Red Hat OpenShift AI Cloud Service 1

Installing and uninstalling OpenShift AI Cloud Service

Install and uninstall OpenShift AI as an add-on to an OpenShift cluster

Last Updated: 2024-03-21

Red Hat OpenShift AI Cloud Service 1 Installing and uninstalling OpenShift AI Cloud Service

Install and uninstall OpenShift AI as an add-on to an OpenShift cluster

Legal Notice

Abstract

Install and uninstall OpenShift AI as an add-on to a Red Hat managed environment, such as Red Hat OpenShift Dedicated or Red Hat OpenShift Service on Amazon Web Services (ROSA).

Table of Contents

CHAPTER 1. ARCHITECTURE OF OPENSHIFT AI

Red Hat OpenShift AI is a fully Red Hat managed cloud service that is available as an Add-on to Red Hat OpenShift Dedicated and to Red Hat OpenShift Service on Amazon Web Services (ROSA).

OpenShift AI integrates the following components and services:

- At the service layer:

  OpenShift AI dashboard

  A customer-facing dashboard that shows available and installed applications for the OpenShift AI environment as well as learning resources such as tutorials, quick start examples, and documentation. You can also access administrative functionality from the dashboard, such as user management, cluster settings, accelerator profiles, and notebook image settings. In addition, data scientists can create their own projects from the dashboard. This enables them to organize their data science work into a single project.

  Model serving

  Data scientists can deploy trained machine-learning models to serve intelligent applications in production. After deployment, applications can send requests to the model using its deployed API endpoint.

  Data science pipelines

  Data scientists can build portable machine learning (ML) workflows with data science pipelines, using Docker containers. This enables your data scientists to automate workflows as they develop their data science models.

  Jupyter (Red Hat managed)

  A Red Hat managed application that allows data scientists to configure their own notebook server environment and develop machine learning models in JupyterLab.

Distributed workloads

Data scientists can use multiple nodes in parallel to train machine-learning models or process data more quickly. This approach significantly reduces the task completion time, and enables the use of larger datasets and more complex models.

IMPORTANT

The distributed workloads feature is currently available in Red Hat OpenShift AI as a Technology Preview feature only. 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 .

- At the management layer:

  The Red Hat OpenShift AI Operator

  A meta-operator that deploys and maintains all components and sub-operators that are part of OpenShift AI.

  Monitoring services

  Alertmanager, OpenShift Telemetry, and Prometheus work together to gather metrics from

OpenShift AI and organize and display those metrics in useful ways for monitoring and billing purposes. Alerts from Alertmanager are sent to PagerDuty, responsible for notifying Red Hat of any issues with your managed cloud service.

When you install the Red Hat OpenShift AI Add-on in the Cluster Manager, the following new projects are created:

- The **redhat-ods-operator** project contains the Red Hat OpenShift AI Operator.
- The **redhat-ods-applications** project installs the dashboard and other required components of OpenShift AI.
- The **redhat-ods-monitoring** project contains services for monitoring and billing.
- The **rhods-notebooks** project is where notebook environments are deployed by default.

You or your data scientists must create additional projects for the applications that will use your machine learning models.

Do not install independent software vendor (ISV) applications in namespaces associated with OpenShift AI add-ons unless you are specifically directed to do so on the application tile on the dashboard.

CHAPTER 2. INSTALLING AND DEPLOYING OPENSHIFT AI

Red Hat OpenShift AI is a platform for data scientists and developers of artificial intelligence (AI) applications. It provides a fully supported environment that lets you rapidly develop, train, test, and deploy machine learning models on-premises and/or in the public cloud.

OpenShift AI is provided as a managed cloud service add-on for Red Hat OpenShift or as self-managed software that you can install on-premise or in the public cloud on OpenShift.

For information about installing OpenShift AI as self-managed software on your OpenShift cluster in a connected or a disconnected environment, see Product Documentation for Red Hat OpenShift AI Self-Managed.

There are two deployment options for Red Hat OpenShift AI as a managed cloud service add-on:

- **OpenShift Dedicated with a Customer Cloud Subscription on Amazon Web Services or Google Cloud Platform**
  OpenShift Dedicated is a complete OpenShift Container Platform cluster provided as a cloud service, configured for high availability, and dedicated to a single customer. OpenShift Dedicated is professionally managed by Red Hat and hosted on Amazon Web Services (AWS) or Google Cloud Platform (GCP). The Customer Cloud Subscription (CCS) model allows Red Hat to deploy and manage clusters into a customer's AWS or GCP account. Contact your Red Hat account manager to get OpenShift Dedicated through a CCS.

- **Red Hat OpenShift Service on AWS (ROSA)**
  ROSA is a fully-managed, turnkey application platform that allows you to focus on delivering value to your customers by building and deploying applications. You subscribe to the service directly from your AWS account.

Installing OpenShift AI as a managed cloud service involves the following high-level tasks:

1. Confirm that your OpenShift cluster meets all requirements.
2. Configure an identity provider for your OpenShift cluster.
3. Add administrative users for your OpenShift cluster.
4. Subscribe to the Red Hat OpenShift AI Add-on.
   For OpenShift Dedicated with a CCS for AWS or GCP, get a subscription through Red Hat.

   For ROSA, get a subscription through the AWS Marketplace.
5. Install the Red Hat OpenShift AI Add-on.
6. Access the OpenShift AI dashboard.
7. Optionally, enable graphics processing units (GPUs) in OpenShift AI to ensure that your data scientists can use compute-heavy workloads in their models.

2.1. REQUIREMENTS FOR OPENSHIFT AI

You must meet the following requirements before you can install OpenShift AI on your Red Hat OpenShift Dedicated or Red Hat OpenShift Service on Amazon Web Services (ROSA) cluster.

- **A subscription for Red Hat OpenShift Dedicated** or **a subscription for ROSA**
  You can deploy Red Hat OpenShift Dedicated on your Amazon Web Services (AWS) or Google

Cloud Platform (GCP) account by using the Customer Cloud Subscription on AWS or Customer Cloud Subscription on GCP model. Note that while Red Hat provides an option to install OpenShift Dedicated on a Red Hat cloud account, if you want to install OpenShift AI then you must install OpenShift Dedicated on your own cloud account.

Contact your Red Hat account manager to purchase a new Red Hat OpenShift Dedicated subscription. If you do not yet have an account manager, complete the form at https://cloud.redhat.com/products/dedicated/contact/ to request one.

You can subscribe to Red Hat OpenShift Service on AWS (ROSA) directly from your AWS account or by contacting your Red Hat account manager.

- **A Red Hat customer account**
  Go to OpenShift Cluster Manager (http://console.redhat.com/openshift) and log in or register for a new account.

- **Cluster administrator access to your OpenShift cluster**
  Use an existing cluster or create a new cluster by following the steps in the relevant documentation:

  - Creating an OpenShift Dedicated cluster
  - Creating a ROSA cluster with STS

- **An OpenShift Dedicated or ROSA cluster configuration that meets the following configuration requirements.**
  At least 2 worker nodes with at least 8 CPUs and 32 GiB RAM available for OpenShift AI to use when you install the Add-on. If this requirement is not met, the installation process fails to start and an error is displayed.

