



Red Hat OpenShift AI Cloud Service 1

Getting started with Red Hat OpenShift AI Cloud Service

Learn how to work in an OpenShift AI environment

Red Hat OpenShift AI Cloud Service 1 Getting started with Red Hat OpenShift AI Cloud Service

Learn how to work in an OpenShift AI environment

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

Learn how to work in an OpenShift AI environment.


Table of Contents

CHAPTER 1. LOGGING IN TO OPENSIFT AI	3
CHAPTER 2. THE OPENSIFT AI USER INTERFACE	4
2.1. GLOBAL NAVIGATION	5
2.2. SIDE NAVIGATION	5
CHAPTER 3. NOTIFICATIONS IN OPENSIFT AI	8
CHAPTER 4. CREATING A DATA SCIENCE PROJECT	9
CHAPTER 5. CREATING A PROJECT WORKBENCH	10
5.1. LAUNCHING JUPYTER AND STARTING A NOTEBOOK SERVER	11
5.2. OPTIONS FOR NOTEBOOK SERVER ENVIRONMENTS	13
CHAPTER 6. TUTORIALS FOR DATA SCIENTISTS	32
6.1. ACCESSING TUTORIALS	35
CHAPTER 7. CONFIGURING YOUR IDE	36
7.1. CONFIGURING YOUR CODE-SERVER WORKBENCH	36
7.1.1. Installing extensions with code-server	36
7.1.1.1. Extensions	37
7.2. BUILDING THE RSTUDIO SERVER NOTEBOOK IMAGES	37
CHAPTER 8. ENABLING SERVICES CONNECTED TO OPENSIFT AI	39
CHAPTER 9. DISABLING APPLICATIONS CONNECTED TO OPENSIFT AI	41
9.1. REMOVING DISABLED APPLICATIONS FROM OPENSIFT AI	42
CHAPTER 10. SUPPORT REQUIREMENTS AND LIMITATIONS	43
10.1. SUPPORTED BROWSERS	43
10.2. SUPPORTED SERVICES	43
10.3. SUPPORTED PACKAGES	44
CHAPTER 11. COMMON QUESTIONS	45
CHAPTER 12. TROUBLESHOOTING COMMON PROBLEMS IN JUPYTER FOR ADMINISTRATORS	46
12.1. A USER RECEIVES A 404: PAGE NOT FOUND ERROR WHEN LOGGING IN TO JUPYTER	46
12.2. A USER'S NOTEBOOK SERVER DOES NOT START	46
12.3. THE USER RECEIVES A DATABASE OR DISK IS FULL ERROR OR A NO SPACE LEFT ON DEVICE ERROR WHEN THEY RUN NOTEBOOK CELLS	47
CHAPTER 13. TROUBLESHOOTING COMMON PROBLEMS IN JUPYTER FOR USERS	49
13.1. I SEE A 403: FORBIDDEN ERROR WHEN I LOG IN TO JUPYTER	49
13.2. MY NOTEBOOK SERVER DOES NOT START	49
13.3. I SEE A DATABASE OR DISK IS FULL ERROR OR A NO SPACE LEFT ON DEVICE ERROR WHEN I RUN MY NOTEBOOK CELLS	49

CHAPTER 1. LOGGING IN TO OPENSIFT AI

Log in to OpenShift AI from a browser for easy access to Jupyter and your data science projects.

Procedure

1. Browse to the OpenShift AI instance URL and click **Log in with OpenShift**
 - If you are a data scientist user, your administrator must provide you with the OpenShift AI instance URL, for example, **`https://rhods-dashboard-redhat-oai-applications.apps.example.abc1.p1.openshiftapps.com/`**
 - If you have access to OpenShift, you can browse to the OpenShift web console and click the **Application Launcher** () → **Red Hat OpenShift AI**.
2. Click the name of your identity provider, for example, **GitHub**.
3. Enter your credentials and click **Log in** (or equivalent for your identity provider).

Verification

- OpenShift AI opens on the **Enabled applications** page.

Troubleshooting

- If you see **An authentication error occurred** or **Could not create user** when you try to log in:
 - You might have entered your credentials incorrectly. Confirm that your credentials are correct.
 - You might have an account in more than one configured identity provider. If you have logged in with a different identity provider previously, try again with that identity provider.

Additional resources

- [Launching Jupyter and starting a notebook server](#)

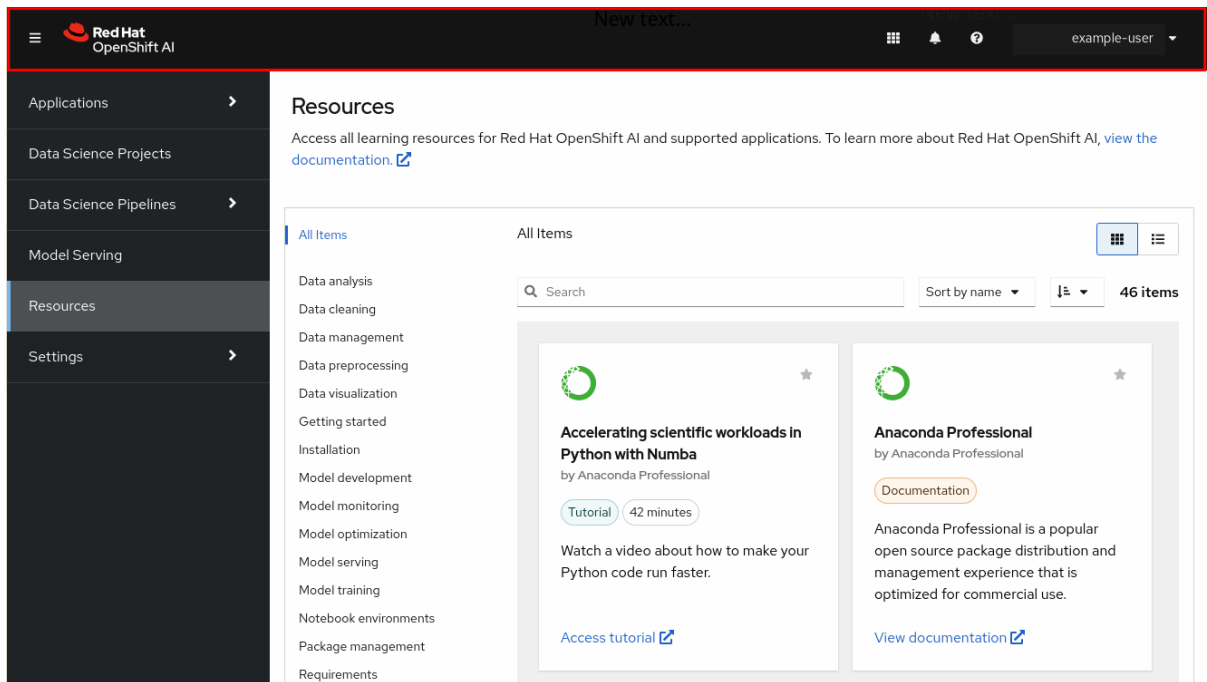
CHAPTER 2. THE OPENSIFT AI USER INTERFACE

The Red Hat OpenShift AI interface is based on the OpenShift web console user interface.

