



Red Hat OpenShift AI Cloud Service 1

Working on data science projects

Organize your work in projects and workbenches.

Red Hat OpenShift AI Cloud Service 1 Working on data science projects

Organize your work in projects and workbenches.

Legal Notice

Copyright © 2024 Red Hat, Inc.

The text of and illustrations in this document are licensed by Red Hat under a Creative Commons Attribution–Share Alike 3.0 Unported license ("CC-BY-SA"). An explanation of CC-BY-SA is available at

<http://creativecommons.org/licenses/by-sa/3.0/>

. In accordance with CC-BY-SA, if you distribute this document or an adaptation of it, you must provide the URL for the original version.

Red Hat, as the licensor of this document, waives the right to enforce, and agrees not to assert, Section 4d of CC-BY-SA to the fullest extent permitted by applicable law.

Red Hat, Red Hat Enterprise Linux, the Shadowman logo, the Red Hat logo, JBoss, OpenShift, Fedora, the Infinity logo, and RHCE are trademarks of Red Hat, Inc., registered in the United States and other countries.

Linux[®] is the registered trademark of Linus Torvalds in the United States and other countries.

Java[®] is a registered trademark of Oracle and/or its affiliates.

XFS[®] is a trademark of Silicon Graphics International Corp. or its subsidiaries in the United States and/or other countries.

MySQL[®] is a registered trademark of MySQL AB in the United States, the European Union and other countries.

Node.js[®] is an official trademark of Joyent. Red Hat is not formally related to or endorsed by the official Joyent Node.js open source or commercial project.

The OpenStack[®] Word Mark and OpenStack logo are either registered trademarks/service marks or trademarks/service marks of the OpenStack Foundation, in the United States and other countries and are used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation, or the OpenStack community.

All other trademarks are the property of their respective owners.

Abstract

Organize your work in projects and workbenches.

Table of Contents

PREFACE	3
CHAPTER 1. USING DATA SCIENCE PROJECTS	4
1.1. CREATING A DATA SCIENCE PROJECT	4
1.2. UPDATING A DATA SCIENCE PROJECT	4
1.3. DELETING A DATA SCIENCE PROJECT	5
CHAPTER 2. USING PROJECT WORKBENCHES	7
2.1. CREATING A WORKBENCH AND SELECTING AN IDE	7
2.1.1. About workbench images	7
2.1.2. Creating a workbench	10
2.2. STARTING A WORKBENCH	12
2.3. UPDATING A PROJECT WORKBENCH	13
2.4. DELETING A WORKBENCH FROM A DATA SCIENCE PROJECT	14
CHAPTER 3. USING DATA CONNECTIONS	15
3.1. ADDING A DATA CONNECTION TO YOUR DATA SCIENCE PROJECT	15
3.2. DELETING A DATA CONNECTION	16
3.3. UPDATING A CONNECTED DATA SOURCE	16
CHAPTER 4. CONFIGURING CLUSTER STORAGE	18
4.1. ADDING CLUSTER STORAGE TO YOUR DATA SCIENCE PROJECT	18
4.2. UPDATING CLUSTER STORAGE	18
4.3. DELETING CLUSTER STORAGE FROM A DATA SCIENCE PROJECT	19
CHAPTER 5. MANAGING ACCESS TO DATA SCIENCE PROJECTS	21
5.1. CONFIGURING ACCESS TO A DATA SCIENCE PROJECT	21
5.2. SHARING ACCESS TO A DATA SCIENCE PROJECT	21
5.3. UPDATING ACCESS TO A DATA SCIENCE PROJECT	22
5.4. REMOVING ACCESS TO A DATA SCIENCE PROJECT	24

PREFACE

As a data scientist, you can organize your data science work into a single project. A data science project in OpenShift AI can consist of the following components:

Workbenches

Creating a workbench allows you to add a Jupyter notebook to your project.

Cluster storage

For data science projects that require data retention, you can add cluster storage to the project.

Data connections

Adding a data connection to your project allows you to connect data inputs to your workbenches.

Pipelines

Standardize and automate machine learning workflows to enable you to further enhance and deploy your data science models.

Models and model servers

Deploy a trained data science model to serve intelligent applications. Your model is deployed with an endpoint that allows applications to send requests to the model.

CHAPTER 1. USING DATA SCIENCE PROJECTS

1.1. CREATING A DATA SCIENCE PROJECT

To implement a data science workflow, you must create a project. In OpenShift, a project is a Kubernetes namespace with additional annotations, and is the main way that you can manage user access to resources. A project organizes your data science work in one place and also allows you to collaborate with other developers and data scientists in your organization.

