Red Hat OpenShift Data Science Self-managed 1.29

Managing users and user resources

Learn to manage user permissions and environments in Red Hat OpenShift Data Science

Red Hat OpenShift Data Science Self-managed 1.29 Managing users and user resources

Learn to manage user permissions and environments in Red Hat OpenShift Data Science

Legal Notice

Abstract

Learn to manage user permissions and environments in Red Hat OpenShift Data Science.

Table of Contents

CHAPTER 1. PROVIDING FEEDBACK ON RED HAT DOCUMENTATION

Let Red Hat know how we can make our documentation better. You can provide feedback directly from a documentation page by following the steps below.

1. Make sure that you are logged in to the Customer Portal.

2. Make sure that you are looking at the **Multi-page HTML** format of this document.

3. Highlight the text that you want to provide feedback on. The **Add Feedback** prompt appears.

4. Click **Add Feedback**.

5. Enter your comments in the **Feedback** text box and click **Submit**.
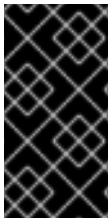
IMPORTANT

Some ad blockers might impede your ability to provide feedback on Red Hat documentation. If you are using a web browser that has an ad blocker enabled and you are unable to leave feedback, consider disabling your ad blocker. For more information about how to disable your ad blocker, see the documentation for your web browser.

Red Hat automatically creates a tracking issue each time you submit feedback. Open the link that is displayed after you click Submit and start watching the issue, or add more comments to give us more information about the problem.

Thank you for taking the time to provide your feedback.

CHAPTER 2. USAGE DATA COLLECTION

Red Hat OpenShift Data Science administrators can choose whether to allow Red Hat to collect data about OpenShift Data Science usage in their cluster. Collecting this data allows Red Hat to monitor and improve our software and support. For further details about the data Red Hat collects, see Usage data collection notice for OpenShift Data Science.

Usage data collection is enabled by default when you install OpenShift Data Science on your OpenShift Container Platform cluster except when clusters are installed in a disconnected environment.

See Disabling usage data collection for instructions on disabling the collection of this data in your cluster. If you have disabled data collection on your cluster, and you want to enable it again, see Enabling usage data collection for more information.

2.1. USAGE DATA COLLECTION NOTICE FOR OPENSHIFT DATA SCIENCE

In connection with your use of this Red Hat offering, Red Hat may collect usage data about your use of the software. This data allows Red Hat to monitor the software and to improve Red Hat offerings and support, including identifying, troubleshooting, and responding to issues that impact users.

What information does Red Hat collect?

Tools within the software monitor various metrics and this information is transmitted to Red Hat. Metrics include information such as:

- Information about applications enabled in the product dashboard.

- The deployment sizes used (that is, the CPU and memory resources allocated).

- Information about documentation resources accessed from the product dashboard.

- The name of the notebook images used (that is, Minimal Python, Standard Data Science, and other images.).

- A unique random identifier that generates during the initial user login to associate data to a particular username.

- Usage information about components, features, and extensions.

Third Party Service Providers

Red Hat uses certain third party service providers to collect the telemetry data.

Security

Red Hat employs technical and organizational measures designed to protect the usage data.

Personal Data

Red Hat does not intend to collect personal information. If Red Hat discovers that personal information has been inadvertently received, Red Hat will delete such personal information and treat such personal information in accordance with Red Hat's Privacy Statement. For more information about Red Hat's privacy practices, see Red Hat's Privacy Statement.

Enabling and Disabling Usage Data

You can disable or enable usage data by following the instructions in Disabling usage data collection or Enabling usage data collection.

2.2. ENABLING USAGE DATA COLLECTION

Red Hat OpenShift Data Science administrators can choose whether to allow Red Hat to collect data about OpenShift Data Science usage in their cluster. Usage data collection is enabled by default when you install OpenShift Data Science on your OpenShift Container Platform cluster except when clusters are installed in a disconnected environment. If you have disabled data collection previously, you can re-enable it by following these steps.

Prerequisites

- You have logged in to Red Hat OpenShift Data Science.

- You are part of the OpenShift Data Science administrator group in OpenShift Container Platform except when clusters are installed in a disconnected environment.