The Red Hat OpenShift AI user interface is divided into several areas:

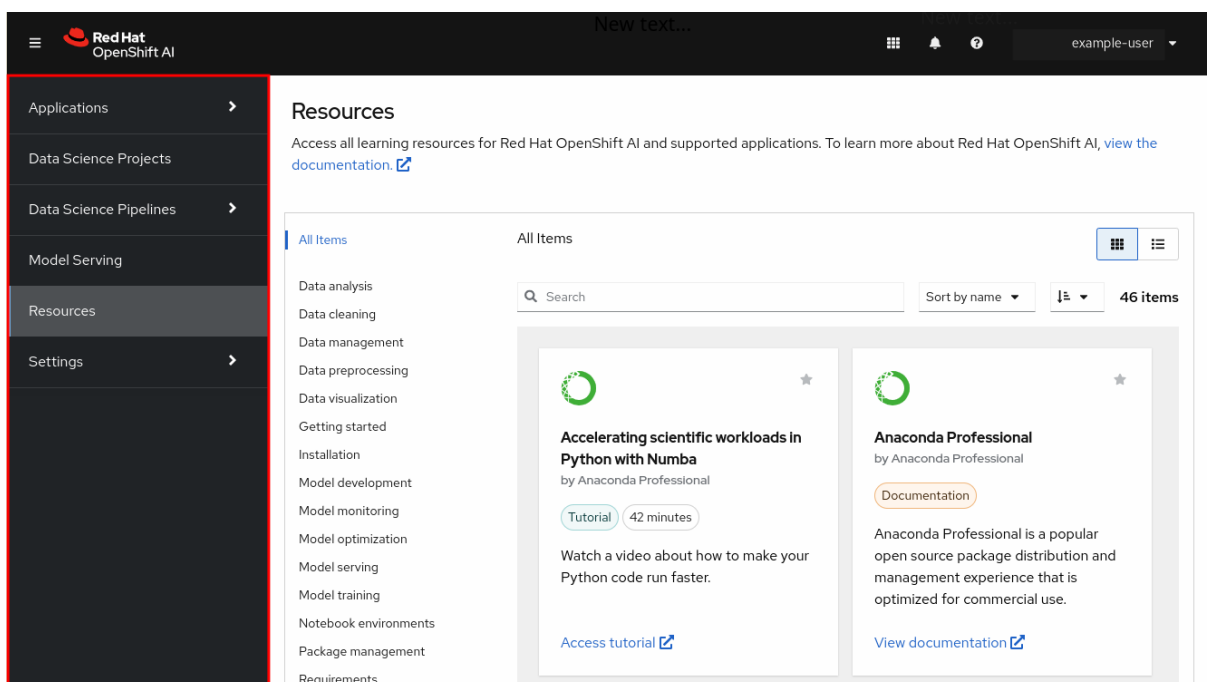
- The global navigation bar, which provides access to useful controls, such as **Help** and **Notifications**.

Figure 2.1. The global navigation bar



- The side navigation menu, which contains different categories of pages available in OpenShift AI.

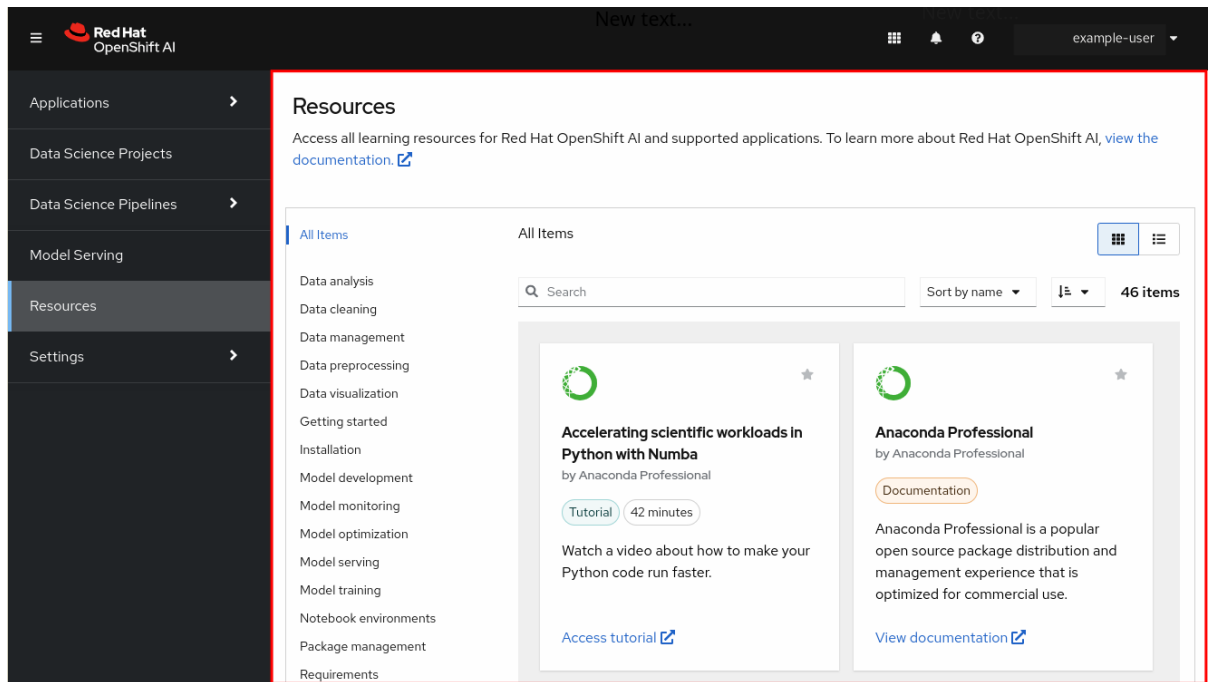
Figure 2.2. The side navigation menu



- The main display area, which displays the current page and shares space with any drawers




currently displaying information, such as notifications or quick start guides. The main display area also displays the **Notebook server control panel** where you can launch Jupyter by starting and configuring a notebook server. Administrators can also use the **Notebook server control panel** to manage other users' notebook servers.

Figure 2.3. The main display area



2.1. GLOBAL NAVIGATION

There are four items in the top navigation:

- The **Toggle side navigation menu** button () toggles whether or not the side navigation is displayed.
- The **Notifications** button () opens and closes the **Notifications** drawer, letting you read current and previous notifications in more detail.
- The **Help** menu () provides a link to create a ticket with Red Hat Support and access the OpenShift AI documentation.
- The **User** menu displays the name of the currently logged-in user and provides access to the **Log out** button.

2.2. SIDE NAVIGATION

There are several different pages in the side navigation:

Applications → Enabled

The **Enabled** page displays applications that are enabled and ready to use on OpenShift AI. This page is the default landing page for OpenShift AI.

Click the **Launch application** button on an application tile to open the application interface in a new tab. If an application has an associated quick start tour, click the drop-down menu on the application tile and select **Open quick start** to access it. This page also displays applications and components

that have been disabled by your administrator. Disabled applications are denoted with **Disabled** on the application tile. Click **Disabled** on the application tile to access links allowing you to remove the tile itself, and to revalidate its license, if the license had previously expired.

Applications → Explore

The **Explore** page displays applications that are available for use with OpenShift AI. Click a tile for more information about the application or to access the **Enable** button. The **Enable** button is visible only if an application does not require an OpenShift Operator installation.

Data Science Projects

The **Data Science Projects** page allows you to organize your data science work into a single project. From this page, you can create and manage data science projects. You can also enhance the capabilities of your data science project by adding workbenches, adding storage to your project's cluster, adding data connections, and adding model servers.

Data Science Pipelines → Pipelines

The **Pipelines** page allows you to import, manage, track, and view data science pipelines. Using Red Hat OpenShift AI pipelines, you can standardize and automate machine learning workflows to enable you to develop and deploy your data science models.

Data Science Pipelines → Runs

The **Runs** page allows you to define, manage, and track executions of a data science pipeline. A pipeline run is a single execution of a data science pipeline. You can also view a record of previously executed and scheduled runs for your data science project.

Model Serving

The **Model Serving** page allows you to manage and view the status of your deployed models. You can use this page to deploy data science models to serve intelligent applications, or to view existing deployed models. You can also determine the inference endpoint of a deployed model.

Resources

The **Resources** page displays learning resources such as documentation, how-to material, and quick start tours. You can filter visible resources using the options displayed on the left, or enter terms into the search bar.

Settings → Notebook images

The **Notebook images** page allows you to configure custom notebook images that cater to your project's specific requirements. After you have added custom notebook images to your deployment of OpenShift AI, they are available for selection when creating a notebook server.

