presto server)

### Connecting to Presto within IBM watsonx.data

To connect to a Presto server within watsonx.data on IBM Cloud, use these connection details:

- Username: **ibmlhapikey**
- Password (for SSL-enabled, which is the default): An IBM Cloud API key. For more information, see [Connecting to Presto server](#).

To connect to a Presto server within watsonx.data on Cloud Pak for Data or stand-alone watsonx.data, use the username and password that you use for the watsonx.data console.

---

## Presto setup

To set up Presto, see [Presto installation](#).

---

## Restrictions

- You can use this connection only for source data. You cannot write to data or export data with this connection.

---

## Limitation

- The Presto connection does not support the Apache Cassandra Time data type.
- The Presto connection does not support **TIME** and **TIMESTAMP** data types when the Presto C++ engine is used

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the [SQL Statement Syntax](#) for the correct syntax.

---

## Learn more

[Presto documentation](#)

**Related connection:** [IBM watsonx.data Presto connection](#)

---

## Salesforce.com connection

To access your data in Salesforce.com, create a connection asset for it.

Salesforce.com is a cloud-based software company which provides customer relationship management (CRM). The Salesforce.com connection supports the standard SQL query language to select, insert, update, and delete data from Salesforce.com products and other supported products that use the Salesforce API.

---

## Other supported products that use the Salesforce API

- Salesforce AppExchange
- FinancialForce
- Service Cloud
- ServiceMax
- Veeva CRM

---

## Create a connection to Salesforce.com

To create the connection asset, you need these connection details:

- The username to access the Salesforce.com server.
- The password and security token to access the Salesforce.com server. In the **Password** field, append your security token to the end of your password. For example, **MyPasswordMyAccessToken**. For information about access tokens, see [Reset Your Security Token](#).
- The Salesforce.com server name. The default is `login.salesforce.com`.

---

## Restriction

You can only use this connection for source data. You cannot write to data or export data with this connection.

---

## Known issue

The following objects in the SFORCE schema are not supported: APPTABMEMBER, CONTENTDOCUMENTLINK, CONTENTFOLDERITEM, CONTENTFOLDERMEMBER, DATACLOUDADDRESS, DATACLOUDCOMPANY, DATACLOUDCONTACT, DATACLOUDANDBCOMPANY, DATASTATISTICS, ENTITYPARTICLE, EVENTBUSSUBSCRIBER, FIELDDEFINITION, FLEXQUEUEITEM, ICONDEFINITION, IDEACOMMENT, LISTVIEWCHARINSTANCE, LOGINEVENT, OUTGOINGEMAIL, OUTGOINGEMAILRELATION, OWNERCHANGEOPTIONINFO, PICKLISTVALUEINFO, PLATFORMACTION, RECORDACTIONHISTORY, RELATIONSHIPDOMAIN, RELATIONSHIPINFO, SEARCHLAYOUT, SITEDETAIL, USERAPPMENUITEM, USERENTITYACCESS, USERFIELDACCESS, USERRECORDACCESS, VOTE.

---

## Learn more

- [Get Started with Salesforce](#)
- [Salesforce editions with API access](#)

Parent topic: [Supported connections](#)

---

## SAP OData connection

To access your data in SAP OData, create a connection asset for it.

Use the SAP OData connection to extract data from a SAP system through its exposed OData services.

## Supported SAP OData products

---

The SAP OData connection is supported on SAP products that support the OData protocol version 2. Example products are S4/HANA (on premises or cloud), ERP, and CRM.

## Create a connection to SAP OData

---

To create the connection asset, you need these connection details:

Credentials type:

- API Key
- Basic
- None

Encryption:

SSL certificate (if required by the database server)

## SAP OData setup

---

See [Prerequisites for using the SAP ODATA Connector](#) for the SAP Gateway setup instructions.

## Restrictions

---

- For Data Refinery, you can use this connection only as a source. You cannot use this connection as a target connection or as a target connected data asset.
- For SPSS Modeler, you cannot create new entity sets.

**Parent topic:** [Supported connections](#)

---

## SingleStoreDB connection

To access your data in SingleStoreDB, create a connection asset for it.

SingleStoreDB is a fast, distributed, and highly scalable cloud-based SQL database. You can use SingleStoreDB to power real-time and data-intensive applications.

## Create a connection to SingleStoreDB

---

To create the connection asset, you need these connection details:

- Database (optional): If you do not enter a database name, you must enter the catalog name, schema name, and the table name in the properties for SQL queries.
- Hostname or IP address
- Port number
- Username and password
- SSL certificate (if required by the database server)

## SingleStoreDB setup

---

To set up SingleStoreDB, see [Getting Started with SingleStoreDB Cloud](#).

