Red Hat OpenShift AI Self-Managed 2-latest

Working with distributed workloads

Use distributed workloads for faster and more efficient data processing and model training

Last Updated: 2024-08-27

Red Hat OpenShift AI Self-Managed 2-latest Working with distributed workloads

Use distributed workloads for faster and more efficient data processing and model training

Legal Notice

Abstract

Distributed workloads enable data scientists to use multiple cluster nodes in parallel for faster and more efficient data processing and model training. The CodeFlare framework simplifies task orchestration and monitoring, and offers seamless integration for automated resource scaling and optimal node utilization with advanced GPU support.

Table of Contents

PREFACE 3

CHAPTER 1. OVERVIEW OF DISTRIBUTED WORKLOADS 4
1.1. OVERVIEW OF KUEUE RESOURCES 4
1.1.1. Resource flavor 4
1.1.2. Cluster queue 5
1.1.3. Local queue 6

CHAPTER 2. CONFIGURING DISTRIBUTED WORKLOADS 8
2.1. CONFIGURING THE DISTRIBUTED WORKLOADS COMPONENTS 8
2.2. CONFIGURING QUOTA MANAGEMENT FOR DISTRIBUTED WORKLOADS 10
2.3. CONFIGURING THE CODEFLARE OPERATOR 13

CHAPTER 3. RUNNING DISTRIBUTED WORKLOADS 15
3.1. RUNNING DISTRIBUTED DATA SCIENCE WORKLOADS FROM NOTEBOOKS 15
3.2. RUNNING DISTRIBUTED DATA SCIENCE WORKLOADS FROM DATA SCIENCE PIPELINES 17
3.3. RUNNING DISTRIBUTED DATA SCIENCE WORKLOADS IN A DISCONNECTED ENVIRONMENT 21

CHAPTER 4. MONITORING DISTRIBUTED WORKLOADS 23
4.1. VIEWING PROJECT METRICS FOR DISTRIBUTED WORKLOADS 23
4.2. VIEWING THE STATUS OF DISTRIBUTED WORKLOADS 24

CHAPTER 5. TUNING A MODEL BY USING THE TRAINING OPERATOR 26
5.1. CONFIGURING THE TRAINING OPERATOR PERMISSIONS WHEN NOT USING KUEUE 26
5.2. CONFIGURING THE TRAINING JOB 28
5.3. RUNNING THE TRAINING JOB 32
5.4. MONITORING THE TRAINING JOB 33

CHAPTER 6. TROUBLESHOOTING COMMON PROBLEMS WITH DISTRIBUTED WORKLOADS FOR USERS .. 36
6.1. MY RAY CLUSTER IS IN A SUSPENDED STATE 36
6.2. MY RAY CLUSTER IS IN A FAILED STATE 37
6.3. I SEE A FAILED TO CALL WEBHOOK ERROR MESSAGE FOR THE CODEFLARE OPERATOR 37
6.4. I SEE A FAILED TO CALL WEBHOOK ERROR MESSAGE FOR KUEUE 37
6.5. MY RAY CLUSTER DOESN'T START 38
6.6. I SEE A DEFAULT LOCAL QUEUE ... NOT FOUND ERROR MESSAGE 39
6.7. I SEE A LOCAL_QUEUE PROVIDED DOES NOT EXIST ERROR MESSAGE 39
6.8. I CANNOT CREATE A RAY CLUSTER OR SUBMIT JOBS 40
6.9. MY POD PROVISIONED BY KUEUE IS TERMINATED BEFORE MY IMAGE IS PULLED 40

PREFACE

To train complex machine-learning models or process data more quickly, data scientists can use the distributed workloads feature to run their jobs on multiple OpenShift worker nodes in parallel. This approach significantly reduces the task completion time, and enables the use of larger datasets and more complex models.

CHAPTER 1. OVERVIEW OF DISTRIBUTED WORKLOADS

You can use the distributed workloads feature to queue, scale, and manage the resources required to run data science workloads across multiple nodes in an OpenShift cluster simultaneously. Typically, data science workloads include several types of artificial intelligence (AI) workloads, including machine learning (ML) and Python workloads.

Distributed workloads provide the following benefits:

- You can iterate faster and experiment more frequently because of the reduced processing time.
- You can use larger datasets, which can lead to more accurate models.
- You can use complex models that could not be trained on a single node.
- You can submit distributed workloads at any time, and the system then schedules the distributed workload when the required resources are available.

The distributed workloads infrastructure includes the following components:

CodeFlare Operator

Secures deployed Ray clusters and grants access to their URLs

CodeFlare SDK

Defines and controls the remote distributed compute jobs and infrastructure for any Python-based environment

NOTE

The CodeFlare SDK is not installed as part of OpenShift AI, but it is contained in some of the notebook images provided by OpenShift AI.

KubeRay

Manages remote Ray clusters on OpenShift for running distributed compute workloads

Kueue

Manages quotas and how distributed workloads consume them, and manages the queueing of distributed workloads with respect to quotas

You can run distributed workloads from data science pipelines, from Jupyter notebooks, or from Microsoft Visual Studio Code files.

NOTE

Data science pipelines workloads are not managed by the distributed workloads feature, and are not included in the distributed workloads metrics.

1.1. OVERVIEW OF KUEUE RESOURCES

Cluster administrators can configure Kueue resource flavors, cluster queues, and local queues to manage distributed workload resources across multiple nodes in an OpenShift cluster.

1.1.1. Resource flavor

The Kueue **ResourceFlavor** object describes the resource variations that are available in a cluster.

Resources in a cluster can be *homogenous* or *heterogeneous*:

- Homogeneous resources are identical across the cluster: same node type, CPUs, memory, accelerators, and so on.
- Heterogeneous resources have variations across the cluster.

If a cluster has homogeneous resources, or if it is not necessary to manage separate quotas for different flavors of a resource, a cluster administrator can create an empty **ResourceFlavor** object named **default-flavor**, without any labels or taints, as follows:

Empty Kueue resource flavor for homegeneous resources

```
apiVersion: kueue.x-k8s.io/v1beta1
kind: ResourceFlavor
metadata:
  name: default-flavor
```

If a cluster has heterogeneous resources, cluster administrators can define a different resource flavor for each variation in the resources available. Example variations include different CPUs, different memory, or different accelerators. Cluster administrators can then associate the resource flavors with cluster nodes by using labels, taints, and tolerations, as shown in the following example.

Example Kueue resource flavor for heterogeneous resources

```
apiVersion: kueue.x-k8s.io/v1beta1
kind: ResourceFlavor
metadata:
  name: "spot"
spec:
  nodeLabels:
    instance-type: spot
  nodeTaints:
  - effect: NoSchedule
    key: spot
    value: "true"
  tolerations:
  - key: "spot-taint"
    operator: "Exists"
    effect: "NoSchedule"
```

NOTE

In OpenShift AI 2-latest, Red Hat supports only a single cluster queue per cluster (that is, homogenous clusters), and only empty resource flavors.

For more information about configuring resource flavors, see Resource Flavor in the Kueue documentation.

1.1.2. Cluster queue

The Kueue **ClusterQueue** object manages a pool of cluster resources such as pods, CPUs, memory, and accelerators. A cluster can have multiple cluster queues, and each cluster queue can reference multiple resource flavors.

Cluster administrators can configure cluster queues to define the resource flavors that the queue manages, and assign a quota for each resource in each resource flavor. Cluster administrators can also configure usage limits and queueing strategies to apply fair sharing rules across multiple cluster queues in a cluster.

The following example configures a cluster queue to assign a quota of 9 CPUs, 36 GiB memory, 5 pods, and 5 NVIDIA GPUs.

Example cluster queue

```
apiVersion: kueue.x-k8s.io/v1beta1
kind: ClusterQueue
metadata:
  name: "cluster-queue"
spec:
  namespaceSelector: {} # match all.
  resourceGroups:
  - coveredResources: ["cpu", "memory", "pods", "nvidia.com/gpu"]
    flavors:
    - name: "default-flavor"
      resources:
      - name: "cpu"
        nominalQuota: 9
      - name: "memory"
        nominalQuota: 36Gi
      - name: "pods"
        nominalQuota: 5
      - name: "nvidia.com/gpu"
        nominalQuota: '5'
```

The cluster queue starts a distributed workload only if the total required resources are within these quota limits. If the sum of the requests for a resource in a distributed workload is greater than the specified quota for that resource in the cluster queue, the cluster queue does not admit the distributed workload.