Procedure

1. From the OpenShift Data Science dashboard, click **Settings → Cluster settings**.

2. Locate the **Usage data collection** section.

3. Select the **Allow collection of usage data** checkbox.

4. Click **Save changes**.

Verification

- A notification is shown when settings are updated: **Settings changes saved.**

Additional resources

- [Usage data collection notice for OpenShift Data Science](#)

2.3. DISABLING USAGE DATA COLLECTION

Red Hat OpenShift Data Science administrators can choose whether to allow Red Hat to collect data about OpenShift Data Science usage in their cluster. Usage data collection is enabled by default when you install OpenShift Data Science on your OpenShift Container Platform cluster except when clusters are installed in a disconnected environment.

You can disable data collection by following these steps.

Prerequisites

- You have logged in to Red Hat OpenShift Data Science.

- You are part of the OpenShift Data Science administrator group in OpenShift Container Platform except when clusters are installed in a disconnected environment.

Procedure

1. From the OpenShift Data Science dashboard, click **Settings → Cluster settings**.

2. Locate the **Usage data collection** section.

3. Deselect the **Allow collection of usage data** checkbox.

4. Click **Save changes**.

Verification

- A notification is shown when settings are updated: **Settings changes saved.**

Additional resources

- Usage data collection notice for OpenShift Data Science

CHAPTER 3. OVERVIEW OF USER PERMISSIONS

By default, all OpenShift users have access to Red Hat OpenShift Data Science. In addition, users with the **cluster-admin** role, automatically have administrator access in OpenShift Data Science.

Alternatively, you can create specialized user groups to restrict access to OpenShift Data Science for users and administrators. Therefore, you must decide if you want to restrict access to your OpenShift Data Science deployment using specialized user groups, as opposed to allowing all OpenShift users access.

If you decide to restrict access, and you already have user groups defined in your configured identity provider, you can add these user groups to your OpenShift Data Science deployment. If you decide to use specialized user groups without adding these groups from an identity provider, you must create the groups in OpenShift Data Science and then add the appropriate users to them.

There are some operations relevant to OpenShift Data Science that require the **cluster-admin** role. Those operations include:

- Adding users to the OpenShift Data Science user and administrator groups, if you are using specialized groups.

- Removing users from the OpenShift Data Science user and administrator groups, if you are using specialized groups.

- Managing custom environment and storage configuration for users in OpenShift, such as Jupyter notebook resources, ConfigMaps, and persistent volume claims (PVCs).

IMPORTANT

Although users of OpenShift Data Science and its components are authenticated through OpenShift, session management is separate from authentication. This means that logging out of OpenShift or OpenShift Data Science does not affect a logged in Jupyter session running on those platforms. This means that when a user's permissions change, that user must log out of all current sessions in order for the changes to take effect.

CHAPTER 4. USER TYPES

Red Hat OpenShift Data Science has the following user types:

Table 4.1. User types

User Type	Permissions
Data scientists	Data scientists can access and use individual components of Red Hat OpenShift Data Science, such as Jupyter.
IT operations administrators	In addition to the actions permitted to a data scientist, IT operations administrators can: • Configure Red Hat OpenShift Data Science settings. • Access and manage notebook servers.

Additional resources

- OpenShift Container Platform Authentication and authorization

CHAPTER 5. DEFINING OPENSHIFT DATA SCIENCE ADMIN AND USER GROUPS

By default, users with the **cluster-admin** role are OpenShift Data Science administrators, but all users authenticated in OpenShift can access OpenShift Data Science. A cluster admin is a superuser that can perform any action in any project in the OpenShift cluster. When bound to a user with a local binding, they have full control over quota and every action on every resource in the project. You can also define additional OpenShift Data Science admin and user groups using the dashboard.

Prerequisites

- You have logged in to Red Hat OpenShift Data Science as described in Logging in to OpenShift Data Science.

- You have the **cluster-admin** role in OpenShift Container Platform.

- The groups that you want to define as admin and user groups exist in OpenShift Container Platform.

Procedure

1. From the OpenShift Data Science dashboard, click **Settings → User management**.

2. Define your OpenShift Data Science admin groups: Under **Data science administrator groups**, click the text box and select an OpenShift group. Repeat this process to define multiple admin groups.