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the SingleStoreDB Docs [SQL Reference](#) for the correct syntax.

## Learn more

---

- [SingleStoreDB Cloud](#)
- [SingleStoreDB with IBM](#) for information about the IBM partnership with SingleStoreDB that provides a single source of procurement, support, and security.

Parent topic: [Supported connections](#)

---

## Snowflake connection

To access your data in Snowflake, create a connection asset for it.

Snowflake is a cloud-based data storage and analytics service.

## Create a connection to Snowflake

---

To create the connection asset, you need the following connection details:

- Account name: The full name of your account
- Database name
- Role: The default access control role to use in the Snowflake session
- Warehouse: The virtual warehouse

### Credentials

Authentication method:

- Username and password
- Key-Pair: Enter the contents of the private key and the key passphrase (if configured). These properties must be set up by the Snowflake administrator. For information, see [Key Pair Authentication & Key Pair Rotation](#) in the Snowflake documentation.
- Okta URL endpoint: If your company uses native Okta SSO authentication, enter the Okta URL endpoint for your Okta account. Example: `https://<okta_account_name>.okta.com`. Leave this field blank if you want to use the default authentication of Snowflake. For information about federated authentication provided by Okta, see [Native SSO](#).

## Snowflake setup

---

[General Configuration](#)

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Snowflake SQL Command Reference](#) for the correct syntax.

## Learn more

---

Parent topic: [Supported connections](#)

---

## Teradata connection

To access your data in Teradata, create a connection asset for it.

Teradata provides database and analytics-related services and products.

## Supported versions

---

Teradata databases 15.10, 16.10, 17.00, 17.10, and 17.20

## Create a connection to Teradata

---

To create the connection asset, you need these connection details:

- **Authentication method:** Select the security mechanism to use to authenticate the user:
  - **TD2 (Teradata Method 2):** Use the Teradata security mechanism.
  - **LDAP:** Use an LDAP security mechanism for external authentication.
- **Username** and **password**
- **SSL certificate** (if required by the database server)

### Running SQL statements

To ensure that your SQL statements run correctly, refer to the [Teradata SQL documentation](#) for the correct syntax.

## Learn more

---

- [Teradata documentation](#)
- [Teradata Community](#)

Parent topic: [Supported connections](#)

*Teradata JDBC Driver 17.00.00.03 Copyright (C) 2024 by Teradata. All rights reserved. IBM provides embedded usage of the Teradata JDBC Driver under license from Teradata solely for use as part of the IBM Watson service offering.*

---

## Working with delivery methods

Data Product Hub provides several options for delivering data products to the consumer. Each item in a data product can be delivered by using different methods. The connector type determines the available delivery methods.

## Types of delivery methods for Data Product Hub

---

You can configure your data product to be delivered by the following delivery methods, depending upon the connector type:

- Downloading directly from a URL
- Downloading a data extract from a target connected data source
- Opening a URL
- Accessing a data product with Flight service

Types of delivery methods

Name	Description	Notes
<a href="#">Download</a>	Consumers receive a URL to download a data product from a connection.	Requires a data source connection that supports download.
<a href="#">Data extract</a>	Consumers can download a data product as a file extract from a target connection.	Requires two connections: A source connection and a target connection. Also requires a default project.
<a href="#">Open URL</a>	Consumers receive a URL to directly access a data product.	Does not require a data source connection.

Name	Description	Notes
<a href="#">Live access with Flight service</a>	Consumers receive a code snippet to connect with a data source and receive a data product.	The provided code snippet can be added to notebooks and other applications.

## Download

The Download delivery method generates a URL that allows consumers to download directly from a connected data source.

## Open URL

The Open URL delivery method provides a URL that points to a website containing the items in a data product.

## Data extract

You can extract data from a database and deliver the extract as a file to consumers. The data extract delivery method requires two connections, one for connecting to the database you are extracting from (the source) and another connection for delivering the file to the consumer (the target). To view which connectors support source and target connections for data extracts, see [Delivery methods for connectors](#).

The data extract delivery method requires a default project. The default project is created automatically when you create a bucket in **Configurations and settings** > **Storage**.

The target connection is a file storage location that is used exclusively for storing data extracts. This connection delivers the extract to consumers through a download URL. Since data extracts can be large, a file storage location with adequate capacity is recommended. The target connections require read/write credentials.

Create two connections for the data extract delivery method:

- Source connection: The source connection is the location where you extract data.
- Target connection: The target connection is the location where the extracted file is stored and where consumers download a data extract from a URL.

The data extract delivery method supports these file types: Avro, CSV, Delimited text, Excel, JSON, ORC, Parquet, SAV, and XML.