For more information about configuring cluster queues, see Cluster Queue in the Kueue documentation.

1.1.3. Local queue

The Kueue **LocalQueue** object groups closely related distributed workloads in a project. Cluster administrators can configure local queues to specify the project name and the associated cluster queue. Each local queue then grants access to the resources that its specified cluster queue manages. A cluster administrator can optionally define one local queue in a project as the default local queue for that project.

When configuring a distributed workload, the user specifies the local queue name. If a cluster administrator configured a default local queue, the user can omit the local queue specification from the distributed workload code.

Kueue allocates the resources for a distributed workload from the cluster queue that is associated with the local queue, if the total requested resources are within the quota limits specified in that cluster queue.

The following example configures a local queue called **team-a-queue** for the **team-a** project, and specifies **cluster-queue** as the associated cluster queue.

Example local queue

```
apiVersion: kueue.x-k8s.io/v1beta1
kind: LocalQueue
metadata:
  namespace: team-a
  name: team-a-queue
  annotations:
    kueue.x-k8s.io/default-queue: "true"
spec:
  clusterQueue: cluster-queue
```

In this example, the **kueue.x-k8s.io/default-queue: "true"** annotation defines this local queue as the default local queue for the **team-a** project. If a user submits a distributed workload in the **team-a** project and that distributed workload does not specify a local queue in the cluster configuration, Kueue automatically routes the distributed workload to the **team-a-queue** local queue. The distributed workload can then access the resources that the **cluster-queue** cluster queue manages.

For more information about configuring local queues, see Local Queue in the Kueue documentation.

CHAPTER 2. CONFIGURING DISTRIBUTED WORKLOADS

To configure the distributed workloads feature for your data scientists to use in OpenShift AI, you must enable several components in the Red Hat OpenShift AI Operator, create the required Kueue resources, and optionally configure the CodeFlare Operator.

2.1. CONFIGURING THE DISTRIBUTED WORKLOADS COMPONENTS

To configure the distributed workloads feature for your data scientists to use in OpenShift AI, you must enable several components.

Prerequisites

- You have logged in to OpenShift with the **cluster-admin** role.
- You have access to the data science cluster.
- You have installed Red Hat OpenShift AI.
- You have sufficient resources. In addition to the minimum OpenShift AI resources described in Installing and deploying OpenShift AI (for disconnected environments, see Deploying OpenShift AI in a disconnected environment), you need 1.6 vCPU and 2 GiB memory to deploy the distributed workloads infrastructure.
- You have access to a Ray cluster image. For information about how to create a Ray cluster, see the Ray Clusters documentation.

NOTE

Mutual Transport Layer Security (mTLS) is enabled by default in the CodeFlare component in OpenShift AI. OpenShift AI 2-latest does not support the **submissionMode=K8sJobMode** setting in the Ray job specification, so the KubeRay Operator cannot create a submitter Kubernetes Job to submit the Ray job. Instead, users must configure the Ray job specification to set **submissionMode=HTTPMode** only, so that the KubeRay Operator sends a request to the RayCluster to create a Ray job.

- You have access to the data sets and models that the distributed workload uses.
- You have access to the Python dependencies for the distributed workload.
- You have removed any previously installed instances of the CodeFlare Operator, as described in the Knowledgebase solution How to migrate from a separately installed CodeFlare Operator in your data science cluster.
- If you want to use graphics processing units (GPUs), you have enabled GPU support in OpenShift AI. See Enabling NVIDIA GPUs.

NOTE

In OpenShift AI 2-latest, for distributed workloads, Red Hat supports only NVIDIA GPU accelerators. Red Hat supports the use of accelerators within the same cluster only. Red Hat does not support remote direct memory access (RDMA) between accelerators, or the use of accelerators across a network, for example, by using technology such as NVIDIA GPUDirect or NVLink.

- 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 (for disconnected environments, see Working with certificates). No additional configuration is necessary to use those certificates with distributed workloads. The centrally configured self-signed certificates are automatically available in the workload pods at the following mount points:
  - Cluster-wide CA bundle:

    ```
    /etc/pki/tls/certs/odh-trusted-ca-bundle.crt
    /etc/ssl/certs/odh-trusted-ca-bundle.crt
    ```

  - Custom CA bundle:

    ```
    /etc/pki/tls/certs/odh-ca-bundle.crt
    /etc/ssl/certs/odh-ca-bundle.crt
    ```

Procedure

1. In the OpenShift console, click **Operators → Installed Operators**.
2. Search for the **Red Hat OpenShift AI** Operator, and then click the Operator name to open the Operator details page.
3. Click the **Data Science Cluster** tab.
4. Click the default instance name (for example, **default-dsc**) to open the instance details page.
5. Click the **YAML** tab to show the instance specifications.
6. Enable the required distributed workloads components. In the **spec:components** section, set the **managementState** field correctly for the required components. The **trainingoperator** component is required only if you want to use the Kubeflow Training Operator to tune models. The list of required components depends on whether the distributed workload is run from a pipeline or notebook or both, as shown in the following table.

Table 2.1. Components required for distributed workloads

Component	Pipelines only	Notebooks only	Pipelines and notebooks
codeflare	**Managed**	**Managed**	**Managed**
dashboard	**Managed**	**Managed**	**Managed**
datasciencepipelines	**Managed**	**Removed**	**Managed**

Component	Pipelines only	Notebooks only	Pipelines and notebooks
kueue	**Managed**	**Managed**	**Managed**
ray	**Managed**	**Managed**	**Managed**
trainingoperator	**Managed**	**Managed**	**Managed**
workbenches	**Removed**	**Managed**	**Managed**

7. Click **Save**. After a short time, the components with a **Managed** state are ready.

Verification

Check the status of the **codeflare-operator-manager**, **kuberay-operator**, and **kueue-controller-manager** pods, as follows:

1. In the OpenShift console, from the **Project** list, select **redhat-ods-applications**.
2. Click **Workloads → Deployments**.
3. Search for the **codeflare-operator-manager**, **kuberay-operator**, and **kueue-controller-manager** deployments. In each case, check the status as follows:
   a. Click the deployment name to open the deployment details page.
   b. Click the **Pods** tab.
   c. Check the pod status.
   When the status of the **codeflare-operator-manager-*<pod-id>***, **kuberay-operator-*<pod-id>***, and **kueue-controller-manager-*<pod-id>*** pods is **Running**, the pods are ready to use.
   d. To see more information about each pod, click the pod name to open the pod details page, and then click the **Logs** tab.

2.2. CONFIGURING QUOTA MANAGEMENT FOR DISTRIBUTED WORKLOADS

Configure quotas for distributed workloads on a cluster, so that you can share resources between several data science projects.

Prerequisites

- You have cluster administrator privileges for your OpenShift cluster.
- You have downloaded and installed the OpenShift command-line interface (CLI). See Installing the OpenShift CLI.
- You have enabled the required distributed workloads components as described in Configuring the distributed workloads components.
- You have created a data science project that contains a workbench, and the workbench is

running a default notebook image that contains the CodeFlare SDK, for example, the **Standard Data Science** notebook. For information about how to create a project, see Creating a data science project.

- You have sufficient resources. In addition to the base OpenShift AI resources, you need 1.6 vCPU and 2 GiB memory to deploy the distributed workloads infrastructure.
- The resources are physically available in the cluster.

NOTE

In OpenShift AI 2-latest, Red Hat supports only a single cluster queue per cluster (that is, homogenous clusters), and only empty resource flavors. For more information about Kueue resources, see Overview of Kueue resources.

- If you want to use graphics processing units (GPUs), you have enabled GPU support in OpenShift AI. See Enabling NVIDIA GPUs.

NOTE

In OpenShift AI 2-latest, Red Hat supports only NVIDIA GPU accelerators for distributed workloads.