  When you create a new cluster, select **m6a.2xlarge** for the computer node instance type to satisfy the requirements.

  For an existing ROSA cluster, you can get the compute node instance type by using this command:

  rosa list machinepools --cluster=cluster-name

  You cannot alter a cluster's compute node instance type, but you can add an additional machine pool or modify the default pool to meet the minimum requirements. However, the minimum resource requirements must be met by a single machine pool in the cluster.

  For more information, see the relevant documentation:

  - Creating a machine pool in OpenShift Dedicated
  - OpenShift AI Service Definition
  - Creating a machine pool in ROSA
  - Prepare your environment (ROSA)

- **For a ROSA cluster, select an access management strategy**
  For installing OpenShift AI on a ROSA cluster, decide whether you want to install on a ROSA cluster that uses AWS Security Token Service (STS) or one that uses AWS Identity and Access Management (IAM) credentials. See Install ROSA Classic clusters for advice on deploying a ROSA cluster with or without AWS STS.

- **Install the Red Hat OpenShift Pipelines Operator**
  OpenShift AI supports data science pipelines. A pipeline is a collection of Task resources that are arranged in a specific order of execution. By using Red Hat OpenShift AI pipelines, you can standardize and automate machine learning workflows to automate the build and deployment of your data science models. Before you can use pipelines with OpenShift AI, install the Red Hat OpenShift Pipelines Operator as described in Installing OpenShift Pipelines.

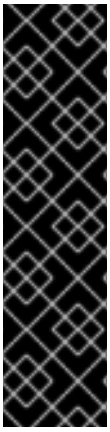
- **Install KServe dependencies**
  To support KServe components, you must install dependent Operators, including the Red Hat OpenShift Serverless and Red Hat OpenShift Service Mesh Operators. For more information, see Serving large models.

2.2. CONFIGURING AN IDENTITY PROVIDER FOR YOUR OPENSHIFT CLUSTER

Configure an identity provider for your OpenShift Dedicated or Red Hat OpenShift Service on Amazon Web Services (ROSA) cluster to manage users and groups.

Red Hat OpenShift AI supports the same authentication systems as Red Hat OpenShift Dedicated and ROSA. Check the appropriate documentation for your cluster for more information.

- Supported identity providers on OpenShift Dedicated
- Supported identity providers on ROSA



IMPORTANT

Adding more than one OpenShift Identity Provider can create problems when the same user name exists in multiple providers.

When **mappingMethod** is set to **claim** (the default mapping method for identity providers) and multiple providers have credentials associated with the same user name, the first provider used to log in to OpenShift is the one that works for that user, regardless of the order in which identity providers are configured.

For more information about mapping methods, see Identity provider parameters in OpenShift Dedicated or Identity provider parameters in ROSA.

Prerequisites

- Credentials for OpenShift Cluster Manager (https://console.redhat.com/openshift/).
- An existing OpenShift Dedicated or ROSA cluster.

Procedure

1. Log in to OpenShift Cluster Manager (https://console.redhat.com/openshift/).
2. Click **Clusters**. The **Clusters** page opens.
3. Click the name of the cluster to configure.
4. Click the **Access control** tab.
5. Click **Identity providers**.

6. Click **Add identity provider**.
    a. Select your provider from the **Identity Provider** list.
    b. Complete the remaining fields relevant to the identity provider that you selected. For more information, see Configuring identity providers in OpenShift Dedicated or Configuring identity providers in ROSA.
7. Click **Confirm**.

Verification

- The configured identity providers are visible on the **Access control** tab of the **Cluster details** page.

Additional resources

- Configuring identity providers in ROSA
- Configuring identity providers in OpenShift Dedicated

2.3. ADDING ADMINISTRATIVE USERS

Before you can install and configure OpenShift AI for your data scientist users, you must define administrative users. Only administrative users can install and configure OpenShift AI.

Prerequisites

- Credentials for OpenShift Cluster Manager (https://console.redhat.com/openshift/).
- An existing OpenShift Dedicated or Red Hat OpenShift Service on AWS (ROSA) cluster with an identity provider configured.

Procedure

1. Log in to OpenShift Cluster Manager (https://console.redhat.com/openshift/).
2. Click **Clusters**. The **Clusters** page opens.
3. Click the name of the cluster to configure.
4. Click the **Access control** tab.
5. Click **Cluster Roles and Access**.
6. Under **Cluster administrative users** click the **Add user** button.
   The **Add cluster user** popover appears.
7. Enter the user name in the **User ID** field.
8. Select an appropriate **Group** for the user.

IMPORTANT

If this user needs to use existing groups in an identity provider to control OpenShift AI access, select **cluster-admins**.

For more information about these user types, see Managing administration roles and users in the OpenShift Dedicated documentation or Default cluster roles in the ROSA documentation.

9. Click **Add user**.

Verification

- The user name and selected group are visible in the list of **Cluster administrative users**.

Additional resources

- OpenShift Dedicated Cluster administration
- ROSA Cluster administration

2.4. SUBSCRIBING TO THE RED HAT OPENSHIFT AI CLOUD SERVICE

You can subscribe to the Red Hat OpenShift AI managed cloud service in the following ways:

- Subscribe through Red Hat if you have a Red Hat OpenShift Dedicated cluster deployed with a Customer Cloud Subscription (CCS) on Amazon Web Services (AWS) or Google Cloud Platform (GCP).
- Subscribe through the AWS Marketplace if you have a Red Hat OpenShift Service on AWS (ROSA) cluster.

NOTE

You can also purchase Red Hat OpenShift AI as self-managed software. To purchase a new subscription, contact your Red Hat account manager. If you do not yet have an account manager, complete the form at https://www.redhat.com/en/contact to request one.

2.4.1. Subscribing to the OpenShift AI managed cloud service on AWS or GCP

For a Red Hat OpenShift Dedicated cluster that is deployed on AWS or GCP, contact your Red Hat account manager to purchase a new subscription. If you do not yet have an account manager, complete the form at https://cloud.redhat.com/products/dedicated/contact/ to request one.

Prerequisite

- You have worked with Red Hat Sales to enable a private offer of OpenShift AI, follow these steps to accept your offer and deploy the solution.

Procedure

1. Visit your Private Offer with the URL link provided by your Red Hat Sales representative.

2. Click **Accept Terms** to subscribe to the AMI Private Offer named **Openshift Data Science from AWS Marketplace**.

3. After accepting the offer terms, click **Continue to Configuration**.

2.4.2. Subscribing to the OpenShift AI managed cloud service on Red Hat OpenShift Service on AWS (ROSA)

For a ROSA cluster, you can subscribe to the OpenShift AI managed cloud service through the Amazon Web Services (AWS) Marketplace.