3. Define your OpenShift Data Science user groups: Under **Data science user groups**, click the text box and select an OpenShift group. Repeat this process to define multiple user groups.

> **IMPORTANT**
>
> The **system:authenticated** setting allows all users authenticated in OpenShift to access OpenShift Data Science.

4. Click **Save changes**.

Verification

- Admin users can successfully log in to OpenShift Data Science and perform administrative functions.

- Non-admin users can successfully log in to OpenShift Data Science, and can access and use individual components, such as Jupyter.

CHAPTER 6. ADDING USERS FOR OPENSHIFT DATA SCIENCE

By default, all OpenShift users have access to Red Hat OpenShift Data Science. If you are using these default permission settings, no further action is required. However, if you plan to restrict access to your OpenShift Data Science instance by defining specialized user groups, you must grant users permission to access Red Hat OpenShift Data Science by adding user accounts to the Red Hat OpenShift Data Science user group, administrator group, or both. You can either use the default group name, or specify a group name that already exists in your identity provider.

The **user group** provides the user with access to developer functions in the Red Hat OpenShift Data Science dashboard, and associated services, such as Jupyter.

The **administrator group** provides the user with access to developer and administrator functions in the Red Hat OpenShift Data Science dashboard and associated services, such as Jupyter.

If you have restricted access using specialized user groups, users that are not in the OpenShift Data Science user group or administrator group can still view the dashboard, but are unable to use associated services, such as Jupyter. They are also unable to access the **Cluster settings** page.

To use the default group names, see *Adding users to specialized OpenShift Data Science user groups* . This method is easy to set up, but you must manually configure user lists in the OpenShift Container Platform web console.

6.1. ADDING USERS TO SPECIALIZED OPENSHIFT DATA SCIENCE USER GROUPS

All OpenShift users have access to Red Hat OpenShift Data Science by default. Additionally, users with the **cluster-admin** role automatically have administrator access to OpenShift Data Science. To further restrict access to OpenShift Data Science, you can continue to create specialized OpenShift Data Science administrator and user groups.

Follow the steps in this section to add users to your specialized OpenShift Data Science administrator and user groups. This method is easy to set up, but you must manage the user lists manually in the OpenShift Container Platform web console.

Prerequisites

- You have configured a supported identity provider for OpenShift Container Platform.

- You are assigned the **cluster-admin** role in OpenShift Container Platform.

- You have defined an OpenShift Data Science administrator group and user group.

Procedure

1. In the OpenShift Container Platform web console, click **User Management → Groups**.

2. Click the name of the group you want to add users to.

   - For administrative users, click the administrator group, for example, **rhods-admins**.

   - For normal users, click the user group, for example, **rhods-users**.
     The **Group details** page for that group appears.

3. Click **Actions → Add Users**.

The **Add Users** dialog appears.

4. In the the **Users** field, enter the relevant user name to add to the group.

5. Click **Save**.

Verification

- Click the **Details** tab for each group and confirm that the **Users** section contains the user names that you added.

6.2. ADDITIONAL RESOURCES

- Adding users to specialized OpenShift Data Science user groups

- Accessing the OpenShift Data Science dashboard

CHAPTER 7. VIEWING OPENSHIFT DATA SCIENCE USERS

By default, all OpenShift users have access to Red Hat OpenShift Data Science. In addition, users with the **cluster-admin** role automatically have administrator access in OpenShift Data Science. However, you can create specialized user groups to restrict access to OpenShift Data Science for users and administrators. Follow these steps if you have defined specialized OpenShift Data Science user groups, so that you can view the users that belong to these groups.

Prerequisites

- The Red Hat OpenShift Data Science user group, administrator group, or both exist.

- You have the **cluster-admin** role in OpenShift Container Platform.

- You have configured a supported identity provider for OpenShift Container Platform.

Procedure

1. In the OpenShift Container Platform web console, click **User Management → Groups**.

2. Click the name of the group containing the users that you want to view.

   - For administrative users, click the name of your administrator group. for example, **rhods-admins**.

   - For normal users, click the name of your user group, for example, **rhods-users**.

   The **Group details** page for the group appears.