Procedure

1. In a terminal window, if you are not already logged in to your OpenShift cluster as a cluster administrator, log in to the OpenShift CLI as shown in the following example:

```
$ oc login <openshift_cluster_url> -u <admin_username> -p <password>
```

2. Create an empty Kueue resource flavor, as follows:
   a. Create a file called **default_flavor.yaml** and populate it with the following content:

Empty Kueue resource flavor

```
apiVersion: kueue.x-k8s.io/v1beta1
kind: ResourceFlavor
metadata:
  name: default-flavor
```

   b. Apply the configuration to create the **default-flavor** object:

```
$ oc apply -f default_flavor.yaml
```

3. Create a cluster queue to manage the empty Kueue resource flavor, as follows:
   a. Create a file called **cluster_queue.yaml** and populate it with the following content:

Example cluster queue

```
apiVersion: kueue.x-k8s.io/v1beta1
kind: ClusterQueue
metadata:
```

```
  name: "cluster-queue"
spec:
  namespaceSelector: {}  # match all.
  resourceGroups:
  - coveredResources: ["cpu", "memory", "nvidia.com/gpu"]
    flavors:
    - name: "default-flavor"
      resources:
      - name: "cpu"
        nominalQuota: 9
      - name: "memory"
        nominalQuota: 36Gi
      - name: "nvidia.com/gpu"
        nominalQuota: 5
```

b. Replace the example quota values (9 CPUs, 36 GiB memory, and 5 NVIDIA GPUs) with the appropriate values for your cluster queue. The cluster queue will start a distributed workload only if the total required resources are within these quota limits.
You must specify a quota for each resource that the user can request, even if the requested value is 0, by updating the **spec.resourceGroups** section as follows:

- Include the resource name in the **coveredResources** list.
- Specify the resource **name** and **nominalQuota** in the **flavors.resources** section, even if the **nominalQuota** value is 0.

c. Apply the configuration to create the **cluster-queue** object:

```
$ oc apply -f cluster_queue.yaml
```

4. Create a local queue that points to your cluster queue, as follows:

a. Create a file called **local_queue.yaml** and populate it with the following content:

Example local queue

```
apiVersion: kueue.x-k8s.io/v1beta1
kind: LocalQueue
metadata:
  namespace: test
  name: local-queue-test
  annotations:
    kueue.x-k8s.io/default-queue: 'true'
spec:
  clusterQueue: cluster-queue
```

The **kueue.x-k8s.io/default-queue: 'true'** annotation defines this queue as the default queue. Distributed workloads are submitted to this queue if no **local_queue** value is specified in the **ClusterConfiguration** section of the data science pipeline or Jupyter notebook or Microsoft Visual Studio Code file.

b. Update the **namespace** value to specify the same namespace as in the **ClusterConfiguration** section that creates the Ray cluster.

c. Optional: Update the **name** value accordingly.

d. Apply the configuration to create the local-queue object:

```
$ oc apply -f local_queue.yaml
```

The cluster queue allocates the resources to run distributed workloads in the local queue.

Verification

Check the status of the local queue in a project, as follows:

```
$ oc get -n <project-name> localqueues
```

Additional resources

- Kueue documentation

2.3. CONFIGURING THE CODEFLARE OPERATOR

If you want to change the default configuration of the CodeFlare Operator for distributed workloads in OpenShift AI, you can edit the associated config map.

Prerequisites

- You have logged in to OpenShift with the **cluster-admin** role.
- You have enabled the required distributed workloads components as described in Configuring the distributed workloads components.

Procedure

1. In the OpenShift console, click **Workloads → ConfigMaps**.
2. From the **Project** list, select **redhat-ods-applications**.
3. Search for the **codeflare-operator-config** config map, and click the config map name to open the **ConfigMap details** page.
4. Click the **YAML** tab to show the config map specifications.
5. In the **data:config.yaml:kuberay** section, you can edit the following entries:

ingressDomain

This configuration option is null (**ingressDomain: ""**) by default. Do not change this option unless the Ingress Controller is not running on OpenShift. OpenShift AI uses this value to generate the dashboard and client routes for every Ray Cluster, as shown in the following examples:

Example dashboard and client routes

```
ray-dashboard-<clustername>-<namespace>.<your.ingress.domain>
ray-client-<clustername>-<namespace>.<your.ingress.domain>
```

mTLSEnabled

This configuration option is enabled (**mTLSEnabled: true**) by default. When this option is enabled, the Ray Cluster pods create certificates that are used for mutual Transport Layer Security (mTLS), a form of mutual authentication, between Ray Cluster nodes. When this option is enabled, Ray clients cannot connect to the Ray head node unless they download the generated certificates from the **ca-secret-_<cluster_name>_** secret, generate the necessary certificates for mTLS communication, and then set the required Ray environment variables. Users must then re-initialize the Ray clients to apply the changes. The CodeFlare SDK provides the following functions to simplify the authentication process for Ray clients:

Example Ray client authentication code

```
from codeflare_sdk import generate_cert

generate_cert.generate_tls_cert(cluster.config.name, cluster.config.namespace)
generate_cert.export_env(cluster.config.name, cluster.config.namespace)

ray.init(cluster.cluster_uri())
```

rayDashboardOauthEnabled

This configuration option is enabled (**rayDashboardOAuthEnabled: true**) by default. When this option is enabled, OpenShift AI places an OpenShift OAuth proxy in front of the Ray Cluster head node. Users must then authenticate by using their OpenShift cluster login credentials when accessing the Ray Dashboard through the browser. If users want to access the Ray Dashboard in another way (for example, by using the Ray **JobSubmissionClient** class), they must set an authorization header as part of their request, as shown in the following example:

Example authorization header

```
{Authorization: "Bearer <your-openshift-token>"}
```

6. To save your changes, click **Save**.
7. To apply your changes, delete the pod:
   a. Click **Workloads → Pods**.
   b. Find the **codeflare-operator-manager-*<pod-id>*** pod.
   c. Click the options menu (⋮) for that pod, and then click **Delete Pod**. The pod restarts with your changes applied.

Verification

Check the status of the **codeflare-operator-manager** pod, as follows:

1. In the OpenShift console, click **Workloads → Deployments**.
2. Search for the **codeflare-operator-manager** deployment, and then click the deployment name to open the deployment details page.
3. Click the **Pods** tab. When the status of the **codeflare-operator-manager-*<pod-id>*** pod is **Running**, the pod is ready to use. To see more information about the pod, click the pod name to open the pod details page, and then click the **Logs** tab.

CHAPTER 3. RUNNING DISTRIBUTED WORKLOADS

In OpenShift AI, you can run a distributed workload from a notebook or from a pipeline. You can also run distributed workloads in a disconnected environment if you have access to all of the required software.

3.1. RUNNING DISTRIBUTED DATA SCIENCE WORKLOADS FROM NOTEBOOKS

To run a distributed data science workload from a notebook, you must first provide the link to your Ray cluster image.

Prerequisites

- You have access to a data science cluster that is configured to run distributed workloads as described in Configuring distributed workloads.
- Your cluster administrator has created the required Kueue resources as described in Configuring quota management for distributed workloads.
- Optional: Your cluster administrator has defined a *default* local queue for the Ray cluster by creating a **LocalQueue** resource and adding the following annotation to its configuration details, as described in Configuring quota management for distributed workloads:

```
"kueue.x-k8s.io/default-queue": "true"
```

NOTE

If your cluster administrator does not define a default local queue, you must specify a local queue in each notebook.

- You have created a data science project that contains a workbench, and the workbench is running a default notebook image that contains the CodeFlare SDK, for example, the **Standard Data Science** notebook. For information about how to create a project, see Creating a data science project.
- You have Admin access for the data science project.
  - If you created the project, you automatically have Admin access.
  - If you did not create the project, your cluster administrator must give you Admin access.
- You have launched your notebook server and logged in to your notebook editor. The examples in this procedure refer to the JupyterLab integrated development environment (IDE).

Procedure

1. Download the demo notebooks provided by the CodeFlare SDK. The demo notebooks provide guidelines for how to use the CodeFlare stack in your own notebooks.
   To access the demo notebooks, run the **codeflare_sdk.copy_demo_nbs()** function in a Jupyter notebook. The function copies the demo notebooks that are packaged with the currently installed version of the CodeFlare SDK, and clones them into the target location.

   The **codeflare_sdk.copy_demo_nbs()** function accepts two arguments:

- **dir: str = "./demo-notebooks"** specifies the target location for the cloned demo notebooks. If the leading character is a forward slash (/), the path is absolute. Otherwise, the path is relative to the current directory. The default value is **"./demo-notebooks"**.
- **overwrite: bool = False** specifies whether the function should overwrite the contents of an existing directory. When this argument is set to **True**, the function overwrites your customizations in the original demo files. The default value is **False**.
  Examples:

```
codeflare_sdk.copy_demo_nbs("my-dir")
codeflare_sdk.copy_demo_nbs("/absolute/path/to/target-dir")
codeflare_sdk.copy_demo_nbs("my-dir", True)
codeflare_sdk.copy_demo_nbs(dir="my-dir", overwrite=True)
```

2. Locate the downloaded demo notebooks in the JupyterLab interface. In the left navigation pane, double-click **guided-demos**.
3. Update each example demo notebook as follows:
   a. If not already specified, update the **import** section to import the **generate_cert** component:

   Updated import section

   ```
   from codeflare_sdk import generate_cert
   ```

   b. Replace the default namespace value with the name of your data science project.
   c. In the **TokenAuthentication** section of your notebook code, provide the token and server details to authenticate to the OpenShift cluster by using the CodeFlare SDK.
   d. Replace the link to the example community image with a link to your Ray cluster image.
   e. Ensure that the following Ray cluster authentication code is included after the Ray cluster creation section.