Prerequisites

- Access to a ROSA cluster, including permissions to view and install add-ons.
- An AWS account with permission to view and subscribe to offerings in the AWS marketplace.

Procedure

1. In the AWS Console, navigate to the AWS Marketplace. For example:
   a. Click the help icon and then select Getting Started Resource Center.
   b. Select AWS Marketplace > Browse AWS Marketplace.
2. In the top **Search** field, type: **Red Hat OpenShift AI**.
3. Select one of the two options depending on the geographical location of the billing address for your AWS account (note that this location might differ from the geographical location of the cluster):
   - Europe, the Middle East, and Africa (EMEA region)
   - North America and regions outside EMEA
4. Click **Continue to Subscribe**.
5. Click **Continue to Configuration** and then select the appropriate fulfillment options. Note that some selectors might have only one option.
6. Click **Continue to Launch**.
7. Link your AWS account with your Red Hat account to complete your registration:
   a. In the AWS Marketplace console, navigate to the **Manage Subscriptions** page.
   b. On the **Red Hat OpenShift AI** tile, click **Set up product**.
   c. On the top banner, click **Set up account**
   This link takes you to the Red Hat Hybrid console.
   d. If you are not already logged in, log in.
   e. Review and then accept the terms and agreements.
   f. Click **Connect accounts**.

Verification

The Data Science product page opens.

2.5. INSTALLING OPENSHIFT AI ON YOUR OPENSHIFT CLUSTER

You can use Red Hat OpenShift Cluster Manager to install Red Hat OpenShift AI as an Add-on to your Red Hat OpenShift cluster.

Prerequisites

- A subscription to the Red Hat OpenShift AI Add-on, as described in Subscribing to the OpenShift AI managed cloud service on AWS or GCP.
- If you purchased the Red Hat OpenShift AI Add-on for ROSA by using the AWS Marketplace, you have associated your AWS account with your Red Hat account as described in Subscribing to the OpenShift AI managed cloud service on Red Hat OpenShift Service on AWS (ROSA).
- Credentials for Red Hat OpenShift Cluster Manager (https://console.redhat.com/openshift/).
- Administrator access to the OpenShift cluster.
- To support KServe components, you installed the dependent Operators, including the Red Hat OpenShift Serverless and Red Hat OpenShift Service Mesh Operators. For more information, see Serving large models.

NOTE

For information about the lifecycle associated with Red Hat OpenShift AI, see Red Hat OpenShift AI Life Cycle.

Procedure

1. Log in to OpenShift Cluster Manager (https://console.redhat.com/openshift/).
2. Click **Clusters**.
   The **Clusters** page opens.
3. Click the name of the cluster you want to install OpenShift AI on.
   The **Details** page for the cluster opens.
4. Click the **Add-ons** tab and locate the **Red Hat OpenShift AI** tile.

NOTE

If there is a **Prerequisites not met** warning message, click the **Prerequisites** tab. Note down the error message. If the error message states that you require a new machine pool, or that more resources are required, take the appropriate action to resolve the problem. You might need to add more resources to your cluster, or increase the size of your default machine pool. To increase your cluster's resources, contact your infrastructure administrator. For more information about increasing the size of your machine pool, see Allocating additional resources to OpenShift AI users.

5. Select a **Subscription type**:

If you obtained your RHODS subscription through your Red Hat account manager, select **Standard** and then skip to Step 7.

If you obtained your RHODS subscription directly from the AWS Marketplace, select **Marketplace** and then continue to Step 6.

6. For a Marketplace subscription, select your AWS account number from the list.

NOTE

If your AWS account number is not in the list, you might need to link your Red Hat and AWS accounts, as described in Subscribing to the OpenShift AI managed cloud service on Red Hat OpenShift Service on AWS (ROSA).

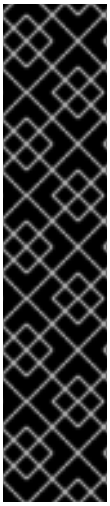
7. Click **Install**. The **Configure Red Hat OpenShift AI** pane appears.
8. In the **Notification email** field, enter any email addresses that you want to receive important alerts about the state of Red Hat OpenShift AI, such as outage alerts.
9. Click **Install**.

Verification

- In OpenShift Cluster Manager, on the **Add-ons** tab for the cluster, confirm that the OpenShift Data Science tile shows one of the following states:
  - **Installing** - installation is in progress; wait for this to change to **Installed**. This takes around 30 minutes.
  - **Installed** - installation is complete; verify that the **View in console** button is visible.
- In OpenShift Dedicated, click **Home → Projects** and confirm that the following project namespaces are visible and listed as **Active**:
  - **redhat-ods-applications**
  - **redhat-ods-monitoring**
  - **redhat-ods-operator**
  - **rhods-notebooks**

2.6. INSTALLING RED HAT OPENSHIFT AI COMPONENTS BY USING THE WEB CONSOLE

The following procedure shows how to use the OpenShift web console to install specific components of Red Hat OpenShift AI on your cluster.



IMPORTANT

When you install Red Hat OpenShift AI as an add-on to your OpenShift cluster, the install process automatically creates a default **DataScienceCluster** object. The following procedure describes how to configure the **DataScienceCluster** object to install Red Hat OpenShift AI components as part of a *new* installation.

If you upgraded from version 1 of OpenShift AI (previously OpenShift Data Science), the upgrade process also automatically creates a default **DataScienceCluster** object. If you upgraded from a previous minor version, the upgrade process uses the settings from the previous version's **DataScienceCluster** object. To inspect the **DataScienceCluster** object and change the installation status of Red Hat OpenShift AI components, see Updating the installation status of Red Hat OpenShift AI components by using the web console.

Prerequisites

- To support the KServe component, you installed dependent Operators, including the Red Hat OpenShift Serverless and Red Hat OpenShift Service Mesh Operators. For more information, see Serving large models.
- Red Hat OpenShift AI is installed as an add-on to your Red Hat OpenShift cluster.
- You have cluster administrator privileges for your OpenShift cluster.

Procedure

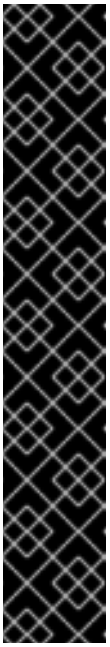
1. Log in to the OpenShift web console as a cluster administrator.
2. In the web console, click **Operators → Installed Operators** and then click the Red Hat OpenShift AI Operator.
3. Configure the **DataScienceCluster** object to install OpenShift AI components by performing the following actions:
   a. Click the **Data Science Cluster** tab.
   b. Click the **default-dsc** object.
   c. Select the **YAML** tab.
   An embedded YAML editor opens showing a default custom resource (CR) for the **DataScienceCluster** object.
   d. In the **spec.components** section of the CR, for each OpenShift AI component shown, set the value of the **managementState** field to either **Managed** or **Removed**. These values are defined as follows:

   Managed

   The Operator actively manages the component, installs it, and tries to keep it active. The Operator will upgrade the component only if it is safe to do so.