Within a project, you can add the following functionality:

- Data connections so that you can access data without having to hardcode information like endpoints or credentials.
- Workbenches for working with and processing data, and for developing models.
- Deployed models so that you can test them and then integrate them into intelligent applications. Deploying a model makes it available as a service that you can access by using an API.
- Pipelines for automating your ML workflow.

Prerequisites

- You have logged in to Red Hat OpenShift AI.
- If you are using specialized OpenShift AI groups, you are part of the user group or admin group (for example, **rhoai-users** or **rhoai-admins**) in OpenShift.

Procedure

1. From the OpenShift AI dashboard, select **Data Science Projects**.
2. Click **Create data science project**
3. In the **Create a data science project** dialog, enter a display **Name** for your project.
4. Optional: Edit the **Resource name** for your data science project. The resource name must consist of lowercase alphanumeric characters, -, and must start and end with an alphanumeric character.
Note: After you create a project, you can change the project display name but you cannot change the resource name.
5. Enter a **description** for your data science project.
6. Click **Create**.

Verification

- A project details page opens. From this page, you can add data connections, create workbenches, configure pipelines, and deploy models.


1.2. UPDATING A DATA SCIENCE PROJECT

You can update your data science project's details by changing your project's name and description text.

Prerequisites

- You have logged in to Red Hat OpenShift AI.
- If you are using specialized OpenShift AI groups, you are part of the user group or admin group (for example, **rhoai-users** or **rhoai-admins**) in OpenShift.
- You have created a data science project.

Procedure

1. From the OpenShift AI dashboard, click **Data Science Projects**.
The **Data Science Projects** page opens.
2. Click the action menu () beside the project whose details you want to update and click **Edit project**.
The **Edit data science project** dialog opens.
3. Optional: Update the **name** for your data science project.
4. Optional: Update the **description** for your data science project.
5. Click **Update**.

Verification

- The data science project that you updated is displayed on the **Data Science Projects** page.


1.3. DELETING A DATA SCIENCE PROJECT

You can delete data science projects so that they do not appear on the OpenShift AI **Data Science Projects** page when you no longer want to use them.

Prerequisites

- You have logged in to Red Hat OpenShift AI.
- If you are using specialized OpenShift AI groups, you are part of the user group or admin group (for example, **{oai-user-group}**) in OpenShift.
- You have created a data science project.

Procedure

1. From the OpenShift AI dashboard, click **Data Science Projects**.
The **Data Science Projects** page opens.
2. Click the action menu () beside the project that you want to delete and then click **Delete project**.
The **Delete project** dialog opens.
3. Enter the project name in the text field to confirm that you intend to delete it.

4. Click **Delete project**.

Verification

- The data science project that you deleted is no longer displayed on the **Data Science Projects** page.
- Deleting a data science project deletes any associated workbenches, data science pipelines, cluster storage, and data connections. This data is permanently deleted and is not recoverable.

CHAPTER 2. USING PROJECT WORKBENCHES

2.1. CREATING A WORKBENCH AND SELECTING AN IDE

A workbench is an isolated area where you can examine and work with ML models. You can also work with data and run programs, for example to prepare and clean data. While a workbench is not required if, for example, you only want to service an existing model, one is needed for most data science workflow tasks, such as writing code to process data, or training a model.

When you create a workbench, you specify an image (an IDE, packages, and other dependencies). Supported IDEs include JupyterLab, code-server (Technology Preview), and RStudio (Technology Preview).

The IDEs are based on a server-client architecture. Each IDE provides a server that runs in a container on the OpenShift cluster, while the user interface (the client) is displayed in your web browser. For example, the Jupyter notebook server runs in a container on the Red Hat OpenShift cluster. The client is the JupyterLab interface that opens in your web browser on your local computer. All of the commands that you enter in JupyterLab are executed by the notebook server. Similarly, other IDEs like code-server or RStudio Server provide a server that runs in a container on the OpenShift cluster, while the user interface is displayed in your web browser. This architecture allows you to interact through your local computer in a browser environment, while all processing occurs on the cluster. The cluster provides the benefits of larger available resources and security because the data being processed never leaves the cluster.

In a workbench, you also configure data connections (to access external data for training models and to save models so that you can deploy them) and cluster storage (for persisting data). Workbenches within the same project can share models and data through object storage with the data science pipelines and model servers.