   Ray cluster authentication code

   ```
   generate_cert.generate_tls_cert(cluster.config.name, cluster.config.namespace)
   generate_cert.export_env(cluster.config.name, cluster.config.namespace)
   ```

   NOTE

   Mutual Transport Layer Security (mTLS) is enabled by default in the CodeFlare component in OpenShift AI. You must include the Ray cluster authentication code to enable the Ray client that runs within a notebook to connect to a secure Ray cluster that has mTLS enabled.

   f. If you have not configured a default local queue by including the **kueue.x-k8s.io/default-queue: 'true'** annotation as described in Configuring quota management for distributed workloads, update the **ClusterConfiguration** section to specify the local queue for the Ray cluster, as shown in the following example:

   Example local queue assignment

```
local_queue="your_local_queue_name"
```

g. Optional: In the **ClusterConfiguration** section, assign a dictionary of **labels** parameters to the Ray cluster for identification and management purposes, as shown in the following example:

Example labels assignment

```
labels = {"exampleLabel1": "exampleLabel1Value", "exampleLabel2": "exampleLabel2Value"}
```

4. Run the notebooks.

Verification

The notebooks run to completion without errors. In the notebooks, the output from the **cluster.status()** function or **cluster.details()** function indicates that the Ray cluster is **Active**.

3.2. RUNNING DISTRIBUTED DATA SCIENCE WORKLOADS FROM DATA SCIENCE PIPELINES

To run a distributed data science workload from a data science pipeline, you must first update the pipeline to include a link to your Ray cluster image.

Prerequisites

- You have access to a data science cluster that is configured to run distributed workloads as described in Configuring distributed workloads.
- Your cluster administrator has created the required Kueue resources as described in Configuring quota management for distributed workloads.
- Optional: Your cluster administrator has defined a *default* local queue for the Ray cluster by creating a **LocalQueue** resource and adding the following annotation to the configuration details for that **LocalQueue** resource, as described in Configuring quota management for distributed workloads:

```
"kueue.x-k8s.io/default-queue": "true"
```

NOTE

If your cluster administrator does not define a default local queue, you must specify a local queue in each pipeline.

- You have access to S3-compatible object storage.
- You have logged in to Red Hat OpenShift AI.
- You have created a data science project that contains a workbench, and the workbench is running a default notebook image that contains the CodeFlare SDK, for example, the **Standard Data Science** notebook. For information about how to create a project, see Creating a data science project.

- You have Admin access for the data science project.
  - If you created the project, you automatically have Admin access.
  - If you did not create the project, your cluster administrator must give you Admin access.

Procedure

1. Create a data connection to connect the object storage to your data science project, as described in Adding a data connection to your data science project .
2. Configure a pipeline server to use the data connection, as described in Configuring a pipeline server.
3. Create the data science pipeline as follows:
   a. Install the **kfp** Python package, which is required for all pipelines:

```
$ pip install kfp
```

   b. Install any other dependencies that are required for your pipeline.
   c. Build your data science pipeline in Python code.
   For example, create a file named **compile_example.py** with the following content:

```
from kfp import dsl


@dsl.component(
    base_image="registry.redhat.io/ubi8/python-39:latest",
    packages_to_install=['codeflare-sdk']
)


def ray_fn():
  import ray (1)
  from codeflare_sdk import Cluster, ClusterConfiguration, generate_cert (2)


  cluster = Cluster( (3)
      ClusterConfiguration(
          namespace="my_project", (4)
          name="raytest",
          num_workers=1,
          head_cpus="500m",
          min_memory=1,
          max_memory=1,
          worker_extended_resource_requests={“nvidia.com/gpu”: 1}, (5)
          image="quay.io/rhoai/ray:2.23.0-py39-cu121", (6)
          local_queue="local_queue_name", (7)
      )
  )


  print(cluster.status())
```

```
    cluster.up() 8
    cluster.wait_ready() 9
    print(cluster.status())
    print(cluster.details())


    ray_dashboard_uri = cluster.cluster_dashboard_uri()
    ray_cluster_uri = cluster.cluster_uri()
    print(ray_dashboard_uri, ray_cluster_uri)

    # Enable Ray client to connect to secure Ray cluster that has mTLS enabled
    generate_cert.generate_tls_cert(cluster.config.name, cluster.config.namespace) 10
    generate_cert.export_env(cluster.config.name, cluster.config.namespace)


    ray.init(address=ray_cluster_uri)
    print("Ray cluster is up and running: ", ray.is_initialized())


    @ray.remote
    def train_fn(): 11
        # complex training function
        return 100


    result = ray.get(train_fn.remote())
    assert 100 == result
    ray.shutdown()
    cluster.down() 12
    auth.logout()
    return result


@dsl.pipeline( 13
    name="Ray Simple Example",
    description="Ray Simple Example",
)


def ray_integration():
    ray_fn()


if __name__ == '__main__': 14
    from kfp.compiler import Compiler
    Compiler().compile(ray_integration, 'compiled-example.yaml')
```

1. Imports Ray.
2. Imports packages from the CodeFlare SDK to define the cluster functions.
3. Specifies the Ray cluster configuration: replace these example values with the values for your Ray cluster.
4. Optional: Specifies the project where the Ray cluster is created. Replace the example value with the name of your project. If you omit this line, the Ray cluster is created in

the current project.

5. Optional: Specifies the requested accelerators for the Ray cluster (in this example, 1 NVIDIA GPU). If no accelerators are required, set the value to 0 or omit the line. Note: To specify the requested accelerators for the Ray cluster, use the **worker_extended_resource_requests** parameter instead of the deprecated **num_gpus** parameter. For more details, see the CodeFlare SDK documentation.
6. Specifies the location of the Ray cluster image. If you are running this code in a disconnected environment, replace the default value with the location for your environment.
7. Specifies the local queue to which the Ray cluster will be submitted. If a default local queue is configured, you can omit this line.
8. Creates a Ray cluster by using the specified image and configuration.
9. Waits until the Ray cluster is ready before proceeding.
10. Enables the Ray client to connect to a secure Ray cluster that has mutual Transport Layer Security (mTLS) enabled. mTLS is enabled by default in the CodeFlare component in OpenShift AI.
11. Replace the example details in this section with the details for your workload.
12. Removes the Ray cluster when your workload is finished.
13. Replace the example name and description with the values for your workload.
14. Compiles the Python code and saves the output in a YAML file.

d. Compile the Python file (in this example, the **compile_example.py** file):

```
$ python compile_example.py
```

This command creates a YAML file (in this example, **compiled-example.yaml**), which you can import in the next step.

4. Import your data science pipeline, as described in Importing a data science pipeline.
5. Schedule the pipeline run, as described in Scheduling a pipeline run.
6. When the pipeline run is complete, confirm that it is included in the list of triggered pipeline runs, as described in Viewing the details of a pipeline run.

Verification

The YAML file is created and the pipeline run completes without errors.

You can view the run details, as described in Viewing the details of a pipeline run.

Additional resources

- Working with data science pipelines
- Ray Clusters documentation

3.3. RUNNING DISTRIBUTED DATA SCIENCE WORKLOADS IN A DISCONNECTED ENVIRONMENT

To run a distributed data science workload in a disconnected environment, you must be able to access a Ray cluster image, and the data sets and Python dependencies used by the workload, from the disconnected environment.