   Removed

   The Operator actively manages the component but does not install it. If the component is already installed, the Operator will try to remove it.



IMPORTANT

- To learn how to install the KServe component, which is used by the single model serving platform to serve large models, see Serving large models.
- The CodeFlare and KubeRay components are Technology Preview features only. 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 learn how to configure the distributed workloads feature that uses the CodeFlare and KubeRay components, see Configuring distributed workloads.

4. Click **Save**.

Verification

- Confirm that there is a running pod for each component:
  1. In the OpenShift web console, click **Workloads → Pods**.
  2. In the **Project** list at the top of the page, select **redhat-ods-applications**.
  3. In the applications namespace, confirm that there are running pods for each of the OpenShift AI components that you installed.
- Confirm the status of all installed components:
  1. In the OpenShift web console, click **Operators → Installed Operators**.
  2. Click the Red Hat OpenShift AI Operator.
  3. Click the **Data Science Cluster** tab and select the **DataScienceCluster** object called **default-dsc**.
  4. Select the **YAML** tab.
  5. In the **installedComponents** section, confirm that the components you installed have a status value of **true**.

NOTE

If a component shows with the **component-name: {}** format in the **spec.components** section of the CR, the component is not installed.

2.7. TROUBLESHOOTING COMMON INSTALLATION PROBLEMS

If you are experiencing difficulties installing the Red Hat OpenShift AI Add-on, 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.

2.7.1. The Red Hat OpenShift AI Operator cannot be retrieved from the image registry

Problem

When attempting to retrieve the Red Hat OpenShift AI Operator from the image registry, an **Failure to pull from quay** error message appears. The Red Hat OpenShift AI Operator might be unavailable for retrieval in the following circumstances:

- The image registry is unavailable.
- There is a problem with your network connection.
- Your cluster is not operational and is therefore unable to retrieve the image registry.

Diagnosis

Check the logs in the **Events** section in OpenShift Dedicated for further information about the **Failure to pull from quay** error message.

Resolution

- To resolve this issue, contact Red Hat support.

2.7.2. OpenShift AI cannot be installed due to insufficient cluster resources

Problem

When attempting to install OpenShift AI, an error message appears stating that installation prerequisites have not been met.

Diagnosis

1. Log in to OpenShift Cluster Manager (https://console.redhat.com/openshift/).
2. Click **Clusters**.
   The **Clusters** page opens.
3. Click the name of the cluster you want to install OpenShift AI on.
   The **Details** page for the cluster opens.
4. Click the **Add-ons** tab and locate the **Red Hat OpenShift AI** tile.
5. Click **Install**. The **Configure Red Hat OpenShift AI** pane appears.
6. If the installation fails, click the **Prerequisites** tab.
7. Note down the error message. If the error message states that you require a new machine pool, or that more resources are required, take the appropriate action to resolve the problem.

Resolution

- You might need to add more resources to your cluster, or increase the size of your machine pool. To increase your cluster's resources, contact your infrastructure administrator. For more

information about increasing the size of your machine pool, see Nodes and Allocating additional resources to OpenShift AI users.

2.7.3. The dedicated-admins Role-based access control (RBAC) policy cannot be created

Problem

The Role-based access control (RBAC) policy for the dedicated-admins group in the target project cannot be created. This issue occurs in unknown circumstances.

Diagnosis

1. In the OpenShift Dedicated web console, change into the **Administrator** perspective.
2. Click **Workloads → Pods**.
3. Set the **Project** to **All Projects** or **redhat-ods-operator**.
4. Click the **rhods-operator-<random string>** pod.
   The **Pod details** page appears.
5. Click **Logs**.
6. Select **rhods-deployer** from the drop-down list
7. Check the log for the **ERROR: Attempt to create the RBAC policy for dedicated admins group in $target_project failed.** error message.

Resolution

- Contact Red Hat support.

2.7.4. OpenShift AI does not install on unsupported infrastructure

Problem

Customer deploying on an environment not documented as being supported by the RHODS operator.

Diagnosis

1. In the OpenShift Dedicated web console, change into the **Administrator** perspective.
2. Click **Workloads → Pods**.
3. Set the **Project** to **All Projects** or **redhat-ods-operator**.
4. Click the **rhods-operator-<random string>** pod.
   The **Pod details** page appears.
5. Click **Logs**.
6. Select **rhods-deployer** from the drop-down list
7. Check the log for the **ERROR: Deploying on $infrastructure, which is not supported. Failing Installation** error message.

Resolution

Before proceeding with a new installation, ensure that you have a fully supported environment on which to install OpenShift AI. For more information, see Requirements for OpenShift AI.

2.7.5. The creation of the OpenShift AI Custom Resource (CR) fails

Problem

During the installation process, the OpenShift AI Custom Resource (CR) does not get created. This issue occurs in unknown circumstances.

Diagnosis

1. In the OpenShift Dedicated web console, change into the **Administrator** perspective.
2. Click **Workloads → Pods**.
3. Set the **Project** to **All Projects** or **redhat-ods-operator**.
4. Click the **rhods-operator-<random string>** pod.
   The **Pod details** page appears.
5. Click **Logs**.
6. Select **rhods-deployer** from the drop-down list
7. Check the log for the **ERROR: Attempt to create the ODH CR failed.** error message.

Resolution

Contact Red Hat support.

2.7.6. The creation of the OpenShift AI Notebooks Custom Resource (CR) fails

Problem

During the installation process, the OpenShift AI Notebooks Custom Resource (CR) does not get created. This issue occurs in unknown circumstances.

Diagnosis

1. In the OpenShift Dedicated web console, change into the **Administrator** perspective.
2. Click **Workloads → Pods**.
3. Set the **Project** to **All Projects** or **redhat-ods-operator**.
4. Click the **rhods-operator-<random string>** pod.
   The **Pod details** page appears.
5. Click **Logs**.
6. Select **rhods-deployer** from the drop-down list
7. Check the log for the **ERROR: Attempt to create the RHODS Notebooks CR failed.** error message.

Resolution

Contact Red Hat support.

2.7.7. The Dead Man's Snitch operator's secret does not get created

Problem

An issue with Managed Tenants SRE automation process causes the Dead Man's Snitch operator's secret to not get created.

Diagnosis

1. In the OpenShift Dedicated web console, change into the **Administrator** perspective.
2. Click **Workloads → Pods**.
3. Set the **Project** to **All Projects** or **redhat-ods-operator**.
4. Click the **rhods-operator-<random string>** pod.
   The **Pod details** page appears.
5. Click **Logs**.
6. Select **rhods-deployer** from the drop-down list
7. Check the log for the **ERROR: Dead Man Snitch secret does not exist.** error message.

Resolution

Contact Red Hat support.

2.7.8. The PagerDuty secret does not get created

Problem

An issue with Managed Tenants SRE automation process causes the PagerDuty's secret to not get created.