For data science projects that require data retention, you can add container storage to the workbench you are creating.

Within a project, you can create multiple workbenches. When to create a new workbench depends on considerations, such as the following:

- The workbench configuration (for example, CPU, RAM, or IDE). If you want to avoid editing the configuration of an existing workbench's configuration to accommodate a new task, you can create a new workbench instead.
- Separation of tasks or activities. For example, you might want to use one workbench for your Large Language Models (LLM) experimentation activities, another workbench dedicated to a demo, and another workbench for testing.

2.1.1. About workbench images

A workbench image (sometimes referred to as a notebook image) is optimized with the tools and libraries that you need for model development. You can use the provided workbench images or an OpenShift AI admin user can create custom workbench images adapted to your needs.

To provide a consistent, stable platform for your model development, many provided workbench images contain the same version of Python. Most workbench images available on OpenShift AI are pre-built and ready for you to use immediately after OpenShift AI is installed or upgraded.

For information about Red Hat support of workbench images and packages, see [Red Hat OpenShift AI: Supported Configurations](#).

Red Hat OpenShift AI contains the following notebook images that are available by default.



IMPORTANT

Notebook images denoted with **(Technology Preview)** in this table are not supported with Red Hat production service level agreements (SLAs) and might not be functionally complete. Red Hat does not recommend using Technology Preview features in production. These features provide early access to upcoming product features, enabling customers to test functionality and provide feedback during the development process. For more information about the support scope of Red Hat Technology Preview features, see [Technology Preview Features Support Scope](#).

Table 2.1. Default notebook images

Image name	Description
CUDA	If you are working with compute-intensive data science models that require GPU support, use the Compute Unified Device Architecture (CUDA) notebook image to gain access to the NVIDIA CUDA Toolkit. Using this toolkit, you can optimize your work by using GPU-accelerated libraries and optimization tools.
Standard Data Science	Use the Standard Data Science notebook image for models that do not require TensorFlow or PyTorch. This image contains commonly-used libraries to assist you in developing your machine learning models.
TensorFlow	TensorFlow is an open source platform for machine learning. With TensorFlow, you can build, train and deploy your machine learning models. TensorFlow contains advanced data visualization features, such as computational graph visualizations. It also allows you to easily monitor and track the progress of your models.
PyTorch	PyTorch is an open source machine learning library optimized for deep learning. If you are working with computer vision or natural language processing models, use the Pytorch notebook image.
Minimal Python	If you do not require advanced machine learning features, or additional resources for compute-intensive data science work, you can use the Minimal Python image to develop your models.
TrustyAI	Use the TrustyAI notebook image to leverage your data science work with model explainability, tracing, and accountability, and runtime monitoring.
HabanaAI	The HabanaAI notebook image optimizes high-performance deep learning (DL) with Intel Gaudi AI accelerators. Intel Gaudi AI accelerators accelerate DL training workloads and maximize training throughput and efficiency.

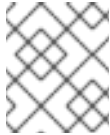

Image name	Description
code-server (Technology Preview)	<p>With the code-server notebook image, you can customize your notebook environment to meet your needs using a variety of extensions to add new languages, themes, debuggers, and connect to additional services. Enhance the efficiency of your data science work with syntax highlighting, auto-indentation, and bracket matching, as well as an automatic task runner for seamless automation. For more information, see code-server in GitHub.</p> <div data-bbox="384 443 494 577" style="display: inline-block; vertical-align: middle;">  </div> <p style="margin-left: 20px;">NOTE</p> <p>Elyra-based pipelines are not available with the code-server notebook image.</p>
RStudio Server (Technology preview)	<p>Use the RStudio Server notebook image to access the RStudio IDE, an integrated development environment for R, a programming language for statistical computing and graphics. For more information, see the RStudio Server site.</p> <p>To use the RStudio Server notebook image, you must first build it by creating a secret and triggering the BuildConfig, and then enable it in the OpenShift AI UI by editing the rstudio-rhel9 image stream. For more information, see Building the RStudio Server workbench images.</p> <div data-bbox="384 1032 494 1288" style="display: inline-block; vertical-align: middle;">  </div> <p style="margin-left: 20px;">IMPORTANT</p> <p>Disclaimer: Red Hat supports managing workbenches in OpenShift AI. However, Red Hat does not provide support for the RStudio software. RStudio Server is available through https://rstudio.org/ and is subject to RStudio licensing terms. Review the licensing terms before you use this sample workbench.</p>