Prerequisites

- You have logged in to OpenShift with the **cluster-admin** role.
- You have access to the disconnected data science cluster.
- You have installed Red Hat OpenShift AI and created a mirror image as described in Installing and uninstalling OpenShift AI Self-Managed in a disconnected environment.
- You can access the following software from the disconnected cluster:
  - A Ray cluster image
  - The data sets and models to be used by the workload
  - The Python dependencies for the workload, either in a Ray image or in your own Python Package Index (PyPI) server that is available from the disconnected cluster
- You have logged in to Red Hat OpenShift AI.
- You have created a data science project that contains a workbench, and the workbench is running a default notebook image that contains the CodeFlare SDK, for example, the **Standard Data Science** notebook. For information about how to create a project, see Creating a data science project.
- You have Admin access for the data science project.
  - If you created the project, you automatically have Admin access.
  - If you did not create the project, your cluster administrator must give you Admin access.

Procedure

1. Configure the disconnected data science cluster to run distributed workloads as described in Configuring distributed workloads.
2. In the **ClusterConfiguration** section of the notebook or pipeline, ensure that the **image** value specifies a Ray cluster image that you can access from the disconnected environment:
   - Notebooks use the Ray cluster image to create a Ray cluster when running the notebook.
   - Pipelines use the Ray cluster image to create a Ray cluster during the pipeline run.
3. If any of the Python packages required by the workload are not available in the Ray cluster, configure the Ray cluster to download the Python packages from a private PyPI server. For example, set the **PIP_INDEX_URL** and **PIP_TRUSTED_HOST** environment variables for the Ray cluster, to specify the location of the Python dependencies, as shown in the following example:

```
PIP_INDEX_URL: https://pypi-notebook.apps.mylocation.com/simple
PIP_TRUSTED_HOST: pypi-notebook.apps.mylocation.com
```

where

- **PIP_INDEX_URL** specifies the base URL of your private PyPI server (the default value is https://pypi.org).
- **PIP_TRUSTED_HOST** configures Python to mark the specified host as trusted, regardless of whether that host has a valid SSL certificate or is using a secure channel.

4. Run the distributed data science workload, as described in Running distributed data science workloads from notebooks or Running distributed data science workloads from data science pipelines.

Verification

The notebook or pipeline run completes without errors:

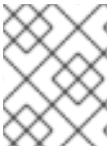
- For notebooks, the output from the **cluster.status()** function or **cluster.details()** function indicates that the Ray cluster is **Active**.
- For pipeline runs, you can view the run details as described in Viewing the details of a pipeline run.

Additional resources

- Installing and uninstalling Red Hat OpenShift AI in a disconnected environment
- Ray Clusters documentation

CHAPTER 4. MONITORING DISTRIBUTED WORKLOADS

In OpenShift AI, you can view project metrics for distributed workloads, and view the status of all distributed workloads in the selected project. You can use these metrics to monitor the resources used by distributed workloads, assess whether project resources are allocated correctly, track the progress of distributed workloads, and identify corrective action when necessary.



NOTE

Data science pipelines workloads are not managed by the distributed workloads feature, and are not included in the distributed workloads metrics.

4.1. VIEWING PROJECT METRICS FOR DISTRIBUTED WORKLOADS

In OpenShift AI, you can view the following project metrics for distributed workloads:

- **CPU** - The number of CPU cores that are currently being used by all distributed workloads in the selected project.
- **Memory** - The amount of memory in gibibytes (GiB) that is currently being used by all distributed workloads in the selected project.

You can use these metrics to monitor the resources used by the distributed workloads, and assess whether project resources are allocated correctly.

Prerequisites

- You have installed Red Hat OpenShift AI.
- On the OpenShift cluster where OpenShift AI is installed, user workload monitoring is enabled.
- You have logged in to 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.
- Your data science project contains distributed workloads.

Procedure

1. In the OpenShift AI left navigation pane, click **Distributed Workloads Metrics**.
2. From the **Project** list, select the project that contains the distributed workloads that you want to monitor.
3. Click the **Project metrics** tab.
4. Optional: From the **Refresh interval** list, select a value to specify how frequently the graphs on the metrics page are refreshed to show the latest data.
   You can select one of these values: **15 seconds**, **30 seconds**, **1 minute**, **5 minutes**, **15 minutes**, **30 minutes**, **1 hour**, **2 hours**, or **1 day**.
5. In the **Requested resources** section, review the **CPU** and **Memory** graphs to identify the resources requested by distributed workloads as follows:

- Requested by the selected project
- Requested by all projects, including the selected project and projects that you cannot access
- Total shared quota for all projects, as provided by the cluster queue

For each resource type (**CPU** and **Memory**), subtract the **Requested by all projects** value from the **Total shared quota** value to calculate how much of that resource quota has not been requested and is available for all projects.

6. Scroll down to the **Top resource-consuming distributed workloads** section to review the following graphs:
   - Top 5 distributed workloads that are consuming the most CPU resources
   - Top 5 distributed workloads that are consuming the most memory

   You can also identify how much CPU or memory is used in each case.
7. Scroll down to view the **Distributed workload resource metrics** table, which lists all of the distributed workloads in the selected project, and indicates the current resource usage and the status of each distributed workload.
   In each table entry, progress bars indicate how much of the requested CPU and memory is currently being used by this distributed workload. To see numeric values for the actual usage and requested usage for CPU (measured in cores) and memory (measured in GiB), hover the cursor over each progress bar. Compare the actual usage with the requested usage to assess the distributed workload configuration. If necessary, reconfigure the distributed workload to reduce or increase the requested resources.

Verification

On the **Project metrics** tab, the graphs and table provide resource-usage data for the distributed workloads in the selected project.

4.2. VIEWING THE STATUS OF DISTRIBUTED WORKLOADS

In OpenShift AI, you can view the status of all distributed workloads in the selected project. You can track the progress of the distributed workloads, and identify corrective action when necessary.

Prerequisites

- You have installed Red Hat OpenShift AI.
- On the OpenShift cluster where OpenShift AI is installed, user workload monitoring is enabled.
- You have logged in to 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.
- Your data science project contains distributed workloads.

Procedure

1. In the OpenShift AI left navigation pane, click **Distributed Workloads Metrics**.

2. From the **Project** list, select the project that contains the distributed workloads that you want to monitor.

3. Click the **Distributed workload status** tab.

4. Optional: From the **Refresh interval** list, select a value to specify how frequently the graphs on the metrics page are refreshed to show the latest data.
   You can select one of these values: **15 seconds**, **30 seconds**, **1 minute**, **5 minutes**, **15 minutes**, **30 minutes**, **1 hour**, **2 hours**, or **1 day**.

5. In the **Status overview** section, review a summary of the status of all distributed workloads in the selected project.
   The status can be **Pending**, **Inadmissible**, **Admitted**, **Running**, **Evicted**, **Succeeded**, or **Failed**.

6. Scroll down to view the **Distributed workloads** table, which lists all of the distributed workloads in the selected project. The table provides the priority, status, creation date, and latest message for each distributed workload.
   The latest message provides more information about the current status of the distributed workload. Review the latest message to identify any corrective action needed. For example, a distributed workload might be **Inadmissible** because the requested resources exceed the available resources. In such cases, you can either reconfigure the distributed workload to reduce the requested resources, or reconfigure the cluster queue for the project to increase the resource quota.

Verification

On the **Distributed workload status** tab, the graph provides a summarized view of the status of all distributed workloads in the selected project, and the table provides more details about the status of each distributed workload.

CHAPTER 5. TUNING A MODEL BY USING THE TRAINING OPERATOR

To tune a model by using the Kubeflow Training Operator, complete the following tasks:

- Configure the Training Operator permissions (you can skip this task if you use Kueue to run PyTorch jobs)
- Configure the training job
- Run the training job
- Monitor the training job

5.1. CONFIGURING THE TRAINING OPERATOR PERMISSIONS WHEN NOT USING KUEUE

If you want to run a PyTorch training job in an environment where Kueue is not installed, you must create a service account and apply access permissions to it.

If you use Kueue to run PyTorch jobs, you can skip this section.

Prerequisites

- You have logged in to OpenShift with the **cluster-admin** role.
- You have access to a data science cluster that is configured to run distributed workloads as described in Configuring distributed workloads.
- You have created a data science project. For information about how to create a project, see Creating a data science project .
- You have Admin access for the data science project.
  - If you created the project, you automatically have Admin access.
  - If you did not create the project, your cluster administrator must give you Admin access.
- You have access to a model.
- You have access to data that you can use to train the model.