Settings → Cluster settings

The **Cluster settings** page allows you to perform the following administrative tasks on your cluster:

- Enable or disable Red Hat's ability to collect data about OpenShift AI usage on your cluster.
- Configure how resources are claimed within your cluster by changing the default size of the cluster's persistent volume claim (PVC).
- Reduce resource usage in your OpenShift AI deployment by stopping notebook servers that have been idle.
- Schedule notebook pods on tainted nodes by adding tolerations.

Settings → Accelerator profiles

The **Accelerator profiles** page allows you to perform the following administrative tasks on your accelerator profiles:

- Enable or disable an existing accelerator profile.
- Create, update, or delete accelerator profiles.
- Schedule pods on tainted nodes by adding tolerations.

Settings → Serving runtimes

The **Serving runtimes** page allows you to manage the model-serving runtimes in your OpenShift AI deployment. You can use this page to add, edit, and enable or disable model-serving runtimes. You specify a model-serving runtime when you configure a model server on the **Data Science Projects** page.

Settings → User management

The **User management** page allows you to define OpenShift AI user group and admin group membership.

CHAPTER 3. NOTIFICATIONS IN OPENSIFT AI

Red Hat OpenShift AI displays notifications when important events happen in the cluster.


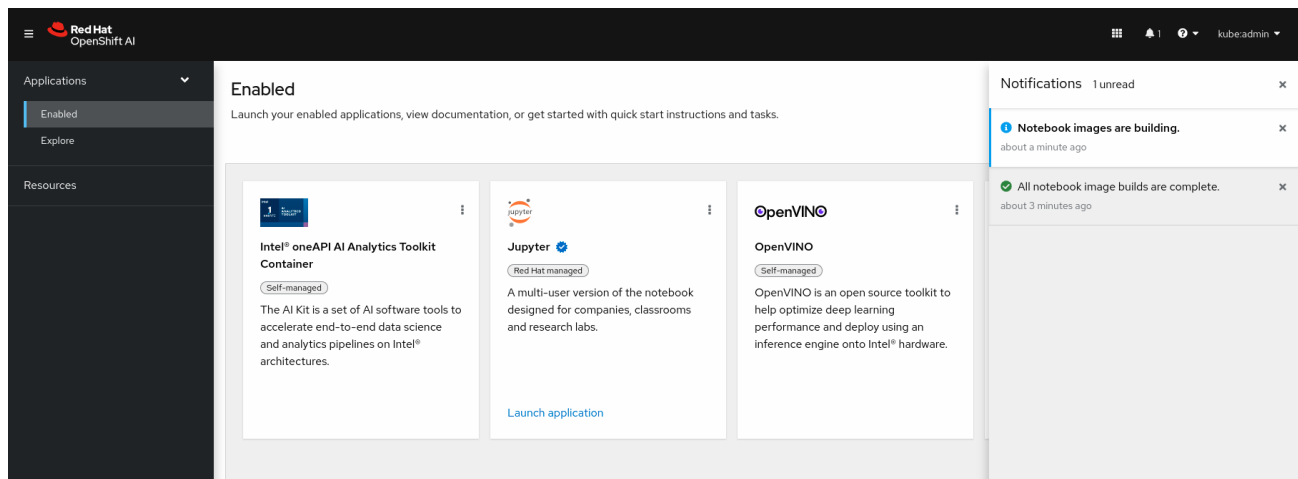
If you miss a notification message, click the **Notifications** button () to open the **Notifications** drawer and view unread messages.

Figure 3.1. The Notifications drawer



CHAPTER 4. CREATING A DATA SCIENCE PROJECT

To start your data science work, create a data science project. Creating a project helps you organize your work in one place. You can also enhance your data science project by adding the following functionality:

- Workbenches
- Storage for your project's cluster
- Data connections
- Data science pipelines
- Model servers

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, click **Data Science Projects**.
The **Data Science Projects** page opens.
2. Click **Create data science project**
The **Create a data science project** dialog opens.
3. Enter a **name** for your data science 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.
5. Enter a **description** for your data science project.
6. Click **Create**.
A project details page opens. From this page, you can create workbenches, add cluster storage and data connections, import pipelines, and deploy models.

Verification

- The project that you created is displayed on the **Data Science Projects** page.

CHAPTER 5. CREATING A PROJECT WORKBENCH

To examine and work with models in an isolated area, you can create a workbench. You can use this workbench to create a Jupyter notebook from an existing notebook container image to access its resources and properties. For data science projects that require data retention, you can add container storage to the workbench you are creating. 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 you can add a workbench 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 workbench to.
A project details page opens.
3. Click the **Workbenches** tab.
4. Click **Create workbench**.
The **Create workbench** page opens.
5. Configure the properties of the workbench 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 notebook image to use with your workbench.
 - i. From the **Image selection** list, select a notebook image.
 - d. In the **Deployment size** section, specify the size of your deployment instance.
 - i. From the **Container size** list, select a container size for your server.
 - ii. Optional: From the **Accelerator** list, select an accelerator.
 - iii. If you selected an accelerator in the preceding step, specify the number of accelerators to use.
 - e. Optional: Select and specify values for any new **environment variables**.
- a. Configure the storage for your OpenShift AI cluster.
 - i. Select **Create new persistent storage** to create storage that is retained after you log out of OpenShift AI. Complete the relevant fields to define the storage.

- ii. Select **Use existing persistent storage** to reuse existing storage and select the storage from the **Persistent storage** list.
- b. To use a data connection, 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.
1. 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.

5.1. LAUNCHING JUPYTER AND STARTING A NOTEBOOK SERVER

Launch Jupyter and start a notebook server to start working with your notebooks. If you require extra power for use with large datasets, you can assign accelerators to your notebook server to optimize performance.

Prerequisites

- You have logged in to Red Hat OpenShift AI.
- You know the names and values you want to use for any environment variables in your notebook server environment, for example, **AWS_SECRET_ACCESS_KEY**.

- If you want to work with a large data set, work with your administrator to proactively increase the storage capacity of your notebook server. If applicable, also consider assigning accelerators to your notebook server.

Procedure

1. Locate the **Jupyter** tile on the **Enabled applications** page.

2. Click **Launch application**.

If you see an **Access permission needed** message, you are not in the default user group or the default administrator group for OpenShift AI. Ask your administrator to add you to the correct group by using [Adding users](#).

If you have not previously authorized the **jupyter-nb-`<username>`** service account to access your account, the **Authorize Access** page appears prompting you to provide authorization. Inspect the permissions selected by default, and click the **Allow selected permissions** button.

If your credentials are accepted, the **Notebook server control panel** opens displaying the **Start a notebook server** page.

3. Start a notebook server.

This is not required if you have previously opened Jupyter.

- a. In the **Notebook image** section, select the notebook image to use for your server.
- b. If the notebook image contains multiple versions, select the version of the notebook image from the **Versions** section.



NOTE

When a new version of a notebook image is released, the previous version remains available and supported on the cluster. This gives you time to migrate your work to the latest version of the notebook image.

- c. From the **Container size** list, select a suitable container size for your server.
- d. Optional: From the **Accelerator** list, select an accelerator.
- e. If you selected an accelerator in the preceding step, specify the number of accelerators to use.



IMPORTANT

Using accelerators is only supported with specific notebook images. For GPUs, only the PyTorch, TensorFlow, and CUDA notebook images are supported. For Habana Gaudi devices, only the HabanaAI notebook image is supported. In addition, you can only specify the number of accelerators required for your notebook server if accelerators are enabled on your cluster. To learn how to enable GPU support, see [Enabling GPU support in OpenShift AI](#).

- f. Optional: Select and specify values for any new **Environment variables**. The interface stores these variables so that you only need to enter them once. Example variable names for common environment variables are automatically provided for frequently integrated environments and frameworks, such as Amazon Web Services (AWS).