Verification

- In the **Users** section for the relevant group, you can view the users who have permission to access Red Hat OpenShift Data Science.

CHAPTER 8. DELETING USERS AND USER RESOURCES

Users assigned the **cluster-admin** role in OpenShift can revoke user access to Jupyter and delete user resources from Red Hat OpenShift Data Science.

> **IMPORTANT**
>
> To completely remove a user from OpenShift Data Science, you must remove them from the allowed group in your OpenShift identity provider.

8.1. BACKING UP STORAGE DATA

Red Hat recommends that you back up the data on your persistent volume claims (PVCs) regularly. Backing up your data is particularly important before deleting a user and before uninstalling OpenShift Data Science, as all PVCs are deleted when OpenShift Data Science is uninstalled.

See the documentation for your cluster platform for more information about backing up your PVCs.

Additional resources

- [Understanding persistent storage](#)

8.2. REVOKING USER ACCESS TO JUPYTER

You can revoke a user's access to Jupyter to prevent them from running notebook servers and consuming resources in your cluster through Jupyter, while still allowing them access to OpenShift Data Science and other services that use OpenShift's identity provider for authentication.

> **IMPORTANT**
>
> Follow these steps only if you have restricted access to OpenShift Data Science using specialized user groups. To completely remove a user from OpenShift Data Science, you must remove them from the allowed group in your OpenShift identity provider.

Prerequisites

- You have stopped any notebook servers owned by the user you want to delete.

- You are assinged the **cluster-admin** role in OpenShift Container Platform.

- If you are using specialized OpenShift Data Science user groups, the user is part of the OpenShift Data Science user group, administrator group, or both.

Procedure

1. In the OpenShift Container Platform web console, click **User Management → Groups**.

2. Click the name of the group that you want to remove the user from.

   - For administrative users, click the name of your administrator group, for example, **rhods-admins**.

   - For normal users, click the name of your user group, for example, **rhods-users**.

The **Group details** page for the group appears.

3. In the **Users** section on the **Details** tab, locate the user that you want to remove.

4. Click the action menu (⋮) beside the user that you want to remove and click **Remove user**.

Verification

- Check the **Users** section on the **Details** tab and confirm that the user that you removed is not visible.

- In the **rhods-notebooks** project, check under **Workload → Pods** and ensure that there is no notebook server pod for this user. If you can see a pod named **jupyter-nb-<username>-*** for the user that you have removed, delete that pod to ensure that the deleted user is not consuming resources on the cluster.

- In the data science dashboard, check the list of data science projects. Delete any projects that belong to the user.

8.3. CLEANING UP AFTER DELETING USERS

After removing a user's access to Red Hat OpenShift Data Science or Jupyter, you must also delete their associated configuration files from OpenShift Container Platform. It is recommended that you back up the user's data before removing their configuration files.

Prerequisites

- (Optional) If you want to completely remove the user's access to OpenShift Data Science, you have removed their credentials from your identity provider.

- You have revoked the user's access to Jupyter.

- You have backed up the user's storage data.

- If you are using specialized OpenShift Data Science groups, you are part of the administrator group (for example, **rhods-admins**). If you are not using specialized groups, you are part of the OpenShift Dedicated administrator group. See Adding administrative users for OpenShift Container Platform for more information.

- You have logged in to the OpenShift Container Platform web console.

- You have logged in to OpenShift Data Science.

Procedure

1. Delete the user's persistent volume claim (PVC).

   a. Click **Storage → PersistentVolumeClaims**.

   b. If it is not already selected, select the **rhods-notebooks** project from the project list.

   c. Locate the **jupyter-nb-<username>** PVC.
      Replace **<username>** with the relevant user name.

   d. Click the action menu (⋮) and select **Delete PersistentVolumeClaim** from the list.
      The **Delete PersistentVolumeClaim** dialog appears.

e. Inspect the dialog and confirm that you are deleting the correct PVC.

f. Click **Delete**.

2. Delete the user's ConfigMap.

   a. Click **Workloads → ConfigMaps**.

   b. If it is not already selected, select the **rhods-notebooks** project from the project list.

   c. Locate the **jupyterhub-singleuser-profile-<username>** ConfigMap.
      Replace **<username>** with the relevant user name.

   d. Click the action menu (⋮) and select **Delete ConfigMap** from the list.
      The **Delete ConfigMap** dialog appears.

   e. Inspect the dialog and confirm that you are deleting the correct ConfigMap.

   f. Click **Delete**.