Image name	Description
CUDA - RStudio Server (Technology preview)	<p>Use the CUDA - RStudio Server notebook image to access the RStudio IDE and NVIDIA CUDA Toolkit. RStudio is an integrated development environment for R, a programming language for statistical computing and graphics. With the NVIDIA CUDA toolkit, you can optimize your work using GPU-accelerated libraries and optimization tools. For more information, see the RStudio Server site.</p> <p>To use the CUDA - RStudio Server notebook image, you must first build it by creating a secret and triggering the BuildConfig, and then enable it in the OpenShift AI UI by editing the cuda-rstudio-rhel9 image stream. For more information, see Building the RStudio Server workbench images.</p> <div data-bbox="386 618 491 1025" style="background-color: black; width: 66px; height: 182px; margin-bottom: 10px;"> </div> <p>IMPORTANT</p> <p>Disclaimer: Red Hat supports managing workbenches in OpenShift AI. However, Red Hat does not provide support for the RStudio software. RStudio Server is available through https://rstudio.org/ and is subject to RStudio licensing terms. Review the licensing terms before you use this sample workbench.</p> <p>The CUDA - RStudio Server notebook image contains NVIDIA CUDA technology. CUDA licensing information is available at https://docs.nvidia.com/cuda/. Review the licensing terms before you use this sample workbench.</p>

2.1.2. Creating a workbench

When you create a workbench, you specify an image (an IDE, packages, and other dependencies). You can also configure data connections, cluster storage, and add container storage.

Prerequisites

- You have logged in to Red Hat OpenShift AI.
- If you use specialized OpenShift AI groups, you are part of the user group or admin group (for example, **rhoai-users** or **rhoai-admins**) in OpenShift.
- You created a project.
- If you created a Simple Storage Service (S3) account outside of Red Hat OpenShift AI and you want to create data connections to your existing S3 storage buckets, you have the following credential information for the storage buckets:
 - Endpoint URL
 - Access key
 - Secret key
 - Region
 - Bucket name

For information about working with data stored in AWS S3, see [Working with data in an S3-compatible object store](#).

Procedure

1. From the OpenShift AI dashboard, click **Data Science Projects**.
2. Click the name of the project that you want to add the workbench to. A **Details** page for the project opens.
3. In the **Workbenches** section, click **Create a workbench**
4. In the **Create workbench** page, configure the properties of the workbench that you are creating.
 - a. In the **Name** field, enter a name for your workbench.
 - b. Optional: In the **Description** field, enter a description to define your workbench.
 - c. In the **Notebook image** section, complete the fields to specify the workbench image to use with your workbench.
From the **Image selection** list, select a workbench image that suits your use case. A workbench image includes an IDE and Python packages (reusable code). Optionally, click the **View package information** option to view a list of packages that are included in the image that you selected.

If the workbench image has multiple versions available, select the workbench image version to use from the **Versions** section. To use the latest package versions, Red Hat recommends that you use the most recently added image.



NOTE

You can change the workbench image after you create the workbench.

- d. In the **Deployment size** section, from the **Container size** list, select a container size for your server. The container size controls the number of CPUs, the amount of memory, and the minimum and maximum request capacity of the container.
- e. Optional: Select and specify values for any environment variables.
Setting environment variables during the workbench configuration helps you save time later because you do not need to define them in the body of your notebooks, or with the IDE command line interface.

If you are using S3-compatible storage, add these recommended environment variables:

- **AWS_ACCESS_KEY_ID** specifies your Access Key ID for Amazon Web Services.
- **AWS_SECRET_ACCESS_KEY** specifies your Secret access key for the account specified in **AWS_ACCESS_KEY_ID**.

OpenShift AI stores the credentials as Kubernetes secrets in a protected namespace if you select **Secret** when you add the variable.

- f. Configure the storage for your workbench. Select one of the following options:

- **Create new persistent storage** to create storage that is retained after you shut down your workbench. Complete the relevant fields to define the storage.
 - **Use existing persistent storage** to reuse existing storage and select the storage from the **Persistent storage** list.
- g. Optionally, you can add a data connection to your workbench. A data connection is a resource that contains the configuration parameters needed to connect to a data source or an object storage bucket. Currently, only S3-Compatible data connections are supported. You can use storage buckets for storing data, models, and pipeline artifacts. You can also use a data connection to specify the location of a model that you want to deploy. In the **Data connections** section, select the **Use a data connection** checkbox.
- Create a new data connection as follows:
 - i. Select **Create new data connection**
 - ii. In the **Name** field, enter a unique name for the data connection.
 - iii. In the **Access key** field, enter the access key ID for the S3-compatible object storage provider.
 - iv. In the **Secret key** field, enter the secret access key for the S3-compatible object storage account that you specified.
 - v. In the **Endpoint** field, enter the endpoint of your S3-compatible object storage bucket.
 - vi. In the **Region** field, enter the default region of your S3-compatible object storage account.
 - vii. In the **Bucket** field, enter the name of your S3-compatible object storage bucket.
 - Use an existing data connection as follows:
 - i. Select **Use existing data connection**
 - ii. From the **Data connection** list, select a data connection that you previously defined.