Procedure

1. In a terminal window, if you are not already logged in to your OpenShift cluster as a cluster administrator, log in to the OpenShift CLI as shown in the following example:

   ```
   $ oc login <openshift_cluster_url> -u <admin_username> -p <password>
   ```

2. Create a cluster role with the required access permissions, as follows:
   a. Create a YAML file named **pytorchjob_role.yaml**.
   b. Add the following **ClusterRole** object definition:

```
apiVersion: rbac.authorization.k8s.io/v1
kind: ClusterRole
metadata:
  name: pytorchjob-role
rules:
  - apiGroups:
      - "kubeflow.org"
    resources:
      - pytorchjobs
    verbs:
      - create
      - delete
      - get
      - list
      - patch
      - update
      - watch
  - apiGroups:
      - "kubeflow.org"
    resources:
      - pytorchjobs/status
    verbs:
      - get
```

c. Save your changes in the **pytorchjob_role.yaml** file.

d. Apply the configuration to create the **pytorchjob-role** object:

```
$ oc apply -f pytorchjob_role.yaml
```

3. Create a service account, as follows:

a. Create a YAML file named **pytorchjob_serviceaccount.yaml**.

b. Add the following **ServiceAccount** object definition:

```
apiVersion: v1
kind: ServiceAccount
metadata:
  name: pytorchjob-sa
  namespace: kfto
```

c. Replace the example namespace value **kfto** with the name of your project.

d. Save your changes in the **pytorchjob_serviceaccount.yaml** file.

e. Apply the configuration to create the **pytorchjob-sa** service account:

```
$ oc apply -f pytorchjob_serviceaccount.yaml
```

4. Assign the cluster role to the service account, as follows:

a. Create a YAML file named **pytorchjob_rolebinding.yaml**.

b. Add the following **ClusterRoleBinding** object definition:

```
apiVersion: rbac.authorization.k8s.io/v1
kind: ClusterRoleBinding
metadata:
  name: pytorchjob-rolebinding
subjects:
- kind: ServiceAccount
  name: pytorchjob-sa
  namespace: kfto
roleRef:
  kind: ClusterRole
  name: pytorchjob-role
  apiGroup: rbac.authorization.k8s.io
```

c. Replace the example namespace value **kfto** with the name of your project.

d. Save your changes in the **pytorchjob_rolebinding.yaml** file.

e. Apply the configuration to assign the **pytorchjob-role** cluster role to the **pytorchjob-sa** service account:

```
$ oc apply -f pytorchjob_rolebinding.yaml
```

Verification

1. In the OpenShift console, select your project from the **Project** list.
2. Click **User Management → ServiceAccounts** and verify that **pytorchjob-sa** is listed.
3. Click **User Management → Roles** and verify that **pytorchjob-role** is correctly created and assigned.
4. Click **User Management → RoleBindings** and verify that **pytorchjob-rolebinding** is correctly created and assigned.

5.2. CONFIGURING THE TRAINING JOB

Before you can use a training job to tune a model, you must configure the training job. The example training job in this section is based on the IBM and Hugging Face tuning example provided here.

Prerequisites

- You have logged in to OpenShift.
- You have access to a data science cluster that is configured to run distributed workloads as described in Configuring distributed workloads.
- You have created a data science project. For information about how to create a project, see Creating a data science project .
- You have Admin access for the data science project.
  - If you created the project, you automatically have Admin access.
  - If you did not create the project, your cluster administrator must give you Admin access.

- You have access to a model.
- You have access to data that you can use to train the model.

Procedure

1. In a terminal window, if you are not already logged in to your OpenShift cluster, log in to the OpenShift CLI as shown in the following example:

```
$ oc login <openshift_cluster_url> -u <username> -p <password>
```

2. Configure a training job, as follows:

   a. Create a YAML file named **config_trainingjob.yaml**.

   b. Add the following **ConfigMap** object definition:

```
kind: ConfigMap
apiVersion: v1
metadata:
  name: my-config-trainingjob
  namespace: kfto
data:
  config.json: |
    {
      "model_name_or_path": "bigscience/bloom-560m",
      "training_data_path": "/data/input/twitter_complaints.json",
      "output_dir": "/data/output/tuning/bloom-twitter",
      "num_train_epochs": 10.0,
      "per_device_train_batch_size": 4,
      "gradient_accumulation_steps": 4,
      "learning_rate": 1e-05,
      "response_template": "\n### Label:",
      "dataset_text_field": "output",
      "use_flash_attn": false
    }
```

   c. Replace the example namespace value **kfto** with the name of your project.

   d. Edit the following parameters of the training job:

      i. In the **model_name_or_path** field, specify the name of the pre-trained model.

      ii. In the **training_data_path** field, specify the path to the training data that you set in the **training_data.yaml** ConfigMap.

      iii. In the **output_dir** field, specify the output directory for the model.

      iv. In the **num_train_epochs** field, specify the number of epochs for training. In this example, the training job is set to run 10 times.

      v. In the **per_device_train_batch_size** field, specify the batch size; that is, how many data set examples to process together. In this example, the training job processes 4 examples at a time.

vi. In the **gradient_accumulation_steps** field, specify the number of gradient accumulation steps.

vii. In the **learning_rate** field, specify the learning rate for the training.

viii. In the **response_template** field, specify the template formatting for the response.

ix. In the **dataset_text_field** field, specify the dataset field for training output. This field is set in the **training_data.yaml** config map.

x. In the **use_flash_attn** field, specify whether to use flash attention.

e. Save your changes in the **config_trainingjob.yaml** file.

f. Apply the configuration to create the **my-config-trainingjob** object:

```
$ oc apply -f config_trainingjob.yaml
```

3. Create the training data, as follows:

a. Create a YAML file named **training_data.yaml**.

b. Add the following **ConfigMap** object definition:

```
kind: ConfigMap
apiVersion: v1
metadata:
  name: twitter-complaints
  namespace: kfto
data:
  twitter_complaints.json: |
    [
        {"Tweet text":"@HMRCcustomers No this is my first
job","ID":0,"Label":2,"text_label":"no complaint","output":"### Text: @HMRCcustomers
No this is my first job\n\n### Label: no complaint"},
        {"Tweet text":"@KristaMariePark Thank you for your interest! If you decide to
cancel, you can call Customer Care at 1-800-NYTIMES.","ID":1,"Label":2,"text_label":"no
complaint","output":"### Text: @KristaMariePark Thank you for your interest! If you
decide to cancel, you can call Customer Care at 1-800-NYTIMES.\n\n### Label: no
complaint"},
        {"Tweet text":"@EE On Rosneath Arial having good upload and download speeds
but terrible latency 200ms. Why is
this.","ID":3,"Label":1,"text_label":"complaint","output":"### Text: @EE On Rosneath Arial
having good upload and download speeds but terrible latency 200ms. Why is this.\n\n###
Label: complaint"},
        {"Tweet text":"Couples wallpaper, so cute. :)
#BrothersAtHome","ID":4,"Label":2,"text_label":"no complaint","output":"### Text:
Couples wallpaper, so cute. :) #BrothersAtHome\n\n### Label: no complaint"},
        {"Tweet text":"@mckelldogs This might just be me, but-- eyedrops? Artificial tears
are so useful when you're sleep-deprived and sp…
https:\/\/t.co\/WRtNsokblG","ID":5,"Label":2,"text_label":"no complaint","output":"### Text:
@mckelldogs This might just be me, but-- eyedrops? Artificial tears are so useful when
you're sleep-deprived and sp… https:\/\/t.co\/WRtNsokblG\n\n### Label: no complaint"},
        {"Tweet text":"@Yelp can we get the exact calculations for a business rating (for
example if its 4 stars but actually 4.2) or do we use a 3rd party
site?","ID":6,"Label":2,"text_label":"no complaint","output":"### Text: @Yelp can we get
```

```
the exact calculations for a business rating (for example if its 4 stars but actually 4.2) or do we use a 3rd party site?\n\n### Label: no complaint"},
    {"Tweet text":"@nationalgridus I have no water and the bill is current and paid. Can you do something about this?","ID":7,"Label":1,"text_label":"complaint","output":"### Text: @nationalgridus I have no water and the bill is current and paid. Can you do something about this?\n\n### Label: complaint"},
    {"Tweet text":"@JenniferTilly Merry Christmas to as well. You get more stunning every year ","ID":9,"Label":2,"text_label":"no complaint","output":"### Text: @JenniferTilly Merry Christmas to as well. You get more stunning every year \n\n### Label: no complaint"}
  ]
```

c. Replace the example namespace value **kfto** with the name of your project.

d. Replace the example training data with your training data.

e. Save your changes in the **training_data.yaml** file.

f. Apply the configuration to create the training data:

```
$ oc apply -f training_data.yaml
```

4. Create a persistent volume claim (PVC), as follows:

a. Create a YAML file named **trainedmodelpvc.yaml**.

b. Add the following **PersistentVolumeClaim** object definition:

```
apiVersion: v1
kind: PersistentVolumeClaim
metadata:
  name: trained-model
  namespace: kfto
spec:
  accessModes:
    - ReadWriteOnce
  resources:
    requests:
      storage: 10Gi
```

c. Replace the example namespace value **kfto** with the name of your project, and update the other parameters to suit your environment.

d. Save your changes in the **trainedmodelpvc.yaml** file.

e. Apply the configuration to create a Persistent Volume Claim (PVC) for the training job:

```
$ oc apply -f trainedmodelpvc.yaml
```

Verification

1. In the OpenShift console, select your project from the **Project** list.

2. Click **ConfigMaps** and verify that the **my-config-trainingjob** and **twitter-complaints** ConfigMaps are listed.

3. Click **Search**. From the **Resources** list, select **PersistentVolumeClaim** and verify that the **trained-model** PVC is listed.