Diagnosis

1. In the OpenShift Dedicated web console, change into the **Administrator** perspective.
2. Click **Workloads → Pods**.
3. Set the **Project** to **All Projects** or **redhat-ods-operator**.
4. Click the **rhods-operator-<random string>** pod.
   The **Pod details** page appears.
5. Click **Logs**.
6. Select **rhods-deployer** from the drop-down list
7. Check the log for the **ERROR: Pagerduty secret does not exist** error message.

Resolution

Contact Red Hat support.

2.7.9. The SMTP secret does not exist

Problem

An issue with Managed Tenants SRE automation process causes the SMTP secret to not get created.

Diagnosis

1. In the OpenShift Dedicated web console, change into the **Administrator** perspective.
2. Click **Workloads → Pods**.
3. Set the **Project** to **All Projects** or **redhat-ods-operator**.
4. Click the **rhods-operator-<random string>** pod.
   The **Pod details** page appears.
5. Click **Logs**.
6. Select **rhods-deployer** from the drop-down list
7. Check the log for the **ERROR: SMTP secret does not exist** error message.

Resolution

Contact Red Hat support.

2.7.10. The ODH parameter secret does not get created

Problem

An issue with the OpenShift AI Add-on's flow could result in the ODH parameter secret to not get created.

Diagnosis

1. In the OpenShift Dedicated web console, change into the **Administrator** perspective.
2. Click **Workloads → Pods**.
3. Set the **Project** to **All Projects** or **redhat-ods-operator**.
4. Click the **rhods-operator-<random string>** pod.
   The **Pod details** page appears.
5. Click **Logs**.
6. Select **rhods-deployer** from the drop-down list
7. Check the log for the **ERROR: Addon managed odh parameter secret does not exist.** error message.

Resolution

Contact Red Hat support.

CHAPTER 3. WORKING WITH CERTIFICATES

Certificates are used by various components in OpenShift Dedicated to validate access to the cluster. For clusters that rely on self-signed certificates, you can add those self-signed certificates to a cluster-wide Certificate Authority (CA) bundle and use the CA bundle in Red Hat OpenShift AI. You can also use self-signed certificates in a custom CA bundle that is separate from the cluster-wide bundle. Administrators can add a CA bundle, remove a CA bundle from all namespaces, remove a CA bundle from individual namespaces, or manually manage certificate changes instead of the system.

3.1. UNDERSTANDING CERTIFICATES IN OPENSHIFT AI

For OpenShift Dedicated clusters that rely on self-signed certificates, you can add those self-signed certificates to a cluster-wide Certificate Authority (CA) bundle (**ca-bundle.crt**) and use the CA bundle in Red Hat OpenShift AI. You can also use self-signed certificates in a custom CA bundle (**odh-ca-bundle.crt**) that is separate from the cluster-wide bundle.

3.1.1. How CA bundles are injected

After installing OpenShift AI, the Red Hat OpenShift AI Operator automatically creates an empty **odh-trusted-ca-bundle** configuration file (ConfigMap), and the Cluster Network Operator (CNO) injects the cluster-wide CA bundle into the **odh-trusted-ca-bundle** configMap with the label "config.openshift.io/inject-trusted-cabundle". The components deployed in the affected namespaces are responsible for mounting this configMap as a volume in the deployment pods.

```
apiVersion: v1
kind: ConfigMap
metadata:
  labels:
    app.kubernetes.io/part-of: opendatahub-operator
    config.openshift.io/inject-trusted-cabundle: 'true'
  name: odh-trusted-ca-bundle
```

After the CNO operator injects the bundle, it updates the ConfigMap with the **ca-bundle.crt** file containing the certificates.

```
apiVersion: v1
kind: ConfigMap
metadata:
  labels:
    app.kubernetes.io/part-of: opendatahub-operator
    config.openshift.io/inject-trusted-cabundle: 'true'
  name: odh-trusted-ca-bundle
data:
  ca-bundle.crt: |
    <BUNDLE OF CLUSTER-WIDE CERTIFICATES>
```

3.1.2. How the ConfigMap is managed

By default, the Red Hat OpenShift AI Operator manages the **odh-trusted-ca-bundle** ConfigMap. If you want to manage or remove the **odh-trusted-ca-bundle** ConfigMap, or add a custom CA bundle (**odh-ca-bundle.crt**) separate from the cluster-wide CA bundle (**ca-bundle.crt**), you can use the **trustedCABundle** property in the Operator's DSC Initialization (DSCI) object.

```
spec:
  trustedCABundle:
    managementState: Managed
    customCABundle: ""
```

In the Operator's DSCI object, you can set the **spec.trustedCABundle.managementState** field to the following values:

- **Managed**: The Red Hat OpenShift AI Operator manages the **odh-trusted-ca-bundle** ConfigMap and adds it to all non-reserved existing and new namespaces (the ConfigMap is not added to any reserved or system namespaces, such as **default**, **openshift-*** or **kube-***). The ConfigMap is automatically updated to reflect any changes made to the **customCABundle** field. This is the default value after installing Red Hat OpenShift AI.

- **Removed**: The Red Hat OpenShift AI Operator removes the **odh-trusted-ca-bundle** ConfigMap (if present) and disables the creation of the ConfigMap in new namespaces. If you change this field from **Managed** to **Removed**, the **odh-trusted-ca-bundle** ConfigMap is also deleted from namespaces. This is the default value after upgrading Red Hat OpenShift AI from 2.7 or earlier versions to 1.

- **Unmanaged**: The Red Hat OpenShift AI Operator does not manage the **odh-trusted-ca-bundle** ConfigMap, allowing for an administrator to manage it instead. Changing the **managementState** from **Managed** to **Unmanaged** does not remove the **odh-trusted-ca-bundle** ConfigMap, but the ConfigMap is not updated if you make changes to the **customCABundle** field.

In the Operator's DSCI object, you can add a custom certificate to the **spec.trustedCABundle.customCABundle** field. This adds the **odh-ca-bundle.crt** file containing the certificates to the **odh-trusted-ca-bundle** ConfigMap, as shown in the following example:

```
apiVersion: v1
kind: ConfigMap
metadata:
  labels:
    app.kubernetes.io/part-of: opendatahub-operator
    config.openshift.io/inject-trusted-cabundle: 'true'
  name: odh-trusted-ca-bundle
data:
  ca-bundle.crt: |
    <BUNDLE OF CLUSTER-WIDE CERTIFICATES>
  odh-ca-bundle.crt: |
    <BUNDLE OF CUSTOM CERTIFICATES>
```

3.2. ADDING A CA BUNDLE

There are two ways to add a Certificate Authority (CA) bundle to OpenShift AI. You can use one or both of these methods:

- For OpenShift Dedicated clusters that rely on self-signed certificates, you can add those self-signed certificates to a cluster-wide Certificate Authority (CA) bundle (**ca-bundle.crt**) and use the CA bundle in Red Hat OpenShift AI. To use this method, log in to the OpenShift Dedicated as a cluster administrator and follow the steps as described in Configuring the cluster-wide proxy during installation.