5. Click **Create workbench**.

Verification

- The workbench that you created appears on the **Workbenches** tab for the project.
- Any cluster storage that you associated with the workbench during the creation process appears on the **Cluster storage** tab for the project.
- The **Status** column on the **Workbenches** tab displays a status of **Starting** when the workbench server is starting, and **Running** when the workbench has successfully started.
- Optionally, click the **Open** link to open the IDE in a new window.

2.2. STARTING A WORKBENCH

You can manually start a data science project's workbench from the **Workbenches** tab on the project details page. By default, workbenches start immediately after you create them.

Prerequisites

- You have logged in to Red Hat OpenShift AI.
- If you are using specialized OpenShift AI groups, you are part of the user group or admin group (for example, **rhoai-users** or **rhoai-admins**) in OpenShift.
- You have created a data science project that contains a workbench.

Procedure

1. From the OpenShift AI dashboard, click **Data Science Projects**.
The **Data Science Projects** page opens.
2. Click the name of the project whose workbench you want to start.
A project details page opens.
3. Click the **Workbenches** tab.
4. Click the toggle in the **Status** column for the relevant workbench to start a workbench that is not running.
The status of the workbench that you started changes from **Stopped** to **Running**. After the workbench has started, click **Open** to open the workbench's notebook.

Verification

- The workbench that you started appears on the **Workbenches** tab for the project, with the status of **Running**.

2.3. UPDATING A PROJECT WORKBENCH


If your data science work requires you to change your workbench's notebook image, container size, or identifying information, you can update the properties of your project's workbench. If you require extra power for use with large datasets, you can assign accelerators to your workbench to optimize performance.

Prerequisites

- You have logged in to Red Hat OpenShift AI.
- If you use specialized OpenShift AI groups, you are part of the user group or admin group (for example, **rhoai-users** or **rhoai-admins**) in OpenShift.
- You have created a data science project that has a workbench.

Procedure

1. From the OpenShift AI dashboard, click **Data Science Projects**.
The **Data Science Projects** page opens.
2. Click the name of the project whose workbench you want to update.
A project details page opens.

3. Click the **Workbenches** tab.
4. Click the action menu () beside the workbench that you want to update and then click **Edit workbench**.
The **Edit workbench** page opens.
5. Update any of the workbench properties and then click **Update workbench**.

Verification

- The workbench that you updated appears on the **Workbenches** tab for the project.


2.4. DELETING A WORKBENCH FROM A DATA SCIENCE PROJECT

You can delete workbenches from your data science projects to help you remove Jupyter notebooks that are no longer relevant to your work.

Prerequisites

- You have logged in to Red Hat OpenShift AI.
- If you are using specialized OpenShift AI groups, you are part of the user group or admin group (for example, **rhoai-users** or **rhoai-admins**) in OpenShift.
- You have created a data science project with a workbench.

Procedure

1. From the OpenShift AI dashboard, click **Data Science Projects**.
The **Data Science Projects** page opens.
2. Click the name of the project that you want to delete the workbench from.
A project details page opens.
3. Click the **Workbenches** tab.
4. Click the action menu () beside the workbench that you want to delete and then click **Delete workbench**.
The **Delete workbench** dialog opens.
5. Enter the name of the workbench in the text field to confirm that you intend to delete it.
6. Click **Delete workbench**.

Verification

- The workbench that you deleted is no longer displayed on the **Workbenches** tab for the project.
- The custom resource (CR) associated with the workbench's Jupyter notebook is deleted.

CHAPTER 3. USING DATA CONNECTIONS

3.1. ADDING A DATA CONNECTION TO YOUR DATA SCIENCE PROJECT

You can enhance your data science project by adding a connection to a data source. When you want to work with a very large data sets, you can store your data in an S3-compatible object storage bucket, so that you do not fill up your local storage. You also have the option of associating the data connection with an existing workbench that does not already have a connection.