## Live access with Flight service

The Flight service provides real-time, read/write access to many data sources through a common open source API. The Flight service provides a single interface for accessing many different data sources. With the Flight service, Data Product Hub no longer needs a connection asset to the data source with personal credentials. The Flight service retrieves the credentials transparently. To determine if your data source supports the Flight service, see [Delivery methods for connectors](#).

When you subscribe to a data product and select delivery by the Flight service, a **Flight URL** is provided. You download a code snippet in your preferred language (Python or R). The code snippet contains an asset ID and a catalog ID to connect to a data source to receive delivery of the items in a data product. The code snippet can be inserted into a notebook and other applications.

You can also access the data programmatically by using an Arrow client. Arrow libraries are available for C, C++, C#, Go, Java, JavaScript, Julia, MATLAB, Python, R and Ruby. See [Apache Arrow](#) for instructions on installing the libraries for each language.

See [Flight client example](#) for an example of how to access your data product using Python.

For more information about Arrow Flight RPC, see [Arrow Flight RPC](#).

## Learn more

[Delivery methods for connectors](#)

Parent topic: [Publishing a data product](#)

# Delivery methods for connectors

You can review the delivery methods available for each connector. Each connector in Data Product Hub supports specific delivery methods which are used to deliver data products.

Review the following table to determine the delivery methods that are supported by each connector.

Delivery methods for connectors

Source connector type	Asset type	Open URL	Direct download	Live access with Flight service	Data extract sources	Data extract targets
--	URL	X				
Amazon RDS for MySQL	Data asset / Query			X	X	
Amazon RDS for Oracle	Data asset / Query				X	
Amazon RDS for PostgreSQL	Data asset / Query				X	
Amazon Redshift	Data asset / Query			X	X	
Amazon S3	Data asset		X	X		X
Apache Cassandra	Data asset			X	X	
Apache Derby	Data asset / Query			X	X	
Apache HDFS	Data asset			X	X	
Apache Hive	Data asset / Query			X	X	
Apache Impala	Data asset			X		
Cloudant	Data asset			X	X	
Data Virtualization Manager for z/OS	Data asset			X	X	
Dremio	Data asset / Query				X	
Dropbox	Data asset				X	
Elasticsearch	Data asset				X	
Exasol	Data asset / Query			X		
Google BigQuery	Data asset / Query				X	
Google Cloud Storage	Data asset		X	X		X
HDFS via Execution Engine for Hadoop	Data asset			X		
Hive via Execution Engine for Hadoop	Data asset			X		
HTTP	Data asset			X		
IBM Cloud Data Engine	Data asset				X	
IBM Cloud Databases for MongoDB	Data asset / Query			X	X	



Source connector type	Asset type	Open URL	Direct download	Live access with Flight service	Data extract sources	Data extract targets
IBM Cloud Databases for PostgreSQL	Data asset / Query			X		
IBM Cloud Object Storage	Data asset		X	X		X
IBM Cognos Analytics	Data asset			X	X	
IBM Data Virtualization	Data asset				X	
IBM Db2 Big SQL	Data asset / Query			X	X	
IBM Db2 for i	Data asset / Query			X		
IBM Db2 for z/OS	Data asset / Query			X		
IBM Db2 on Cloud	Data asset / Query			X	X	
IBM Db2 Warehouse	Data asset			X	X	
IBM Db2	Data asset / Query			X	X	
IBM Informix	Data asset / Query			X		
IBM Planning Analytics	Data asset			X	X	
IBM watsonx.data Presto	Data asset			X	X	
Looker	Data asset			X		
MariaDB	Data asset / Query				X	
Microsoft Azure Blob Storage	Data asset		X	X		X
Microsoft Azure Cosmos DB	Data asset			X	X	
Microsoft Azure File Storage	Data asset		X	X		X
Microsoft Azure Data Lake Storage	Data asset			X		
Microsoft SQL Server	Data asset / Query			X	X	
MinIO	Data asset			X		
MongoDB	Data asset			X		
MySQL	Data asset / Query			X		
Netezza Performance Server	Data asset / Query			X		
OData	Data asset			X	X	
Oracle	Data asset / Query			X	X	

Source connector type	Asset type	Open URL	Direct download	Live access with Flight service	Data extract sources	Data extract targets
Presto	Data asset / Query			X	X	
PostgreSQL	Data asset / Query			X	X	
Salesforce	Data asset / Query			X		
SAP HANA	Data asset			X		
SAP OData	Data asset			X	X	
SingleStoreDB	Data asset / Query			X	X	
Snowflake	Data asset / Query			X	X	
Teradata	Data asset / Query			X		

Parent topic: [Working with delivery methods](#)