IMPORTANT

Select the **Secret** checkbox for variables with sensitive values that must remain private, such as passwords.

- g. Optional: Select the **Start server in current tab** checkbox if necessary.
- h. Click **Start server**.
The **Starting server** progress indicator appears. Click **Expand event log** to view additional information about the server creation process. Depending on the deployment size and resources you requested, starting the server can take up to several minutes. Click **Cancel** to cancel the server creation.

After the server starts, you see one of the following behaviors:

- If you previously selected the **Start server in current tab** checkbox, the JupyterLab interface opens in the current tab of your web browser.
- If you did not previously select the **Start server in current tab** checkbox, the **Starting server** dialog box prompts you to open the server in a new browser tab or in the current browser tab.
The JupyterLab interface opens according to your selection.

Verification

- The JupyterLab interface opens.

Additional resources

[Options for notebook server environments.](#)

Troubleshooting

- If you see the "Unable to load notebook server configuration options" error message, contact your administrator so that they can review the logs associated with your Jupyter pod and determine further details about the problem.

5.2. OPTIONS FOR NOTEBOOK SERVER ENVIRONMENTS

When you start Jupyter for the first time, or after stopping your notebook server, you must select server options in the **Start a notebook server** wizard so that the software and variables that you expect are available on your server. This section explains the options available in the **Start a notebook server** wizard in detail.

The **Start a notebook server** page consists of the following sections:

Notebook image

Specifies the container image that your notebook server is based on. Different notebook images have different packages installed by default. If the notebook image has multiple versions available, you can select the notebook image version to use from the **Versions** section.



NOTE

Notebook images are supported for a minimum of one year. Major updates to preconfigured notebook images occur about every six months. Therefore, two supported notebook image versions are typically available at any given time. Legacy notebook image versions, that is, not the two most recent versions, might still be available for selection. Legacy image versions include a label that indicates the image is out-of-date.

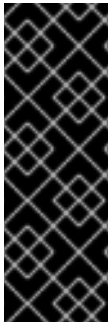
Versions 1.2 and 2023.1 of notebook images are no longer supported. Notebooks that are already running on versions 1.2 or 2023.1 of an image will continue to work normally, but they are not available to select for new users or notebooks.

If you want to use OpenShift AI with Data Science Pipelines (DSP) 2.0, you must use notebook image versions 2024.1 or later.

To use the latest package versions, Red Hat recommends that you use the most recently added notebook image.

After you start a notebook image, you can check which Python packages are installed on your notebook server and which version of the package you have by running the **pip** tool in a notebook cell.

The following table shows the package versions used in the available notebook images.



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 5.1. Notebook image options

Image name	Image version	Preinstalled packages
CUDA	2024.1 (Recommended)	<ul style="list-style-type: none"> ● CUDA 12.1 ● Python 3.9 ● JupyterLab 3.6 ● Notebook 6.5

Image name	Image version	Preinstalled packages
	2023.2	<ul style="list-style-type: none"> ● CUDA 11.8 ● Python 3.9 ● JupyterLab 3.6 ● Notebook 6.5
	2023.1 (Deprecated)	<ul style="list-style-type: none"> ● CUDA 11.8 ● Python 3.9 ● JupyterLab 3.5 ● Notebook 6.5
Minimal Python (default)	2024.1 (Recommended)	<ul style="list-style-type: none"> ● Python 3.9 ● JupyterLab 3.6 ● Notebook 6.5
	2023.2	<ul style="list-style-type: none"> ● Python 3.9 ● JupyterLab 3.6 ● Notebook 6.5
	2023.1 (Deprecated)	<ul style="list-style-type: none"> ● Python 3.9 ● JupyterLab 3.5 ● Notebook 6.5

Image name	Image version	Preinstalled packages
PyTorch	2024.1 (Recommended)	<ul style="list-style-type: none">● CUDA 12.1● Python 3.9● PyTorch 2.2● JupyterLab 3.6● Notebook 6.5● TensorBoard 2.16● Boto3 1.34● Kafka-Python 2.0● Kfp 2.7● Matplotlib 3.8● Numpy 1.26● Pandas 2.2● Scikit-learn 1.4● SciPy 1.12● ODH-Elyra 3.16● PyMongo 4.6● Pyodbc 5.1● Codeflare-SDK 0.16● Sklearn-onnx 1.16● Psycopg 3.1● MySQL Connector/Python 8.3

Image name	Image version	Preinstalled packages
	2023.2	<ul style="list-style-type: none">● CUDA 11.8● Python 3.9● PyTorch 2.0● JupyterLab 3.6● Notebook 6.5● TensorBoard 2.13● Boto3 1.28● Kafka-Python 2.0● Kfp-tekton 1.5● Matplotlib 3.6● Numpy 1.24● Pandas 1.5● Scikit-learn 1.3● SciPy 1.11● Elyra 3.15● PyMongo 4.5● Pyodbc 4.0● Codeflare-SDK 0.12● Sklearn-onnx 1.15● Psycopg 3.1● MySQL Connector/Python 8.0

Image name	Image version	Preinstalled packages
	2023.1 (Deprecated)	<ul style="list-style-type: none">● CUDA 11.8● Python 3.9● PyTorch 1.13● JupyterLab 3.5● Notebook 6.5● TensorBoard 2.11● Boto3 1.26● Kafka-Python 2.0● Kfp-tekton 1.5● Matplotlib 3.6● Numpy 1.24● Pandas 1.5● Scikit-learn 1.2● SciPy 1.10● Elyra 3.15

Image name	Image version	Preinstalled packages
Standard Data Science	2024.1 (Recommended)	<ul style="list-style-type: none">● Python 3.9● JupyterLab 3.6● Notebook 6.5● Boto3 1.34● Kafka-Python 2.0● Kfp 2.7● Matplotlib 3.8● Pandas 2.2● Numpy 1.26● Scikit-learn 1.4● SciPy 1.12● ODH-Elyra 3.16● PyMongo 4.6● Pyodbc 5.1● Codeflare-SDK 0.16● Sklearn-onnx 1.16● Psycopg 3.1● MySQL Connector/Python 8.3

Image name	Image version	Preinstalled packages
	2023.2	<ul style="list-style-type: none">● Python 3.9● JupyterLab 3.6● Notebook 6.5● Boto3 1.28● Kafka-Python 2.0● Kfp-tekton 1.5● Matplotlib 3.6● Pandas 1.5● Numpy 1.24● Scikit-learn 1.3● SciPy 1.11● Elyra 3.15● PyMongo 4.5● Pyodbc 4.0● Codeflare-SDK 0.12● Sklearn-onnx 1.15● Psycopg 3.1● MySQL Connector/Python 8.0

Image name	Image version	Preinstalled packages
	2023.1 (Deprecated)	<ul style="list-style-type: none">● Python 3.9● JupyterLab 3.5● Notebook 6.5● Boto3 1.26● Kafka-Python 2.0● Kfp-tekton 1.5● Matplotlib 3.6● Numpy 1.24● Pandas 1.5● Scikit-learn 1.2● SciPy 1.10● Elyra 3.15

Image name	Image version	Preinstalled packages
TensorFlow	2024.1 (Recommended)	<ul style="list-style-type: none">● CUDA 12.1● Python 3.9● JupyterLab 3.6● Notebook 6.5● TensorFlow 2.15● TensorBoard 2.15● Boto3 1.34● Kafka-Python 2.0● Kfp 2.5● Matplotlib 3.8● Numpy 1.26● Pandas 2.2● Scikit-learn 1.4● SciPy 1.12● ODH-Elyra 3.16● PyMongo 4.6● Pyodbc 5.1● Codeflare-SDK 0.16● Sklearn-onnx 1.16● Psycopg 3.1● MySQL Connector/Python 8.3