Prerequisites

- You have logged in to Red Hat OpenShift AI.
- If you are using specialized OpenShift AI groups, you are part of the user group or admin group (for example, **rhoai-users** or **rhoai-admins**) in OpenShift.
- You have created a data science project that you can add a data connection to.
- You have access to S3-compatible object storage.
- If you intend to add the data connection to an existing workbench, you have saved any data in the workbench to avoid losing work.

Procedure

1. From the OpenShift AI dashboard, click **Data Science Projects**.
The **Data Science Projects** page opens.
2. Click the name of the project that you want to add a data connection to.
A project details page opens.
3. Click the **Data connections** tab.
4. Click **Add data connection**
The **Add data connection** dialog opens.
5. Enter a **name** for the data connection.
6. In the **Access key** field, enter the access key ID for your S3-compatible object storage provider.
7. In the **Secret key** field, enter the secret access key for the S3-compatible object storage account you specified.
8. In the **Endpoint** field, enter the endpoint of your S3-compatible object storage bucket.
9. In the **Region** field, enter the default region of your S3-compatible object storage account.
10. In the **Bucket** field, enter the name of your S3-compatible object storage bucket.
11. Optional: From the **Connected workbench** list, select a workbench to connect.
12. Click **Add data connection**

Verification

- The data connection that you added appears on the **Data connections** tab for the project.
- If you selected a workbench, the workbench is visible in the **Connected workbenches** column on the **Data connections** tab for the project.

3.2. DELETING A DATA CONNECTION

You can delete data connections from your data science projects to help you remove connections that are no longer relevant to your work.

Prerequisites

- You have logged in to Red Hat OpenShift AI.
- If you are using specialized OpenShift AI groups, you are part of the user group or admin group (for example, **rhoai-users** or **rhoai-admins**) in OpenShift.
- You have created a data science project with a data connection.

Procedure

1. From the OpenShift AI dashboard, click **Data Science Projects**.
The **Data Science Projects** page opens.
2. Click the name of the project that you want to delete the data connection from.
A project details page opens.
3. Click the **Data connections** tab.
4. Click the action menu (**⋮**) beside the data connection that you want to delete and then click **Delete data connection**.
The **Delete data connection** dialog opens.
5. Enter the name of the data connection in the text field to confirm that you intend to delete it.
6. Click **Delete data connection**.

Verification

- The data connection that you deleted is no longer displayed on the **Data connections** tab for the project.

3.3. UPDATING A CONNECTED DATA SOURCE


To use an existing data source with a different workbench, you can change the data source that is connected to your project's workbench.

Prerequisites

- You have logged in to Red Hat OpenShift AI.
- If you are using specialized OpenShift AI groups, you are part of the user group or admin group (for example, **rhoai-users** or **rhoai-admins**) in OpenShift.

- You have created a data science project, created a workbench, and you have defined a data connection.

Procedure

1. From the OpenShift AI dashboard, click **Data Science Projects**.
The **Data Science Projects** page opens.
2. Click the name of the project whose data source you want to change.
A project details page opens.
3. Click the **Data connections** tab.
4. Click the action menu () beside the data source that you want to change and then click **Edit data connection**.
The **Edit data connection** dialog opens.
5. In the **Connected workbench** section, select an existing workbench from the list.
6. Click **Update data connection**

Verification

- The updated data connection is displayed on the **Data connections** tab for the project.
- You can access your S3 data source using environment variables in the connected workbench.

CHAPTER 4. CONFIGURING CLUSTER STORAGE

4.1. ADDING CLUSTER STORAGE TO YOUR DATA SCIENCE PROJECT

For data science projects that require data to be retained, you can add cluster storage to the project. Additionally, you can also connect cluster storage to a specific project's workbench.

Prerequisites

- You have logged in to Red Hat OpenShift AI.
- If you are using specialized OpenShift AI groups, you are part of the user group or admin group (for example, **rhoai-users** or **rhoai-admins**) in OpenShift.
- You have created a data science project that you can add cluster storage to.

Procedure

1. From the OpenShift AI dashboard, click **Data Science Projects**.
The **Data Science Projects** page opens.
2. Click the name of the project that you want to add the cluster storage to.
A project details page opens.
3. Click the **Cluster storage** tab.
4. Click **Add cluster storage**.
The **Add storage** dialog opens.
5. Enter a **name** for the cluster storage.
6. Enter a **description** for the cluster storage.
7. Under **Persistent storage size**, enter a new size in gibibytes. The minimum size is 1 GiB, and the maximum size is 16384 GiB.
8. Optional: Select a **workbench** from the list to connect the cluster storage to an existing workbench.
9. If you selected a workbench to connect the storage to, enter the storage directory in the **Mount folder** field.
10. Click **Add storage**.