5.3. RUNNING THE TRAINING JOB

You can run a training job to tune a model. The example training job in this section is based on the IBM and Hugging Face tuning example provided here.

Prerequisites

- You have logged in to OpenShift with the **cluster-admin** role.
- You have access to a data science cluster that is configured to run distributed workloads as described in Configuring distributed workloads.
- You have created a data science project. For information about how to create a project, see Creating a data science project .
- You have Admin access for the data science project.
  - If you created the project, you automatically have Admin access.
  - If you did not create the project, your cluster administrator must give you Admin access.
- You have access to a model.
- You have access to data that you can use to train the model.
- You have configured the training job as described in Configuring the training job.

Procedure

1. In a terminal window, if you are not already logged in to your OpenShift cluster as a cluster administrator, log in to the OpenShift CLI as shown in the following example:

```
$ oc login <openshift_cluster_url> -u <admin_username> -p <password>
```

2. Create a PyTorch training job, as follows:
   a. Create a YAML file named **pytorchjob.yaml**.
   b. Add the following **PyTorchJob** object definition:

```
apiVersion: kubeflow.org/v1
kind: PyTorchJob
metadata:
  name: kfto-demo
  namespace: kfto
spec:
  pytorchReplicaSpecs:
    Master:
      replicas: 1
      restartPolicy: Never
      template:
        spec:
          containers:
```

```
        - env:
            - name: SFT_TRAINER_CONFIG_JSON_PATH
              value: /etc/config/config.json
          image: 'quay.io/modh/fms-hf-tuning:release'
          imagePullPolicy: IfNotPresent
          name: pytorch
          volumeMounts:
            - mountPath: /etc/config
              name: config-volume
            - mountPath: /data/input
              name: dataset-volume
            - mountPath: /data/output
              name: model-volume
      volumes:
        - configMap:
            items:
              - key: config.json
                path: config.json
            name: my-config-trainingjob
          name: config-volume
        - configMap:
            name: twitter-complaints
          name: dataset-volume
        - name: model-volume
          persistentVolumeClaim:
            claimName: trained-model
  runPolicy:
    suspend: false
```

c. Replace the example namespace value **kfto** with the name of your project, and update the other parameters to suit your environment.

d. Edit the parameters of the PyTorch training job, to provide the details for your training job and environment.

e. Save your changes in the **pytorchjob.yaml** file.

f. Apply the configuration to run the PyTorch training job:

```
$ oc apply -f pytorchjob.yaml
```

Verification

1. In the OpenShift console, select your project from the **Project** list.
2. Click **Workloads → Pods** and verify that the ***<project-name>*-demo-master-0** pod is listed.

5.4. MONITORING THE TRAINING JOB

When you run a training job to tune a model, you can monitor the progress of the job. The example training job in this section is based on the IBM and Hugging Face tuning example provided here.

Prerequisites

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

- You have access to a data science cluster that is configured to run distributed workloads as described in Configuring distributed workloads.
- You have created a data science project. For information about how to create a project, see Creating a data science project .
- You have Admin access for the data science project.
  - If you created the project, you automatically have Admin access.
  - If you did not create the project, your cluster administrator must give you Admin access.
- You have access to a model.
- You have access to data that you can use to train the model.
- You are running the training job as described in Running the training job.

Procedure

1. In the OpenShift console, select your project from the **Project** list.
2. Click **Workloads → Pods**.
3. Search for the pod that corresponds to the PyTorch job, that is, ***<project-name>*-demo-master-0**.
   For example, if the project name is **kfto**, the pod name is **kfto-demo-master-0**.
4. Click the pod name to open the pod details page.
5. Click the **Logs** tab to monitor the progress of the job and view status updates, as shown in the following example output:

```
0%| | 0/10 [00:00<?, ?it/s] 10%|▉ | 1/10 [01:10<10:32, 70.32s/it] {'loss': 6.9531, 'grad_norm': 1104.0, 'learning_rate': 9e-06, 'epoch': 1.0}
10%|▉ | 1/10 [01:10<10:32, 70.32s/it] 20%|██ | 2/10 [01:40<06:13, 46.71s/it] 30%|███ | 3/10 [02:26<05:25, 46.55s/it] {'loss': 2.4609, 'grad_norm': 736.0, 'learning_rate': 7e-06, 'epoch': 2.0}
30%|███ | 3/10 [02:26<05:25, 46.55s/it] 40%|████ | 4/10 [03:23<05:02, 50.48s/it] 50%|█████ | 5/10 [03:41<03:13, 38.66s/it] {'loss': 1.7617, 'grad_norm': 328.0, 'learning_rate': 5e-06, 'epoch': 3.0}
50%|█████ | 5/10 [03:41<03:13, 38.66s/it] 60%|██████ | 6/10 [04:54<03:22, 50.58s/it] {'loss': 3.1797, 'grad_norm': 1016.0, 'learning_rate': 4.000000000000001e-06, 'epoch': 4.0}
60%|██████ | 6/10 [04:54<03:22, 50.58s/it] 70%|███████ | 7/10 [06:03<02:49, 56.59s/it] {'loss': 2.9297, 'grad_norm': 984.0, 'learning_rate': 3e-06, 'epoch': 5.0}
70%|███████ | 7/10 [06:03<02:49, 56.59s/it] 80%|████████ | 8/10 [06:38<01:39, 49.57s/it] 90%|█████████ | 9/10 [07:22<00:48, 48.03s/it] {'loss': 1.4219, 'grad_norm': 684.0, 'learning_rate': 1.0000000000000002e-06, 'epoch': 6.0}
90%|█████████ | 9/10 [07:22<00:48, 48.03s/it]100%|██████████| 10/10 [08:25<00:00, 52.53s/it] {'loss': 1.9609, 'grad_norm': 648.0, 'learning_rate': 0.0, 'epoch': 6.67}
100%|██████████| 10/10 [08:25<00:00, 52.53s/it] {'train_runtime': 508.0444, 'train_samples_per_second': 0.197, 'train_steps_per_second': 0.02, 'train_loss': 2.63125, 'epoch': 6.67}
100%|██████████| 10/10 [08:28<00:00, 52.53s/it]100%|██████████| 10/10 [08:28<00:00, 50.80s/it]
```

In the example output, the solid blocks indicate progress bars.

Verification

1. The ***<project-name>*-demo-master-0** pod is running.
2. The **Logs** tab provides information about the job progress and job status.

CHAPTER 6. TROUBLESHOOTING COMMON PROBLEMS WITH DISTRIBUTED WORKLOADS FOR USERS

If you are experiencing errors in Red Hat OpenShift AI relating to distributed workloads, read this section to understand what could be causing the problem, and how to resolve the problem.

If the problem is not documented here or in the release notes, contact Red Hat Support.

6.1. MY RAY CLUSTER IS IN A SUSPENDED STATE

Problem

The resource quota specified in the cluster queue configuration might be insufficient, or the resource flavor might not yet be created.

Diagnosis

The Ray cluster head pod or worker pods remain in a suspended state.