Image name	Image version	Preinstalled packages
	2023.2	<ul style="list-style-type: none">● CUDA 11.8● Python 3.9● JupyterLab 3.6● Notebook 6.5● TensorFlow 2.13● TensorBoard 2.13● Boto3 1.28● Kafka-Python 2.0● Kfp-tekton 1.5● Matplotlib 3.6● Numpy 1.24● Pandas 1.5● Scikit-learn 1.3● SciPy 1.11● Elyra 3.15● PyMongo 4.5● Pyodbc 4.0● Codeflare-SDK 0.12● Sklearn-onnx 1.15● Psycopg 3.1● MySQL Connector/Python 8.0

Image name	Image version	Preinstalled packages
	2023.1 (Deprecated)	<ul style="list-style-type: none">● CUDA 11.8● Python 3.9● JupyterLab 3.5● Notebook 6.5● TensorFlow 2.11● TensorBoard 2.11● Boto3 1.26● Kafka-Python 2.0● Kfp-tekton 1.5● Matplotlib 3.6● Numpy 1.24● Pandas 1.5● Scikit-learn 1.2● SciPy 1.10● Elyra 3.15

Image name	Image version	Preinstalled packages
TrustyAI	2024.1 (Recommended)	<ul style="list-style-type: none">● Python 3.9● JupyterLab 3.6● Notebook 6.5● TrustyAI 0.5● Boto3 1.34● Kafka-Python 2.0● Kfp 2.7● Matplotlib 3.6● Numpy 1.24● Pandas 1.5● Scikit-learn 1.4● SciPy 1.12● ODH-Elyra 3.16● PyMongo 4.6● Pyodbc 5.1● Codeflare-SDK 0.16● Sklearn-onnx 1.16● Psycopg 3.1● MySQL Connector/Python 8.3

Image name	Image version	Preinstalled packages
	2023.2	<ul style="list-style-type: none">● Python 3.9● JupyterLab 3.6● Notebook 6.5● TrustyAI 0.3● Boto3 1.28● Kafka-Python 2.0● Kfp-tekton 1.5● Matplotlib 3.6● Numpy 1.24● Pandas 1.5● Scikit-learn 1.3● SciPy 1.11● Elyra 3.15● PyMongo 4.5● Pyodbc 4.0● Codeflare-SDK 0.12● Sklearn-onnx 1.15● Psycopg 3.1● MySQL Connector/Python 8.0

Image name	Image version	Preinstalled packages
	2023.1 (Deprecated)	<ul style="list-style-type: none"> ● Python 3.9 ● JupyterLab 3.5 ● Notebook 6.5 ● TrustyAI 0.3 ● Boto3 1.26 ● Kafka-Python 2.0 ● Kfp-tekton 1.5 ● Matplotlib 3.6 ● Numpy 1.24 ● Pandas 1.5 ● Scikit-learn 1.2 ● SciPy 1.10 ● Elyra 3.15
HabanaAI	2024.1 (Recommended)	<ul style="list-style-type: none"> ● Python 3.8 ● Habana 1.13 ● JupyterLab 3.6 ● Boto3 1.34 ● Kafka-Python 2.0 ● Kfp 2.7 ● Matplotlib 3.7 ● Numpy 1.23 ● Pandas 2.0 ● Scikit-learn 1.3 ● Scipy 1.10 ● TensorFlow 2.13 ● PyTorch 2.1 ● ODH-Elyra v3.16

Image name	Image version	Preinstalled packages
	2023.2	<ul style="list-style-type: none"> ● Python 3.8 ● Habana 1.10 ● JupyterLab 3.5 ● TensorFlow 2.12 ● Boto3 1.26 ● Kafka-Python 2.0 ● Kfp-tekton 1.5 ● Matplotlib 3.6 ● Numpy 1.23 ● Pandas 1.5 ● Scikit-learn 1.2 ● SciPy 1.10 ● PyTorch 2.0 ● Elyra 3.15
code-server (Technology Preview)	2024.1 (Recommended)	<ul style="list-style-type: none"> ● Python 3.9 ● Boto3 1.29 ● Kafka-Python 2.0 ● Matplotlib 3.8 ● Numpy 1.26 ● Pandas 2.1 ● Plotly 5.18 ● Scikit-learn 1.3 ● Scipy 1.11 ● Sklearn-onnx 1.15 ● Ipykernel 6.26 ● (code-server plugin) Python 2024.2.1 ● (code-server plugin) Jupyter 2023.9.100

Image name	Image version	Preinstalled packages
	2023.2	<ul style="list-style-type: none"> ● Python 3.9 ● Boto3 1.29 ● Kafka-Python 2.0 ● Matplotlib 3.8 ● Numpy 1.26 ● Pandas 2.1 ● Plotly 5.18 ● Scikit-learn 1.3 ● Scipy 1.11 ● Sklearn-onnx 1.15 ● Ipykernel 6.26 ● (code-server plugin) Python 2023.14.0 ● (code-server plugin) Jupyter 2023.3.100
RStudio Server (Technology preview)	2024.1 (Recommended)	<ul style="list-style-type: none"> ● Python 3.9 ● R 4.3 <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 their licensing terms. Review their licensing terms before you use this sample workbench.</p>


Image name	Image version	Preinstalled packages
CUDA - RStudio Server (Technology preview)	2024.1 (Recommended)	<ul style="list-style-type: none">● Python 3.9● CUDA 12.1● R 4.3  <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 their licensing terms. Review their 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 their licensing terms before you use this sample workbench.</p>

Image name	Image version	Preinstalled packages
Deployment size		

specifies the compute resources available on your notebook server.

Container size controls the number of CPUs, the amount of memory, and the minimum and maximum request capacity of the container.

Accelerators specifies the accelerators available on your notebook server.

Number of accelerators specifies the number of accelerators to use.



IMPORTANT

Using accelerators is only supported with specific notebook images. For GPUs, only the PyTorch, TensorFlow, and CUDA notebook images are supported. For Habana Gaudi devices, only the HabanaAI notebook image is supported. In addition, you can only specify the number of accelerators required for your notebook server if accelerators are enabled on your cluster. To learn how to enable GPU support, see [Enabling GPU support in OpenShift AI](#).

Environment variables

Specifies the name and value of variables to be set on the notebook server. Setting environment variables during server startup means that you do not need to define them in the body of your notebooks, or with the Jupyter command line interface. Some recommended environment variables are shown in the table.

Table 5.2. Recommended environment variables

Environment variable option	Recommended variable names
AWS	<ul style="list-style-type: none"> • 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.

Additional resources

- [Launching Jupyter and starting a notebook server](#)

CHAPTER 6. TUTORIALS FOR DATA SCIENTISTS

To help you get started quickly, you can access learning resources for Red Hat OpenShift AI and its supported applications.

The [OpenShift AI tutorial: Fraud detection example](#) provides step-by-step guidance for using RHOAI to develop and train an example model in Jupyter notebooks, deploy the model, integrate the model into a fraud detection application, and refine the model by using automated pipelines.

Additional resources are available on the **Resources** tab of the Red Hat OpenShift AI user interface.

Table 6.1. Tutorials

Resource Name	Description
Accelerating scientific workloads in Python with Numba	Watch a video about how to make your Python code run faster.
Building interactive visualizations and dashboards in Python	Explore a variety of data across multiple notebooks and learn how to deploy full dashboards and applications.
Building machine learning models with scikit-learn	Learn how to build machine learning models with scikit-learn for supervised learning, unsupervised learning, and classification problems.
Building a binary classification model	Train a model to predict if a customer is likely to subscribe to a bank promotion.
Choosing Python tools for data visualization	Use the PyViz.org website to help you decide on the best open source Python data visualization tools for you.
Exploring Anaconda for data science	Learn about Anaconda, a freemium open source distribution of the Python and R programming languages.
Getting started with Pachyderm concepts	Learn Pachyderm's main concepts by creating pipelines that perform edge detection on a few images.
GPU Computing in Python with Numba	Learn how to create GPU accelerated functions using Numba.
Run a Python notebook to generate results in IBM Watson OpenScale	Run a Python notebook to create, train, and deploy a machine learning model.
Running an AutoAI experiment to build a model	Watch a video about building a binary classification model for a marketing campaign.