Verification

- The cluster storage that you added appears on the **Cluster storage** tab for the project.
- A new persistent volume claim (PVC) is created with the storage size that you defined.
- The persistent volume claim (PVC) is visible as an attached storage on the **Workbenches** tab for the project.


4.2. UPDATING CLUSTER STORAGE

If your data science work requires you to change the identifying information of a project's cluster storage or the workbench that the storage is connected to, you can update your project's cluster storage to change these properties.

Prerequisites

- You have logged in to Red Hat OpenShift AI.
- If you are using specialized OpenShift AI groups, you are part of the user group or admin group (for example, **rhoai-users** or **rhoai-admins**) in OpenShift.
- You have created a data science project that contains cluster storage.

Procedure

1. From the OpenShift AI dashboard, click **Data Science Projects**.
The **Data Science Projects** page opens.
2. Click the name of the project whose storage you want to update.
A project details page opens.
3. Click the **Cluster storage** tab.
4. Click the action menu () beside the storage that you want to update and then click **Edit storage**.
The **Edit storage** page opens.
5. Update the storage's properties.
 - a. Update the **name** for the storage, if applicable.
 - b. Update the **description** for the storage, if applicable.
 - c. Increase the **Persistent storage size** for the storage, if applicable.
Note that you can only increase the storage size. Updating the storage size restarts the workbench and makes it unavailable for a period of time that is usually proportional to the size change.
 - d. Update the **workbench** that the storage is connected to, if applicable.
 - e. If you selected a new workbench to connect the storage to, enter the storage directory in the **Mount folder** field.
6. Click **Update storage**.

If you increased the storage size, the workbench restarts and is unavailable for a period of time that is usually proportional to the size change.

Verification

- The storage that you updated appears on the **Cluster storage** tab for the project.

4.3. DELETING CLUSTER STORAGE FROM A DATA SCIENCE PROJECT

You can delete cluster storage from your data science projects to help you free up resources and delete unwanted storage space.

Prerequisites

- You have logged in to Red Hat OpenShift AI.
- If you are using specialized OpenShift AI groups, you are part of the user group or admin group (for example, **rhoai-users** or **rhoai-admins**) in OpenShift.
- You have created a data science project with cluster storage.

Procedure

1. From the OpenShift AI dashboard, click **Data Science Projects**.
The **Data Science Projects** page opens.
2. Click the name of the project that you want to delete the storage from.
A project details page opens.
3. Click the **Cluster storage** tab.
4. Click the action menu (**⋮**) beside the storage that you want to delete and then click **Delete storage**.
The **Delete storage** dialog opens.
5. Enter the name of the storage in the text field to confirm that you intend to delete it.
6. Click **Delete storage**.

Verification

- The storage that you deleted is no longer displayed on the **Cluster storage** tab for the project.
- The persistent volume (PV) and persistent volume claim (PVC) associated with the cluster storage are both permanently deleted. This data is not recoverable.

CHAPTER 5. MANAGING ACCESS TO DATA SCIENCE PROJECTS

5.1. CONFIGURING ACCESS TO A DATA SCIENCE PROJECT

To enable you to work collaboratively on your data science projects with other users, you can share access to your project. After creating your project, you can then set the appropriate access permissions from the OpenShift AI user interface.

You can assign the following access permission levels to your data science projects:

- **Admin** - Users can modify all areas of a project, including its details (project name and description), components, and access permissions.
- **Contributor** - Users can modify a project's components, such as its workbench, but they cannot edit a project's access permissions or its details (project name and description).

5.2. SHARING ACCESS TO A DATA SCIENCE PROJECT


To enable your organization to work collaboratively, you can share access to your data science project with other users and groups.

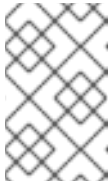
Prerequisites

- You have logged in to Red Hat OpenShift AI.
- If you are using specialized OpenShift AI groups, you are part of the user group or admin group (for example, **rhoai-users** or **rhoai-admins**) in OpenShift.
- You have created a data science project.