Resolution

1. In the OpenShift console, select your project from the **Project** list.
2. Check the workload resource:
   a. Click **Search**, and from the **Resources** list, select **Workload**.
   b. Select the workload resource that is created with the Ray cluster resource, and click the **YAML** tab.
   c. Check the text in the **status.conditions.message** field, which provides the reason for the suspended state, as shown in the following example:

```
status:
 conditions:
   - lastTransitionTime: '2024-05-29T13:05:09Z'
     message: 'couldn''t assign flavors to pod set small-group-jobtest12: insufficient quota
for nvidia.com/gpu in flavor default-flavor in ClusterQueue'
```

3. Check the Ray cluster resource:
   a. Click **Search**, and from the **Resources** list, select **RayCluster**.
   b. Select the Ray cluster resource, and click the **YAML** tab.
   c. Check the text in the **status.conditions.message** field.
4. Check the cluster queue resource:
   a. Click **Search**, and from the **Resources** list, select **ClusterQueue**.
   b. Check your cluster queue configuration to ensure that the resources that you requested are within the limits defined for the project.
   c. Either reduce your requested resources, or contact your administrator to request more resources.

6.2. MY RAY CLUSTER IS IN A FAILED STATE

Problem

You might have insufficient resources.

Diagnosis

The Ray cluster head pod or worker pods are not running. When a Ray cluster is created, it initially enters a **failed** state. This failed state usually resolves after the reconciliation process completes and the Ray cluster pods are running.

Resolution

If the failed state persists, complete the following steps:

1. In the OpenShift console, select your project from the **Project** list.
2. Click **Search**, and from the **Resources** list, select **Pod**.
3. Click your pod name to open the pod details page.
4. Click the **Events** tab, and review the pod events to identify the cause of the problem.
5. If you cannot resolve the problem, contact your administrator to request assistance.

6.3. I SEE A FAILED TO CALL WEBHOOK ERROR MESSAGE FOR THE CODEFLARE OPERATOR

Problem

After you run the **cluster.up()** command, the following error is shown:

```
ApiException: (500)
Reason: Internal Server Error
HTTP response body: {"kind":"Status","apiVersion":"v1","metadata":
{},"status":"Failure","message":"Internal error occurred: failed calling webhook
\"mraycluster.ray.openshift.ai\": failed to call webhook: Post \"https://codeflare-operator-webhook-
service.redhat-ods-applications.svc:443/mutate-ray-io-v1-raycluster?timeout=10s\": no endpoints
available for service \"codeflare-operator-webhook-service\"","reason":"InternalError","details":
{"causes":[{"message":"failed calling webhook \"mraycluster.ray.openshift.ai\": failed to call webhook:
Post \"https://codeflare-operator-webhook-service.redhat-ods-applications.svc:443/mutate-ray-io-v1-
raycluster?timeout=10s\": no endpoints available for service \"codeflare-operator-webhook-
service\""}]},"code":500}
```

Diagnosis

The CodeFlare Operator pod might not be running.

Resolution

Contact your administrator to request assistance.

6.4. I SEE A FAILED TO CALL WEBHOOK ERROR MESSAGE FOR KUEUE

Problem

After you run the **cluster.up()** command, the following error is shown:

```
ApiException: (500)
Reason: Internal Server Error
HTTP response body: {"kind":"Status","apiVersion":"v1","metadata":
{},"status":"Failure","message":"Internal error occurred: failed calling webhook \"mraycluster.kb.io\":
failed to call webhook: Post \"https://kueue-webhook-service.redhat-ods-applications.svc:443/mutate-
ray-io-v1-raycluster?timeout=10s\": no endpoints available for service \"kueue-webhook-
service\"","reason":"InternalError","details":{"causes":[{"message":"failed calling webhook
\"mraycluster.kb.io\": failed to call webhook: Post \"https://kueue-webhook-service.redhat-ods-
applications.svc:443/mutate-ray-io-v1-raycluster?timeout=10s\": no endpoints available for service
\"kueue-webhook-service\""}]},"code":500}
```

Diagnosis

The Kueue pod might not be running.

Resolution

Contact your administrator to request assistance.

6.5. MY RAY CLUSTER DOESN'T START

Problem

After you run the **cluster.up()** command, when you run either the **cluster.details()** command or the **cluster.status()** command, the Ray Cluster remains in the **Starting** status instead of changing to the **Ready** status. No pods are created.

Diagnosis

1. In the OpenShift console, select your project from the **Project** list.
2. Check the workload resource:
   a. Click **Search**, and from the **Resources** list, select **Workload**.
   b. Select the workload resource that is created with the Ray cluster resource, and click the **YAML** tab.
   c. Check the text in the **status.conditions.message** field, which provides the reason for remaining in the **Starting** state.
3. Check the Ray cluster resource:
   a. Click **Search**, and from the **Resources** list, select **RayCluster**.
   b. Select the Ray cluster resource, and click the **YAML** tab.
   c. Check the text in the **status.conditions.message** field.

Resolution

If you cannot resolve the problem, contact your administrator to request assistance.

6.6. I SEE A DEFAULT LOCAL QUEUE ... NOT FOUNDERROR MESSAGE

Problem

After you run the **cluster.up()** command, the following error is shown:

> Default Local Queue with kueue.x-k8s.io/default-queue: true annotation not found please create a default Local Queue or provide the local_queue name in Cluster Configuration.

Diagnosis

No default local queue is defined, and a local queue is not specified in the cluster configuration.

Resolution

1. In the OpenShift console, select your project from the **Project** list.
2. Click **Search**, and from the **Resources** list, select **LocalQueue**.
3. Resolve the problem in one of the following ways:
   - If a local queue exists, add it to your cluster configuration as follows:

     ```
     local_queue="<local_queue_name>"
     ```

   - If no local queue exists, contact your administrator to request assistance.

6.7. I SEE A LOCAL_QUEUE PROVIDED DOES NOT EXISTERROR MESSAGE

Problem

After you run the **cluster.up()** command, the following error is shown:

> local_queue provided does not exist or is not in this namespace. Please provide the correct local_queue name in Cluster Configuration.

Diagnosis

An incorrect value is specified for the local queue in the cluster configuration, or an incorrect default local queue is defined. The specified local queue either does not exist, or exists in a different namespace.

Resolution

1. In the OpenShift console, select your project from the **Project** list.
2. Click **Search**, and from the **Resources** list, select **LocalQueue**.
3. Resolve the problem in one of the following ways:
   - If a local queue exists, ensure that you spelled the local queue name correctly in your cluster configuration, and that the **namespace** value in the cluster configuration matches your project name. If you do not specify a **namespace** value in the cluster configuration, the Ray

cluster is created in the current project.

- If no local queue exists, contact your administrator to request assistance.

6.8. I CANNOT CREATE A RAY CLUSTER OR SUBMIT JOBS

Problem

After you run the **cluster.up()** command, an error similar to the following error is shown:

```
RuntimeError: Failed to get RayCluster CustomResourceDefinition: (403)
Reason: Forbidden
HTTP response body: {"kind":"Status","apiVersion":"v1","metadata":
{},"status":"Failure","message":"rayclusters.ray.io is forbidden: User
\"system:serviceaccount:regularuser-project:regularuser-workbench\" cannot list resource
\"rayclusters\" in API group \"ray.io\" in the namespace \"regularuser-
project\"","reason":"Forbidden","details":{"group":"ray.io","kind":"rayclusters"},"code":403}
```

Diagnosis

The correct OpenShift login credentials are not specified in the **TokenAuthentication** section of your notebook code.

Resolution

1. Identify the correct OpenShift login credentials as follows:
   a. In the OpenShift console header, click your username and click **Copy login command**.
   b. In the new tab that opens, log in as the user whose credentials you want to use.
   c. Click **Display Token**.
   d. From the **Log in with this token** section, copy the **token** and **server** values.
2. In your notebook code, specify the copied **token** and **server** values as follows:

```
auth = TokenAuthentication(
    token = "<token>",
    server = "<server>",
    skip_tls=False
)
auth.login()
```

6.9. MY POD PROVISIONED BY KUEUE IS TERMINATED BEFORE MY IMAGE IS PULLED

Problem

Kueue waits for a period of time before marking a workload as ready, to enable all of the workload pods to become provisioned and running. By default, Kueue waits for 5 minutes. If the pod image is very large and is still being pulled after the 5-minute waiting period elapses, Kueue fails the workload and terminates the related pods.

Diagnosis

1. In the OpenShift console, select your project from the **Project** list.
2. Click **Search**, and from the **Resources** list, select **Pod**.
3. Click the Ray head pod name to open the pod details page.
4. Click the **Events** tab, and review the pod events to check whether the image pull completed successfully.

Resolution

If the pod takes more than 5 minutes to pull the image, contact your administrator to request assistance.