Resource Name	Description
Training a regression model in Pachyderm	Learn how to create a sample housing data repository using a Pachyderm cluster to run experiments, analyze data, and set up regression.
Using Dask for parallel data analysis	Analyze medium-sized datasets in parallel locally using Dask, a parallel computing library that scales the existing Python ecosystem.
Using Jupyter notebooks in Watson Studio	Watch a video about working with Jupyter notebooks in Watson Studio.
Using Pandas for data analysis in Python	Learn how to use pandas, a data analysis library for the Python programming language.

Table 6.2. Quick start guides

Resource Name	Description
Creating a Jupyter notebook	Create a Jupyter notebook in JupyterLab.
Creating an Anaconda-enabled Jupyter notebook	Create an Anaconda-enabled Jupyter notebook and access Anaconda packages that are curated for security and compatibility.
Deploying a model with Watson Studio	Import a notebook in Watson Studio and use AutoAI to build and deploy a model.
Deploying a sample Python application using Flask and OpenShift	Deploy your data science model out of a Jupyter notebook and into a Flask application to use as a development sandbox.
Importing Pachyderm Beginner Tutorial Notebook	Load Pachyderm's beginner tutorial notebook and learn about Pachyderm's main concepts such as data repositories, pipelines, and using the pachctl CLI from your cells.
Querying data with Starburst Galaxy	Learn to query data by using Starburst Galaxy from a Jupyter notebook.
Securing a deployed model using Red Hat OpenShift API Management	Protect a model service API using Red Hat OpenShift API Management.
Using the Intel® oneAPI AI Analytics Toolkit (AI Kit) Notebook	Run a data science notebook sample with the Intel® oneAPI AI Analytics Toolkit.

Resource Name	Description
Using the OpenVINO toolkit	Quantize an ONNX computer vision model using the OpenVINO model optimizer and use the result for inference from a notebook.

Table 6.3. How to guides

Resource Name	Description
How to choose between notebook runtime environment options	Explore available options for configuring your notebook runtime environment.
How to clean, shape, and visualize data	Learn how to clean and shape tabular data using IBM Watson Studio data refinery.
How to create a connection to access data	Learn how to create connections to various data sources across the platform.
How to create a deployment space	Learn how to create a deployment space for machine learning.
How to create a notebook in Watson Studio	Learn how to create a basic Jupyter notebook in Watson Studio.
How to create a project in Watson Studio	Learn how to create an analytics project in Watson Studio.
How to create a project that integrates with Git	Learn how to add assets from a Git repository into a project.
How to install Python packages on your notebook server	Learn how to install additional Python packages on your notebook server.
How to load data into a Jupyter notebook	Learn how to integrate data sources into a Jupyter notebook by loading data.
How to serve a model using OpenVINO Model Server	Learn how to deploy optimized models with the OpenVINO Model Server using OpenVINO custom resources.
How to set up Watson OpenScale	Learn how to track and measure outcomes from models with OpenScale.
How to update notebook server settings	Learn how to update the settings or the notebook image on your notebook server.

Resource Name	Description
How to use data from Amazon S3 buckets	Learn how to connect to data in S3 Storage using environment variables.
How to view installed packages on your notebook server	Learn how to see which packages are installed on your running notebook server.

6.1. ACCESSING TUTORIALS

You can access learning resources for Red Hat OpenShift AI and supported applications.

Prerequisites

- Ensure that you have logged in to Red Hat OpenShift AI.
- You have logged in to the OpenShift web console.

Procedure

1. On the Red Hat OpenShift AI home page, click **Resources**.
The **Resources** page opens.
2. Click **Access tutorial** on the relevant tile.

Verification

- You can view and access the learning resources for Red Hat OpenShift AI and supported applications.

Additional resources

- [Common questions](#).

CHAPTER 7. CONFIGURING YOUR IDE

You can configure some notebook workbenches to get the most out of your data science work.

7.1. CONFIGURING YOUR CODE-SERVER WORKBENCH

You can use extensions to streamline your workflow, add new languages, themes, debuggers, and connect to additional services.

For more information on code-server, see [code-server in GitHub](#).



IMPORTANT

The code-server notebook image is currently available in Red Hat OpenShift AI as a Technology Preview feature. Technology Preview features are not supported with Red Hat production service level agreements (SLAs) and might not be functionally complete. Red Hat does not recommend using them 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](#).

7.1.1. Installing extensions with code-server

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 code-server 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 containing the code-server 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**.
5. After the workbench has started, click **Open** to open the workbench notebook.



6. In the Activity Bar, click the Extensions icon. ()

7. Search for the name of the extension you want to install.
8. Click **Install** to add the extension to your code-server environment.
The extension you installed appears in the **Browser - Installed** list on the **Extensions** panel.

7.1.1.1. Extensions

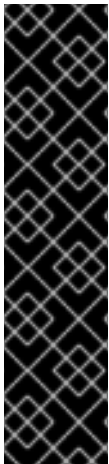
See [Open VSX Registry](#) for available third-party extensions that you can consider installing.

7.2. BUILDING THE RSTUDIO SERVER NOTEBOOK IMAGES

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 their licensing terms. Review their licensing terms before you use this sample workbench.

The **CUDA - RStudio Server** notebook image contains NVIDIA CUDA technology. CUDA licensing information is available at <https://docs.nvidia.com/cuda/>. Review their licensing terms before you use this sample workbench.



IMPORTANT

The **RStudio Server** and **CUDA - RStudio Server** notebook images are currently available in Red Hat OpenShift AI as Technology Preview features.

Technology Preview features are not supported with Red Hat production service level agreements (SLAs) and might not be functionally complete. Red Hat does not recommend using them 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](#).

To use the **RStudio Server** and **CUDA - RStudio Server** notebook images, you must first build them by creating a secret and triggering the BuildConfig, and then enable them in the OpenShift AI UI by editing the **rstudio-rhel9** and **cuda-rstudio-rhel9** image streams.

Prerequisites

- Before starting the RStudio Server build process, you have at least 1 CPU and 2Gi memory available for **rstudio-server-rhel9**, and 1.5 CPUs and 8Gi memory available for **cuda-rstudio-server-rhel9** on your cluster.
- You are logged in to your OpenShift cluster.
- You have the **cluster-admin** role in OpenShift, to the namespace **rhoai-applications**, or with cluster-wide role binding.
- You have an active Red Hat Enterprise Linux (RHEL) subscription.

Procedure

1. Create a secret with Subscription Manager credentials. These are usually your Red Hat Customer Portal username and password.

Note: The secret must be named **rhel-subscription-secret**, and its **USERNAME** and **PASSWORD** keys must be in capital letters.

```
oc create secret generic rhel-subscription-secret --from-literal=USERNAME=<username> --from-literal=PASSWORD=<password> -n redhat-ods-applications
```

2. Start the build:

- a. To start the lightweight RStudio Server build:

```
oc start-build rstudio-server-rhel9 -n redhat-ods-applications --follow
```

- b. To start the CUDA-enabled RStudio Server build, trigger the **cuda-rhel9** BuildConfig:

```
oc start-build cuda-rhel9 -n redhat-ods-applications --follow
```

The `cuda-rhel9` build is a prerequisite for `cuda-rstudio-rhel9`. The `cuda-rstudio-rhel9` build starts automatically.

3. Confirm that the build process has completed successfully using the following command. Successful builds appear as **Complete**.

```
oc get builds -n redhat-ods-applications
```

4. After the builds complete successfully, use the following commands to make the notebook images available in the OpenShift AI UI.