Verification

- The user cannot access Jupyter any more, and sees an "Access permission needed" message if they try.

- The user's single-user profile, persistent volume claim (PVC), and ConfigMap are not visible in OpenShift Container Platform.

CHAPTER 9. ALLOCATING ADDITIONAL RESOURCES TO OPENSHIFT DATA SCIENCE USERS

As a cluster administrator, you can allocate additional resources to a cluster to support compute-intensive data science work. This includes increasing the number of nodes in the cluster and changing the cluster's allocated machine pool.

For more information about allocating additional resources to an OpenShift Container Platform cluster, see Manually scaling a machine set .

CHAPTER 10. ENABLING GPU SUPPORT IN OPENSHIFT DATA SCIENCE

Optionally, to ensure that your data scientists can use compute-heavy workloads in their models, you can enable graphics processing units (GPUs) in OpenShift Data Science.

To make GPUs available in OpenShift Data Science, after you install OpenShift Data Science, you must install the Node Feature Discovery and GPU Operators.

> **IMPORTANT**
>
> Follow the instructions in this chapter only if you want to enable GPU support in an unrestricted self-managed environment. To enable GPU support in a disconnected self-managed environment, see Enabling GPU support in OpenShift Data Science for more information. To enable GPU support in a Red Hat managed cloud service, see Enabling GPU support in OpenShift Data Science for more information.

For more information on GPU enablement on OpenShift, see GPU Operator on OpenShift.

Additional resources

- GPU Operator on OpenShift.

- Cluster NFD Operator

CHAPTER 11. CONFIGURING THE DEFAULT PVC SIZE FOR YOUR CLUSTER

To configure how resources are claimed within your OpenShift Data Science cluster, you can change the default size of the cluster's persistent volume claim (PVC) ensuring that the storage requested matches your common storage workflow. PVCs are requests for resources in your cluster and also act as claim checks to the resource.

Prerequisites

- You have logged in to Red Hat OpenShift Data Science.

Changing this setting restarts the Jupyter pod making Jupyter unavailable for up to 30 seconds. As a workaround, it is recommended that you perform this action outside of your organization's typical working day.

Procedure

1. From the OpenShift Data Science dashboard, click **Settings → Cluster settings**.

2. Under **PVC size**, enter a new size in gibibytes. The minimum size is 1 GiB, and the maximum size is 16384 GiB.

3. Click **Save changes**.

Verification

- New PVCs are created with the default storage size that you configured.

Additional resources

- Understanding persistent storage

CHAPTER 12. RESTORING THE DEFAULT PVC SIZE FOR YOUR CLUSTER

To change the size of resources utilized within your OpenShift Data Science cluster, you can restore the default size of your cluster's persistent volume claim (PVC).

Prerequisites

- You have logged in to Red Hat OpenShift Data Science.

- You are part of the OpenShift Data Science administrator group in OpenShift Container Platform.

Procedure

1. From the OpenShift Data Science dashboard, click **Settings → Cluster settings**.

2. Click **Restore Default** to restore the default PVC size of 20GiB.

3. Click **Save changes**.

Verification

- New PVCs are created with the default storage size of 20 GiB.

Additional resources

- [Understanding persistent storage](#)

CHAPTER 13. MANAGING NOTEBOOK SERVERS

13.1. ACCESSING NOTEBOOK SERVERS OWNED BY OTHER USERS

Administrators can access notebook servers that are owned by other users to correct configuration errors or help a data scientist troubleshoot problems with their environment.

Prerequisites

- You are part of the OpenShift Dedicated administrator group. See *Adding administrative users for OpenShift Dedicated* for more information.

- You have launched the Jupyter application. See Launching Jupyter and starting a notebook server.

- The notebook server that you want to access is running.

Procedure

1. On the page that opens when you launch Jupyter, click the **Administration** tab.

2. On the **Administration** page, perform the following actions:

   a. In the **Users** section, locate the user that the notebook server belongs to.

   b. Click **View server** beside the relevant user.

   c. On the **Notebook server control panel** page, click **Access notebook server**.