- You can use self-signed certificates in a custom CA bundle (**odh-ca-bundle.crt**) that is separate from the cluster-wide bundle. To use this method, follow the steps in this section.

Prerequisites

- You have admin access to the **DSCInitialization** resources in the OpenShift Dedicated cluster.
- You installed the OpenShift command line interface (**oc**) as described in Get Started with the CLI.
- You are working in a new installation of Red Hat OpenShift AI. If you upgraded Red Hat OpenShift AI, see Adding a CA bundle after upgrading .

Procedure

1. Log in to the OpenShift Dedicated.
2. Click **Operators → Installed Operators** and then click the Red Hat OpenShift AI Operator.
3. Click the **DSC Initialization** tab.
4. Click the **default-dsci** object.
5. Click the **YAML** tab.
6. In the **spec** section, add the custom certificate to the **customCABundle** field for **trustedCABundle**, as shown in the following example:

```
spec:
  trustedCABundle:
    managementState: Managed
    customCABundle: |
      -----BEGIN CERTIFICATE-----
      examplebundle123
      -----END CERTIFICATE-----
```

7. Click **Save**.

Verification

- If you are using a cluster-wide CA bundle, run the following command to verify that all non-reserved namespaces contain the **odh-trusted-ca-bundle** ConfigMap:

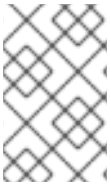
```
$ oc get configmaps --all-namespaces -l app.kubernetes.io/part-of=opendatahub-operator | grep odh-trusted-ca-bundle
```

- If you are using a custom CA bundle, run the following command to verify that a non-reserved namespace contains the **odh-trusted-ca-bundle** ConfigMap and that the ConfigMap contains your **customCABundle** value. In the following command, *example-namespace* is the non-reserved namespace and *examplebundle123* is the customCABundle value.

```
$ oc get configmap odh-trusted-ca-bundle -n example-namespace -o yaml | grep examplebundle123
```

3.3. REMOVING A CA BUNDLE

You can remove a Certificate Authority (CA) bundle from all non-reserved namespaces in OpenShift AI. This process changes the default configuration and disables the creation of the **odh-trusted-ca-bundle** configuration file (ConfigMap), as described in *Understanding certificates in OpenShift AI*.



NOTE

The **odh-trusted-ca-bundle** ConfigMaps are only deleted from namespaces when you set the **managementState** of **trustedCABundle** to **Removed**; deleting the DSC Initialization does not delete the ConfigMaps.

To remove a CA bundle from a single namespace only, see *Removing a CA bundle from a namespace* .

Prerequisites

- You have cluster administrator privileges for your OpenShift Dedicated cluster.
- You installed the OpenShift command line interface (**oc**) as described in Get Started with the CLI.

Procedure

1. In the OpenShift Dedicated web console, click **Operators → Installed Operators** and then click the Red Hat OpenShift AI Operator.
2. Click the **DSC Initialization** tab.
3. Click the **default-dsci** object.
4. Click the **YAML** tab.
5. In the **spec** section, change the value of the **managementState** field for **trustedCABundle** to **Removed**:

```
spec:
  trustedCABundle:
    managementState: Removed
```

6. Click **Save**.

Verification

- Run the following command to verify that the **odh-trusted-ca-bundle** ConfigMap has been removed from all namespaces:

```
$ oc get configmaps --all-namespaces | grep odh-trusted-ca-bundle
```

The command should not return any ConfigMaps.

3.4. REMOVING A CA BUNDLE FROM A NAMESPACE

You can remove a custom Certificate Authority (CA) bundle from individual namespaces in OpenShift AI. This process disables the creation of the **odh-trusted-ca-bundle** configuration file (ConfigMap) for the specified namespace only.

To remove a certificate bundle from all namespaces, see *Removing a CA bundle* .

Prerequisites

- You have cluster administrator privileges for your OpenShift Dedicated cluster.
- You installed the OpenShift command line interface (**oc**) as described in Get Started with the CLI.

Procedure

- Run the following command to remove a CA bundle from a namespace. In the following command, *example-namespace* is the non-reserved namespace.

```
$ oc annotate ns example-namespace security.opendatahub.io/inject-trusted-ca-bundle=false
```

Verification

- Run the following command to verify that the CA bundle has been removed from the namespace. In the following command, *example-namespace* is the non-reserved namespace.

```
$ oc get configmap odh-trusted-ca-bundle -n example-namespace
```

The command should return **configmaps "odh-trusted-ca-bundle" not found**.

3.5. MANAGING CERTIFICATES

After installing OpenShift AI, the Red Hat OpenShift AI Operator creates the **odh-trusted-ca-bundle** configuration file (ConfigMap) that contains the trusted CA bundle and adds it to all new and existing non-reserved namespaces in the cluster. By default, the Red Hat OpenShift AI Operator manages the **odh-trusted-ca-bundle** ConfigMap and automatically updates it if any changes are made to the CA bundle. You can choose to manage the **odh-trusted-ca-bundle** ConfigMap instead of allowing the Red Hat OpenShift AI Operator to manage it.

Prerequisites

- You have cluster administrator privileges for your OpenShift Dedicated cluster.

Procedure

1. In the OpenShift Dedicated web console, click **Operators → Installed Operators** and then click the **Red Hat OpenShift AI Operator**.
2. Click the **DSC Initialization** tab.
3. Click the **default-dsci** object.
4. Click the **YAML** tab.

5. In the **spec** section, change the value of the **managementState** field for **trustedCABundle** to **Unmanaged**, as shown:

```
spec:
  trustedCABundle:
    managementState: Unmanaged
```

6. Click **Save**.
   Note that changing the **managementState** from **Managed** to **Unmanaged** does not remove the **odh-trusted-ca-bundle** ConfigMap, but the ConfigMap is not updated if you make changes to the **customCABundle** field.

Verification

1. In the **spec** section, set or change the value of the **customCABundle** field for **trustedCABundle**, for example:

```
spec:
  trustedCABundle:
    managementState: Unmanaged
    customCABundle: example123
```

2. Click **Save**.
3. Click **Workloads → ConfigMaps**.
4. Select a project from the project list.
5. Click the **odh-trusted-ca-bundle** ConfigMap.
6. Click the **YAML** tab and verify that the value of the **customCABundle** field did not update.

3.6. USING SELF-SIGNED CERTIFICATES WITH OPENSHIFT AI COMPONENTS

Some OpenShift AI components have additional options or required configuration for self-signed certificates.

3.6.1. Using certificates with data science pipelines

If you want to use self-signed certificates, you have added them to a central Certificate Authority (CA) bundle as described in Working with certificates.

No additional configuration is necessary to use those certificates with data science pipelines.

3.6.1.1. Providing a CA bundle only for data science pipelines

Perform the following steps to provide a Certificate Authority (CA) bundle just for data science pipelines.