- a. To enable the RStudio Server notebook image:

```
oc label -n redhat-ods-applications imagestream rstudio-rhel9 opendatahub.io/notebook-image='true'
```

- b. To enable the CUDA - RStudio Server notebook image:

```
oc label -n redhat-ods-applications imagestream cuda-rstudio-rhel9 opendatahub.io/notebook-image='true'
```

Verification

- You can see **RStudio Server** and **CUDA - RStudio Server** images on the **Applications** → **Enabled** menu in the Red Hat OpenShift AI dashboard.
- You can see **R Studio Server** or **CUDA - RStudio Server** in the **Data Science Projects** → **Workbenches** → **Create workbench** → **Notebook image** → **Image selection** dropdown list.

CHAPTER 8. ENABLING SERVICES CONNECTED TO OPENSIFT AI

You must enable SaaS-based services, such as Anaconda Professional Edition, before using them with Red Hat OpenShift AI. On-cluster services are enabled automatically.

Typically, you can install services, or enable services connected to OpenShift AI using one of the following methods:

- Enabling the service from the **Explore** page on the OpenShift AI dashboard, as documented in the following procedure.
- Installing the Operator for the service from OperatorHub. OperatorHub is a web console for cluster administrators to discover and select Operators to install on their cluster. It is deployed by default in OpenShift ([Installing from OperatorHub using the web console](#)).



NOTE

Deployments containing Operators installed from OperatorHub may not be fully supported by Red Hat.

- Installing the Operator for the service from Red Hat Marketplace ([Install Operators](#)).
- Installing the service as an Add-on to your OpenShift Dedicated ([Adding Operators to an OpenShift Dedicated cluster](#)) or ROSA cluster ([Adding Operators to a ROSA cluster](#)).

For some services (such as Jupyter), the service endpoint is available on the tile for the service on the **Enabled** page of OpenShift AI. Certain services cannot be accessed directly from their tiles, for example, OpenVINO and Anaconda provide notebook images for use in Jupyter and do not provide an endpoint link from their tile. Additionally, it may be useful to store these endpoint URLs as environment variables for easy reference in a notebook environment.

Some independent software vendor (ISV) applications must be installed in specific namespaces. In these cases, the tile for the application in the OpenShift AI dashboard specifies the required namespace.

To help you get started quickly, you can access the service's learning resources and documentation on the **Resources** page, or by clicking the relevant link on the tile for the service on the **Enabled** page.

Prerequisites

- You have logged in to OpenShift AI.
- Your administrator has installed or configured the service on your OpenShift cluster.

Procedure

1. On the OpenShift AI home page, click **Explore**.
The **Explore** page opens.
2. Click the tile of the service that you want to enable.
3. Click **Enable** on the drawer for the service.
4. If prompted, enter the service's key and click **Connect**.

5. Click **Enable** to confirm that you are enabling the service.

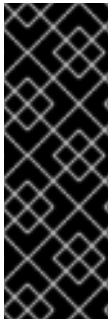
Verification

- The service that you enabled appears on the **Enabled** page.
- The service endpoint is displayed on the tile for the service on the **Enabled** page.

CHAPTER 9. DISABLING APPLICATIONS CONNECTED TO OPENSIFT AI

You can disable applications and components so that they do not appear on the OpenShift AI dashboard when you no longer want to use them, for example, when data scientists no longer use an application or when the application license expires.

Disabling unused applications allows your data scientists to manually remove these application tiles from their OpenShift AI dashboard so that they can focus on the applications that they are most likely to use. See [Removing disabled applications from OpenShift AI](#) for more information about manually removing application tiles.



IMPORTANT

Do not follow this procedure when disabling the following applications:

- Anaconda Professional Edition. You cannot manually disable Anaconda Professional Edition. It is automatically disabled only when its license expires.
- Red Hat OpenShift API Management. You can only uninstall Red Hat OpenShift API Management from OpenShift Cluster Manager.

Prerequisites

- You have logged in to the OpenShift web console.
- You are part of the **cluster-admins** or **dedicated-admins** user group in your OpenShift cluster. The **dedicated-admins** user group applies only to OpenShift Dedicated.
- You have installed or configured the service on your OpenShift cluster.
- The application or component that you want to disable is enabled and appears on the **Enabled** page.

Procedure

1. In the OpenShift web console, switch to the **Administrator** perspective.
2. Switch to the **redhat-ods-applications** project.
3. Click **Operators** → **Installed Operators**.
4. Click on the Operator that you want to uninstall. You can enter a keyword into the **Filter by name** field to help you find the Operator faster.
5. Delete any Operator resources or instances by using the tabs in the Operator interface. During installation, some Operators require the administrator to create resources or start process instances using tabs in the Operator interface. These must be deleted before the Operator can uninstall correctly.
6. On the **Operator Details** page, click the **Actions** drop-down menu and select **Uninstall Operator**.
An **Uninstall Operator?** dialog box is displayed.

7. Select **Uninstall** to uninstall the Operator, Operator deployments, and pods. After this is complete, the Operator stops running and no longer receives updates.



IMPORTANT

Removing an Operator does not remove any custom resource definitions or managed resources for the Operator. Custom resource definitions and managed resources still exist and must be cleaned up manually. Any applications deployed by your Operator and any configured off-cluster resources continue to run and must be cleaned up manually.

Verification

- The Operator is uninstalled from its target clusters.
- The Operator no longer appears on the **Installed Operators** page.
- The disabled application is no longer available for your data scientists to use, and is marked as **Disabled** on the **Enabled** page of the OpenShift AI dashboard. This action may take a few minutes to occur following the removal of the Operator.

9.1. REMOVING DISABLED APPLICATIONS FROM OPENSIFT AI

After your administrator has disabled your unused applications, you can manually remove them from the Red Hat OpenShift AI dashboard. Disabling and removing unused applications allows you to focus on the applications that you are most likely to use.

Prerequisites

- Ensure that you have logged in to Red Hat OpenShift AI.
- You have logged in to the OpenShift web console.
- Your administrator has previously disabled the application that you want to remove.

Procedure

1. In the OpenShift AI interface, click **Enabled**.
The **Enabled** page opens. Disabled applications are denoted with **Disabled** on the tile for the application.
2. Click **Disabled** on the tile for the application that you want to remove.
3. Click the link to remove the application tile.

Verification

- The tile for the disabled application no longer appears on the **Enabled** page.

CHAPTER 10. SUPPORT REQUIREMENTS AND LIMITATIONS

Review this section to understand the requirements for Red Hat support and any limitations to Red Hat support of Red Hat OpenShift AI.

10.1. SUPPORTED BROWSERS

Red Hat OpenShift AI supports the latest version of the following browsers:


- Google Chrome
- Mozilla Firefox
- Safari

10.2. SUPPORTED SERVICES

Red Hat OpenShift AI supports the following services:

Table 10.1. Supported services

Service Name	Description
Anaconda Professional	Anaconda Professional is a popular open source package distribution and management experience that is optimized for commercial use.
IBM Watson Studio	IBM Watson Studio is a platform for embedding AI and machine learning into your business and creating custom models with your own data.
Intel® oneAPI AI Analytics Toolkit Container	The AI Kit is a set of AI software tools to accelerate end-to-end data science and analytics pipelines on Intel® architectures.
Jupyter	<p>Jupyter is a multi-user version of the notebook designed for companies, classrooms, and research labs.</p> <div style="display: flex; align-items: flex-start;"> <div style="width: 20px; height: 100px; background: repeating-linear-gradient(45deg, transparent, transparent 2px, black 2px, black 4px); margin-right: 10px;"></div> <div> <p>IMPORTANT</p> <p>While every effort is made to make Red Hat OpenShift AI resilient to OpenShift node failure, upgrades, and similarly disruptive operations, individual users' notebook environments can be interrupted during these events. If an OpenShift node restarts or becomes unavailable, any user notebook environment on that node is restarted on a different node. When this occurs, any ongoing process executing in the user's notebook environment is interrupted, and the user needs to re-execute it when their environment becomes available again.</p> <p>Due to this limitation, Red Hat recommends that processes for which interruption is unacceptable are not executed in the Jupyter notebook server environment on OpenShift AI.</p> </div> </div>

Service Name	Description
Red Hat OpenShift API Management	OpenShift API Management is a service that accelerates time-to-value and reduces the cost of delivering API-first, microservices-based applications.
OpenVINO	OpenVINO is an open source toolkit to help optimize deep learning performance and deploy using an inference engine onto Intel hardware.
Pachyderm	<p>Use Pachyderm's data versioning, pipeline and lineage capabilities to automate the machine learning life cycle and optimize machine learning operations.</p> <div style="display: flex; align-items: flex-start;">  <div> <p>NOTE</p> <p>The pachd pod that Pachyderm creates intentionally does not provide terminal access from the OpenShift web console. If you try to access the terminal view for the pachd pod, you see an error. This behavior is expected.</p> </div> </div>
Starburst Galaxy	Starburst Galaxy is a fully-managed service for running high-performance queries across your various data sources.