Procedure


1. From the OpenShift AI dashboard, click **Data Science Projects**.
The **Data Science Projects** page opens.
2. From the list of data science projects, click the name of the data science project that you want to share access to.
A project details page opens.
3. Click the **Permissions** tab.
The **Permissions** page for the project opens.
4. Provide one or more users with access to the project.
 - a. In the **Users** section, click **Add user**.
 - b. In the **Name** field, enter the user name of the user whom you want to provide access to the project.
 - c. From the **Permissions** list, select one of the following access permission levels:
 - **Admin**: Users with this access level can edit project details and manage access to the project.

- Contributor: Users with this access level can view and edit project components, such as its workbenches, data connections, and storage.
- d. To confirm your entry, click **Confirm** ().
 - e. Optional: To add an additional user, click **Add user** and repeat the process.
5. Provide one or more OpenShift groups with access to the project.
 - a. In the **Groups** section, click **Add group**.
 - b. From the **Name** list, select a group to provide access to the project.



NOTE

If you do not have **cluster-admin** or **dedicated-admin** permissions, the **Name** list is not visible. Instead, an input field is displayed enabling you to configure group permissions.

- c. From the **Permissions** list, select one of the following access permission levels:
 - Admin: Groups with this access permission level can edit project details and manage access to the project.
 - Edit: Groups with this access permission level can view and edit project components, such as its workbenches, data connections, and storage.
- d. To confirm your entry, click **Confirm** ().
- e. Optional: To add an additional group, click **Add group** and repeat the process.

Verification

- Users to whom you provided access to the project can perform only the actions permitted by their access permission level.
- The **Users** and **Groups** sections on the **Permissions** tab show the respective users and groups that you provided with access to the project.

5.3. UPDATING ACCESS TO A DATA SCIENCE PROJECT


To change the level of collaboration on your data science project, you can update the access permissions of users and groups who have access to your project.

Prerequisites

- You have logged in to Red Hat OpenShift AI.
- If you are using specialized OpenShift AI groups, you are part of the user group or admin group (for example, **rhoai-users** or **rhoai-admins**) in OpenShift.
- You have created a data science project.
- You have previously shared access to your project with other users or groups.

- You have administrator permissions or you are the project owner.


Procedure

1. From the OpenShift AI dashboard, click **Data Science Projects**.
The **Data Science Projects** page opens.
2. Click the name of the project that you want to change the access permissions of.
A project details page opens.
3. Click the **Permissions** tab.
The **Permissions** page for the project opens.
4. Update the user access permissions to the project.
 - a. In the **Name** field, update the user name of the user whom you want to provide access to the project.
 - b. From the **Permissions** list, update the user access permissions by selecting one of the following:
 - Admin: Users with this access level can edit project details and manage access to the project.
 - Contributor: Users with this access level can view and edit project components, such as its workbenches, data connections, and storage.
 - c. To confirm the update to the entry, click **Confirm** ().
5. Update the OpenShift groups access permissions to the project.
 - a. From the **Name** list, update the group that has access to the project by selecting another group from the list.



NOTE

If you do not have **cluster-admin** or **dedicated-admin** permissions, the **Name** list is not visible. Instead, an input field displays enabling you to configure group permissions.

- b. From the **Permissions** list, update the group access permissions by selecting one of the following:
 - Admin: Groups with this access permission level can edit project details and manage access to the project.
 - Edit: Groups with this access permission level can view and edit project components, such as its workbenches, data connections, and storage.
- c. To confirm the update to the entry, click **Confirm** ().

Verification

- The **Users** and **Groups** sections on the **Permissions** tab show the respective users and groups whose project access permissions you changed.


5.4. REMOVING ACCESS TO A DATA SCIENCE PROJECT

If you no longer want to work collaboratively on your data science project, you can restrict access to your project by removing users and groups that you previously provided access to your project.

Prerequisites

- You have logged in to Red Hat OpenShift AI.
- If you are using specialized OpenShift AI groups, you are part of the user group or admin group (for example, **rhoai-users** or **rhoai-admins**) in OpenShift.
- You have created a data science project.
- You have previously shared access to your project with other users or groups.
- You have administrator permissions or you are the project owner.

Procedure

1. From the OpenShift AI dashboard, click **Data Science Projects**.
The **Data Science Projects** page opens.
2. Click the name of the project that you want to change the access permissions of.
A project details page opens.
3. Click the **Permissions** tab.
The **Permissions** page for the project opens.
4. Click the action menu () beside the user or group whose access permissions you want to revoke and click **Delete**.

Verification

- Users whose access you have revoked can no longer perform the actions that were permitted by their access permission level.