Procedure

1. Log in to OpenShift Dedicated.

2. From **Workloads → ConfigMaps**, create a ConfigMap with the required bundle in the same data science project or namespace as the target data science pipeline:

```
kind: ConfigMap
apiVersion: v1
metadata:
  name: custom-ca-bundle
data:
  ca-bundle.crt: |
  # contents of ca-bundle.crt
```

3. Add the following snippet to the **.spec.apiserver.caBundle** field of the underlying Data Science Pipelines Application (DSPA):

```
apiVersion: datasciencepipelinesapplications.opendatahub.io/v1alpha1
kind: DataSciencePipelinesApplication
metadata:
  name: data-science-pipelines-definition
spec:
  ...
  apiServer:
  ...
  cABundle:
    configMapName: custom-ca-bundle
    configMapKey: ca-bundle.crt
```

The pipeline server pod redeploys with the updated bundle and uses it in the newly created pipeline pods.

Verification

Perform the following steps to confirm that your CA bundle was successfully mounted.

1. Log in to the OpenShift Dedicated console.
2. Go to the OpenShift Dedicated project that corresponds to the data science project.
3. Click the **Pods** tab.
4. Click the pipeline server pod with the **ds-pipeline-pipelines-definition-<hash>** prefix.
5. Click **Terminal**.
6. Enter **cat /dsp-custom-certs/dsp-ca.crt**.
7. Verify that your CA bundle is present within this file.

You can also confirm that your CA bundle was successfully mounted by using the CLI:

1. In a terminal window, log in to the OpenShift cluster where OpenShift AI is deployed.

```
oc login
```

2. Set the **dspa** value:

```
dspa=pipelines-definition
```

3. Set the **dsProject** value, replacing **$YOUR_DS_PROJECT** with the name of your data science project:

```
dsProject=$YOUR_DS_PROJECT
```

4. Set the **pod** value:

```
pod=$(oc get pod -n ${dsProject} -l app=ds-pipeline-${dspa} --no-headers | awk '{print $1}')
```

5. Display the contents of the **/dsp-custom-certs/dsp-ca.crt** file:

```
oc -n ${dsProject} exec $pod -- cat /dsp-custom-certs/dsp-ca.crt
```

6. Verify that your CA bundle is present within this file.

3.6.2. Using certificates with workbenches

IMPORTANT

Self-signed certificates apply to workbenches that you create after configuring self-signed certificates centrally as described in Working with certificates. There is no change to workbenches that you created before configuring self-signed certificates.

3.6.2.1. Creating data science pipelines with Elyra and self-signed certificates

To create pipelines using a workbench that contains the Elyra extension and which uses self-signed certificates, see the Workbench workaround for executing a pipeline using Elyra in a disconnected environment knowledgebase article.

CHAPTER 4. ACCESSING THE DASHBOARD

After you have installed OpenShift AI and added users, you can access the URL for your OpenShift AI console and share the URL with the users to let them log in and work on their models.

Prerequisites

- You have installed OpenShift AI on your OpenShift cluster.
- You have added at least one user to the user group for OpenShift AI as described in Adding users.

Procedure

1. Log in to OpenShift web console.
2. Click the application launcher ().
3. Right-click on **Red Hat OpenShift AI** and copy the URL for your OpenShift AI instance.
4. Provide this instance URL to your data scientists to let them log in to OpenShift AI.

Verification

- Confirm that you and your users can log in to OpenShift AI by using the instance URL.

Additional resources

- Logging in to OpenShift AI
- Adding users

CHAPTER 5. ENABLING GPU SUPPORT IN OPENSHIFT AI

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

IMPORTANT

The NVIDIA GPU add-on is no longer supported. Instead, enable GPUs by installing the NVIDIA GPU Operator. If your deployment has a previously-installed NVIDIA GPU add-on, before you install the NVIDIA GPU Operator, use Red Hat OpenShift Cluster Manager to uninstall the NVIDIA GPU add-on from your cluster.

Prerequisites

- You have logged in to your OpenShift cluster.
- You have the **cluster-admin** role in your OpenShift cluster.

Procedure

1. To enable GPU support on an OpenShift cluster, follow the instructions here: NVIDIA GPU Operator on Red Hat OpenShift Container Platform in the NVIDIA documentation.
2. Delete the **migration-gpu-status** ConfigMap.
    a. In the OpenShift web console, switch to the **Administrator** perspective.
    b. Set the **Project** to **All Projects** or **redhat-ods-applications** to ensure you can see the appropriate ConfigMap.
    c. Search for the **migration-gpu-status** ConfigMap.
    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**.
3. Restart the dashboard replicaset.
    a. In the OpenShift web console, switch to the **Administrator** perspective.
    b. Click **Workloads → Deployments**.
    c. Set the **Project** to **All Projects** or **redhat-ods-applications** to ensure you can see the appropriate deployment.
    d. Search for the **rhods-dashboard** deployment.
    e. Click the action menu (⋮) and select **Restart Rollout** from the list.
    f. Wait until the **Status** column indicates that all pods in the rollout have fully restarted.

Verification

- The NVIDIA GPU Operator appears on the **Operators → Installed Operators** page in the OpenShift web console.
- The reset **migration-gpu-status** instance is present in the **Instances** tab on the **AcceleratorProfile** custom resource definition (CRD) details page.

After installing the NVIDIA GPU Operator, create an accelerator profile as described in Working with accelerator profiles.

CHAPTER 6. UNINSTALLING OPENSHIFT AI

Use Red Hat OpenShift Cluster Manager to uninstall Red Hat OpenShift AI from your OpenShift Dedicated cluster.

6.1. UNDERSTANDING THE UNINSTALLATION PROCESS

Installing Red Hat OpenShift AI created several custom resource instances on your OpenShift Dedicated cluster for various components of OpenShift AI. After installation, users likely created several additional resources while using OpenShift AI. Uninstalling OpenShift AI removes the resources that were created by the Operator, but retains the resources created by users to prevent inadvertently deleting information you might want.

What is deleted

Uninstalling OpenShift AI removes the following resources from your OpenShift Dedicated cluster:

- **DataScienceCluster** custom resource instance
- **DSCInitialization** custom resource instance
- **FeatureTracker** custom resource instances created during or after installation
- **ServiceMesh** custom resource instance created by the Operator during or after installation
- **KNativeServing** custom resource instance created by the Operator during or after installation
- **redhat-ods-applications**, **redhat-ods-monitoring**, and **rhods-notebooks** namespaces created by the Operator
- Workloads in the **rhods-notebooks** namespace
- **Subscription**, **ClusterServiceVersion**, and **InstallPlan** objects
- **KfDef** object (version 1 Operator only)

What might remain

Uninstalling OpenShift AI retains the following resources in your OpenShift Dedicated cluster:

- Data science projects created by users
- Custom resource instances created by users
- Custom resource definitions (CRDs) created by users or by the Operator

While these resources might still remain in your OpenShift Dedicated cluster, they are not functional. After uninstalling, Red Hat recommends that you review the data science projects and custom resources in your OpenShift Dedicated cluster and delete anything no longer in use to prevent potential issues, such as pipelines that cannot run, notebooks that cannot be undeployed, or models that cannot be undeployed.