10.3. SUPPORTED PACKAGES

The latest supported notebook server images in Red Hat OpenShift AI are installed with Python by default. See the table in [Options for notebook server environments](#) for a complete list of packages and versions included in these images.

You can install packages that are compatible with the supported version of Python on any notebook server that has the binaries required by that package. If the required binaries are not included on the notebook server image you want to use, contact Red Hat Support to request that the binary be considered for inclusion.

You can install packages on a temporary basis by using the **pip install** command. You can also provide a list of packages to the **pip install** command using a **requirements.txt** file. See [Installing Python packages on your notebook server](#) for more information.

You must re-install these packages each time you start your notebook server.

You can remove packages by using the **pip uninstall** command.

Additional resources

- [Installing Python packages on your notebook server](#)
- [Options for notebook server environments](#)

CHAPTER 11. COMMON QUESTIONS

In addition to documentation, Red Hat provides a rich set of learning resources for OpenShift AI and supported applications.

On the **Resources** page of the OpenShift AI dashboard, you can use the category links to filter the resources for various stages of your data science workflow. For example, click the **Model serving** category to display resources that describe various methods of deploying models. Click **All items** to show the resources for all categories.

For the selected category, you can apply additional options to filter the available resources. For example, you can filter by type, such as how-to articles, quick starts, tutorials; these resources provide the answers to common questions.

CHAPTER 12. TROUBLESHOOTING COMMON PROBLEMS IN JUPYTER FOR ADMINISTRATORS

If your users are experiencing errors in Red Hat OpenShift AI relating to Jupyter, their notebooks, or their notebook server, read this section to understand what could be causing the problem, and how to resolve the problem.

If you cannot see the problem here or in the release notes, contact Red Hat Support.

12.1. A USER RECEIVES A 404: PAGE NOT FOUND ERROR WHEN LOGGING IN TO JUPYTER

Problem

If you have configured specialized user groups for OpenShift AI, the user name might not be added to the default user group for OpenShift AI.

Diagnosis

Check whether the user is part of the default user group.

1. Find the names of groups allowed access to Jupyter.
 - a. Log in to the OpenShift web console.
 - b. Click **User Management** → **Groups**.
 - c. Click the name of your user group, for example, **rhoai-users**.
The **Group details** page for that group appears.
2. Click the **Details** tab for the group and confirm that the **Users** section for the relevant group contains the users who have permission to access Jupyter.

Resolution

- If the user is not added to any of the groups with permission access to Jupyter, follow [Adding users](#) to add them.
- If the user is already added to a group with permission to access Jupyter, contact Red Hat Support.

12.2. A USER'S NOTEBOOK SERVER DOES NOT START

Problem

The OpenShift cluster that hosts the user's notebook server might not have access to enough resources, or the Jupyter pod may have failed.

Diagnosis

1. Log in to the OpenShift web console.
2. Delete and restart the notebook server pod for this user.
 - a. Click **Workloads** → **Pods** and set the **Project** to **rhods-notebooks**.

- b. Search for the notebook server pod that belongs to this user, for example, **jupyter-nb-`<username>-*`**.

If the notebook server pod exists, an intermittent failure may have occurred in the notebook server pod.

If the notebook server pod for the user does not exist, continue with diagnosis.

3. Check the resources currently available in the OpenShift cluster against the resources required by the selected notebook server image.
If worker nodes with sufficient CPU and RAM are available for scheduling in the cluster, continue with diagnosis.
4. Check the state of the Jupyter pod.

Resolution

- If there was an intermittent failure of the notebook server pod:
 - a. Delete the notebook server pod that belongs to the user.
 - b. Ask the user to start their notebook server again.
- If the notebook server does not have sufficient resources to run the selected notebook server image, either add more resources to the OpenShift cluster, or choose a smaller image size.
- If the Jupyter pod is in a **FAILED** state:
 - a. Retrieve the logs for the **jupyter-nb-*** pod and send them to Red Hat Support for further evaluation.
 - b. Delete the **jupyter-nb-*** pod.
- If none of the previous resolutions apply, contact Red Hat Support.

12.3. THE USER RECEIVES A DATABASE OR DISK IS FULL ERROR OR A NO SPACE LEFT ON DEVICE ERROR WHEN THEY RUN NOTEBOOK CELLS

Problem

The user might have run out of storage space on their notebook server.

Diagnosis

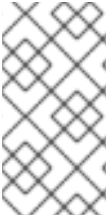
1. Log in to Jupyter and start the notebook server that belongs to the user having problems. If the notebook server does not start, follow these steps to check whether the user has run out of storage space:
 - a. Log in to the OpenShift web console.
 - b. Click **Workloads** → **Pods** and set the **Project** to **rhods-notebooks**.
 - c. Click the notebook server pod that belongs to this user, for example, **jupyter-nb-`<idp>-<username>-*`**.

- d. Click **Logs**. The user has exceeded their available capacity if you see lines similar to the following:

```
Unexpected error while saving file: XXXX database or disk is full
```

Resolution

- Increase the user's available storage by expanding their persistent volume: [Expanding persistent volumes](#)
- Work with the user to identify files that can be deleted from the **/opt/app-root/src** directory on their notebook server to free up their existing storage space.



NOTE

When you delete files using the JupyterLab file explorer, the files move to the hidden **/opt/app-root/src/.local/share/Trash/files** folder in the persistent storage for the notebook. To free up storage space for notebooks, you must permanently delete these files.

CHAPTER 13. TROUBLESHOOTING COMMON PROBLEMS IN JUPYTER FOR USERS

If you are seeing errors in Red Hat OpenShift AI related to Jupyter, your notebooks, or your notebook server, read this section to understand what could be causing the problem.

If you cannot see your problem here or in the release notes, contact Red Hat Support.

13.1. I SEE A 403: FORBIDDEN ERROR WHEN I LOG IN TO JUPYTER

Problem

If your administrator has configured specialized user groups for OpenShift AI, your user name might not be added to the default user group or the default administrator group for OpenShift AI.

Resolution

Contact your administrator so that they can add you to the correct group/s.

13.2. MY NOTEBOOK SERVER DOES NOT START

Problem

The OpenShift cluster that hosts your notebook server might not have access to enough resources, or the Jupyter pod may have failed.

Resolution

Check the logs in the **Events** section in OpenShift for error messages associated with the problem. For example:

```
Server requested
2021-10-28T13:31:29.830991Z [Warning] 0/7 nodes are available: 2 Insufficient memory,
2 node(s) had taint {node-role.kubernetes.io/infra: }, that the pod didn't tolerate, 3 node(s) had taint
{node-role.kubernetes.io/master: },
that the pod didn't tolerate.
```

Contact your administrator with details of any relevant error messages so that they can perform further checks.

13.3. I SEE A DATABASE OR DISK IS FULL ERROR OR ANO SPACE LEFT ON DEVICE ERROR WHEN I RUN MY NOTEBOOK CELLS

Problem

You might have run out of storage space on your notebook server.

Resolution

Contact your administrator so that they can perform further checks.