Verification

- The user's notebook server opens in JupyterLab.

13.2. STOPPING IDLE NOTEBOOKS

You can reduce resource usage in your OpenShift Data Science deployment by stopping notebook servers that have been idle (without logged in users) for a period of time. This is useful when resource demand in the cluster is high. By default, idle notebooks are not stopped after a specific time limit.

> **NOTE**
>
> If you have configured your cluster settings to disconnect all users from a cluster after a specified time limit, then this setting takes precedence over the idle notebook time limit. Users are logged out of the cluster when their session duration reaches the cluster-wide time limit.

Prerequisites

- You have logged in to Red Hat OpenShift Data Science.

- You are part of the OpenShift Data Science administrator group in OpenShift Container Platform.

Procedure

1. From the OpenShift Data Science dashboard, click **Settings → Cluster settings**.

2. Under **Stop idle notebooks**, select **Stop idle notebooks after**.

3. Enter a time limit, in **hours** and **minutes**, for when idle notebooks are stopped.

4. Click **Save changes**.

Verification

- The **notebook-controller-culler-config** ConfigMap, located in the **redhat-ods-applications** project on the **Workloads → ConfigMaps** page, contains the following culling configuration settings:

  - **ENABLE_CULLING**: Specifies if the culling feature is enabled or disabled (this is **false** by default).

  - **IDLENESS_CHECK_PERIOD**: The polling frequency to check for a notebook's last known activity (in minutes).

  - **CULL_IDLE_TIME**: The maximum allotted time to scale an inactive notebook to zero (in minutes).

13.3. CONFIGURING A CUSTOM NOTEBOOK IMAGE

In addition to notebook images provided and supported by Red Hat and independent software vendors (ISVs), you can configure custom notebook images that cater to your project's specific requirements.

Red Hat supports you in adding custom notebook images to your deployment of OpenShift Data Science and ensuring that they are available for selection when creating a notebook server. However, Red Hat does not support the contents of your custom notebook image. That is, if your custom notebook image is available for selection during notebook server creation, but does not create a usable notebook server, Red Hat does not provide support to fix your custom notebook image.

Prerequisites

- You have logged in to Red Hat OpenShift Data Science.

- You are assigned the **cluster-admin** role in OpenShift Container Platform.

- Your custom notebook image exists in an image registry and is accessible.

Procedure

1. From the OpenShift Data Science dashboard, click **Settings → Notebook images**.
   The **Notebook image settings** page appears. Previously imported notebook images are displayed. To enable or disable a previously imported notebook image, on the row containing the relevant notebook image, click the toggle in the **Enabled** column.

2. Click **Import new image**. Alternatively, if no previously imported images were found, click **Import image**.
   The **Import Notebook images** dialog appears.

3. In the **Repository** field, enter the URL of the repository containing the notebook image.

4. In the **Name** field, enter an appropriate name for the notebook image.

5. In the **Description** field, enter an appropriate description for the notebook image.

6. Optional: Add software to the notebook image. After the import has completed, the software is added to the notebook image's meta-data and displayed on the Jupyter server creation page.

   a. Click the **Software** tab.

   b. Click the **Add software** button.

   c. Click **Edit** (✏).

   d. Enter the **Software** name.

   e. Enter the software **Version**.

   f. Click **Confirm** (✔) to confirm your entry.

   g. To add additional software, click **Add software**, complete the relevant fields, and confirm your entry.

7. Optional: Add packages to the notebook images. After the import has completed, the packages are added to the notebook image's meta-data and displayed on the Jupyter server creation page.

   a. Click the **Packages** tab.

   b. Click the **Add package** button.

   c. Click **Edit** (✏).

   d. Enter the **Package** name.

   e. Enter the package **Version**.

   f. Click **Confirm** (✔) to confirm your entry.

   g. To add an additional package, click **Add package**, complete the relevant fields, and confirm your entry.

8. Click **Import**.

Verification

- The notebook image that you imported is displayed in the table on the **Notebook image settings** page.

- Your custom notebook image is available for selection on the **Start a notebook server** page in Jupyter.

Additional resources

- Managing image streams

- Understanding build configurations