Additional resources

Operator Lifecycle Manager (OLM) uninstall documentation

6.2. BACKING UP STORAGE DATA FROM AMAZON EBS

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 AI, as all PVCs are deleted when OpenShift AI is uninstalled.

Prerequisites

- You have credentials for Red Hat OpenShift Cluster Manager (https://console.redhat.com/openshift/).
- You have administrator access to the OpenShift Dedicated cluster.
- You have credentials for the Amazon Web Services (AWS) account that the OpenShift Dedicated cluster is deployed under.

Procedure

1. Determine the IDs of the persistent volumes (PVs) that you want to back up.
   a. In the OpenShift Dedicated web console, change into the **Administrator** perspective.
   b. Click **Home → Projects**.
   c. Click the **rhods-notebooks** project.
   The **Details** page for the project opens.
   d. Click the **PersistentVolumeClaims** in the **Inventory** section.
   The **PersistentVolumeClaims** page opens.
   e. Note the ID of the persistent volume (PV) that you want to back up.

NOTE

The persistent volumes (PV) that you make a note of are required to identify the correct EBS volume to back up in your AWS instance.

2. Locate the EBS volume containing the PVs that you want to back up.
   See Amazon Web Services documentation: Create Amazon EBS snapshots for more information.
   a. Log in to AWS (https://aws.amazon.com) and ensure that you are viewing the region that your OpenShift Dedicated cluster is deployed in.
   b. Click **Services**.
   c. Click **Compute → EC2**.
   d. Click **Elastic Block Storage → Volumes** in the side navigation.
   The **Volumes** page opens.
   e. In the search bar, enter the ID of the persistent volume (PV) that you made a note of earlier.
   The **Volumes** page reloads to display the search results.
   f. Click on the volume shown and verify that any **kubernetes.io/created-for/pvc/namespace** tags contain the value **rhods-notebooks**, and any **kubernetes.io/created-for/pvc/name** tags match the name of the persistent volume that the EC2 volume is being used for, for example, **jupyter-nb-user1-pvc**.

3. Back up the EBS volume that contains your persistent volume (PV).
    a. Right-click on the volume that you want to back up and select **Create Snapshot** from the list.
    The **Create Snapshot** page opens.
    b. Enter a **Description** for the volume.
    c. Click **Create Snapshot**.
    The snapshot of the volume is created.
    d. Click **Close**.

Verification

- The snapshot that you created is visible on the **Snapshots** page in AWS.

Additional resources

- Amazon Web Services documentation: Create Amazon EBS snapshots

6.3. BACKING UP STORAGE DATA FROM GOOGLE PERSISTENT DISK

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 AI, as all PVCs are deleted when OpenShift AI is uninstalled.

Prerequisites

- You have credentials for Red Hat OpenShift Cluster Manager (https://console.redhat.com/openshift/).
- You have administrator access to the OpenShift Dedicated cluster.
- You have credentials for the Google Cloud Platform (GCP) account that the OpenShift Dedicated cluster is deployed under.

Procedure

1. Determine the IDs of the persistent volumes (PVs) that you want to back up.
    a. In the OpenShift Dedicated web console, change into the **Administrator** perspective.
    b. Click **Home → Projects**.
    c. Click the **rhods-notebooks** project.
    The **Details** page for the project opens.
    d. Click the **PersistentVolumeClaims** in the **Inventory** section.
    The **PersistentVolumeClaims** page opens.
    e. Note the ID of the persistent volume (PV) that you want to back up.
    The persistent volume (PV) IDs are required to identify the correct persistent disk to back up in your GCP instance.

2. Locate the persistent disk containing the PVs that you want to back up.
    a. Log in to the Google Cloud console (https://console.cloud.google.com) and ensure that you are viewing the region that your OpenShift Dedicated cluster is deployed in.
    b. Click the navigation menu (≡) and then click **Compute Engine**.
    c. From the side navigation, under **Storage**, click **Disks**.
    The **Disks** page opens.
    d. In the **Filter** query box, enter the ID of the persistent volume (PV) that you made a note of earlier.
    The **Disks** page reloads to display the search results.
    e. Click on the disk shown and verify that any **kubernetes.io/created-for/pvc/namespace** tags contain the value **rhods-notebooks**, and any **kubernetes.io/created-for/pvc/name** tags match the name of the persistent volume that the persistent disk is being used for, for example, **jupyterhub-nb-user1-pvc**.
3. Back up the persistent disk that contains your persistent volume (PV).
    a. Select **CREATE SNAPSHOT** from the top navigation.
    The **Create a snapshot** page opens.
    b. Enter a unique **Name** for the snapshot.
    c. Under **Source disk**, verify the persistent disk you want to back up is displayed.
    d. Change any optional settings as needed.
    e. Click **CREATE**.
    The snapshot of the persistent disk is created.

Verification

- The snapshot that you created is visible on the **Snapshots** page in GCP.

Additional resources

- Google Cloud documentation: Create and manage disk snapshots

6.4. UNINSTALLING OPENSHIFT AI

You can use Red Hat OpenShift Cluster Manager to safely uninstall Red Hat OpenShift AI from your OpenShift cluster.

Prerequisites

- Credentials for Red Hat OpenShift Cluster Manager (https://console.redhat.com/openshift/).
- Administrator access to the OpenShift cluster.
- For AWS clusters, you have backed up the EBS volume containing your Persistent Volume Claims (PVCs). See Amazon Web Services documentation: Create Amazon EBS snapshots for more information.

- For GCP clusters, you have backed up the persistent disk containing your Persistent Volume Claims (PVCs). See Google Cloud documentation: Create and manage disk snapshots for more information.

Procedure

1. Log in to Red Hat OpenShift Cluster Manager (https://console.redhat.com/openshift/).
2. Click **Clusters**.
   The **Clusters** page opens.
3. Click the name of the cluster that hosts the instance OpenShift AI to uninstall.
   The **Details** page for the cluster opens.
4. Click the **Add-ons** tab and locate the **Red Hat OpenShift AI** tile.
5. Click **Uninstall**.
   This process takes approximately 30 minutes to complete. Do not manually delete any resources while uninstalling OpenShift AI, as this can interfere with the uninstall process.

   OpenShift AI is uninstalled and any persistent volume claims (PVCs) associated with your OpenShift AI instance are deleted. However, any user groups for OpenShift AI that you previously created remain on your cluster.

Verification

- In Red Hat OpenShift Cluster Manager, on the **Add-ons** tab for the cluster, confirm that the OpenShift Data Science tile does not show the **Installed** state.
- In your OpenShift cluster, click **Home → Projects** and confirm that the following project namespaces are not visible:
  - **redhat-ods-applications**
  - **redhat-ods-monitoring**
  - **redhat-ods-operator**

Additional resources

- Amazon Web Services documentation: Create Amazon EBS snapshots
- Google Cloud documentation: Create and manage disk snapshots
- Deleting users and user resources

6.5. ADDITIONAL RESOURCES

- Deleting users and user resources