



# OpenShift Container Platform 4.16

## Specialized hardware and driver enablement

Learn about hardware enablement on OpenShift Container Platform



## OpenShift Container Platform 4.16 Specialized hardware and driver enablement

---

Learn about hardware enablement on OpenShift Container Platform

## Legal Notice

Copyright © 2024 Red Hat, Inc.

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

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

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

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

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

Linux<sup>®</sup> is the registered trademark of Linus Torvalds in the United States and other countries.

Java<sup>®</sup> is a registered trademark of Oracle and/or its affiliates.

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

MySQL<sup>®</sup> is a registered trademark of MySQL AB in the United States, the European Union and other countries.

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

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

All other trademarks are the property of their respective owners.

## Abstract

This document provides an overview of hardware enablement in OpenShift Container Platform.

# Table of Contents

<b>CHAPTER 1. ABOUT SPECIALIZED HARDWARE AND DRIVER ENABLEMENT</b> .....	<b>5</b>
<b>CHAPTER 2. DRIVER TOOLKIT</b> .....	<b>6</b>
2.1. ABOUT THE DRIVER TOOLKIT	6
Background	6
Purpose	7
2.2. PULLING THE DRIVER TOOLKIT CONTAINER IMAGE	7
2.2.1. Pulling the Driver Toolkit container image from registry.redhat.io	7
2.2.2. Finding the Driver Toolkit image URL in the payload	7
2.3. USING THE DRIVER TOOLKIT	8
2.3.1. Build and run the simple-kmod driver container on a cluster	8
2.4. ADDITIONAL RESOURCES	12
<b>CHAPTER 3. NODE FEATURE DISCOVERY OPERATOR</b> .....	<b>13</b>
3.1. ABOUT THE NODE FEATURE DISCOVERY OPERATOR	13
3.2. INSTALLING THE NODE FEATURE DISCOVERY OPERATOR	13
3.2.1. Installing the NFD Operator using the CLI	13
3.2.2. Installing the NFD Operator using the web console	14
3.3. USING THE NODE FEATURE DISCOVERY OPERATOR	15
3.3.1. Create a NodeFeatureDiscovery instance using the CLI	15
3.3.2. Create a NodeFeatureDiscovery CR using the web console	18
3.4. CONFIGURING THE NODE FEATURE DISCOVERY OPERATOR	18
3.4.1. core	18
core.sleepInterval	18
core.sources	18
core.labelWhiteList	19
core.noPublish	19
core.klog	19
core.klog.addDirHeader	19
core.klog.alsologtostderr	19
core.klog.logBacktraceAt	19
core.klog.logDir	20
core.klog.logFile	20
core.klog.logFileMaxSize	20
core.klog.logtostderr	20
core.klog.skipHeaders	20
core.klog.skipLogHeaders	20
core.klog.stderrthreshold	20
core.klog.v	20
core.klog.vmodule	21
3.4.2. sources	21
sources.cpu.cpuid.attributeBlacklist	21
sources.cpu.cpuid.attributeWhitelist	21
sources.kernel.kconfigFile	21
sources.kernel.configOpts	22
sources.pci.deviceClassWhitelist	22
sources.pci.deviceLabelFields	22
sources.usb.deviceClassWhitelist	22
sources.usb.deviceLabelFields	22
sources.custom	23
3.5. ABOUT THE NODEFEATURERULE CUSTOM RESOURCE	23

3.6. USING THE NODEFEATURERULE CUSTOM RESOURCE	23
3.7. USING THE NFD TOPOLOGY UPDATER	24
3.7.1. NodeResourceTopology CR	24
3.7.2. NFD Topology Updater command line flags	25
-ca-file	25
-cert-file	26
-h, -help	26
-key-file	26
-kubelet-config-file	26
-no-publish	26
3.7.2.1. -oneshot	27
-podresources-socket	27
-server	27
-server-name-override	27
-sleep-interval	27
-version	28
-watch-namespace	28
<b>CHAPTER 4. KERNEL MODULE MANAGEMENT OPERATOR</b>	<b>29</b>
4.1. ABOUT THE KERNEL MODULE MANAGEMENT OPERATOR	29
4.2. INSTALLING THE KERNEL MODULE MANAGEMENT OPERATOR	29
4.2.1. Installing the Kernel Module Management Operator using the web console	29
4.2.2. Installing the Kernel Module Management Operator by using the CLI	30
4.2.3. Installing the Kernel Module Management Operator on earlier versions of OpenShift Container Platform	31
4.3. CONFIGURING THE KERNEL MODULE MANAGEMENT OPERATOR	33
4.3.1. Unloading the kernel module	36
4.3.2. Setting the kernel firmware search path	36
4.4. UNINSTALLING THE KERNEL MODULE MANAGEMENT OPERATOR	36
4.4.1. Uninstalling a Red Hat catalog installation	37
4.4.2. Uninstalling a CLI installation	37
4.5. KERNEL MODULE DEPLOYMENT	37
4.5.1. The Module custom resource definition	38
4.5.2. Set soft dependencies between kernel modules	38
4.6. SECURITY AND PERMISSIONS	39
4.6.1. ServiceAccounts and SecurityContextConstraints	39
4.6.2. Pod security standards	40
4.7. REPLACING IN-TREE MODULES WITH OUT-OF-TREE MODULES	40
4.7.1. Example Module CR	41
4.8. SYMBOLIC LINKS FOR IN-TREE DEPENDENCIES	43
4.9. CREATING A KMOD IMAGE	44
4.9.1. Running depmod	44
4.9.1.1. Example Dockerfile	45
4.9.2. Building in the cluster	45
4.9.3. Using the Driver Toolkit	46
4.10. USING SIGNING WITH KERNEL MODULE MANAGEMENT (KMM)	47
4.11. ADDING THE KEYS FOR SECUREBOOT	47
4.11.1. Checking the keys	48
4.12. SIGNING KMODS IN A PRE-BUILT IMAGE	49
4.13. BUILDING AND SIGNING A KMOD IMAGE	50
4.14. KMM HUB AND SPOKE	52
4.14.1. KMM-Hub	52
4.14.2. Installing KMM-Hub	52

---

4.14.2.1. Installing KMM-Hub using the Operator Lifecycle Manager	53
4.14.2.2. Installing KMM-Hub by creating KMM resources	53
4.14.3. Using the ManagedClusterModule CRD	53
4.14.4. Running KMM on the spoke	54
4.15. CUSTOMIZING UPGRADES FOR KERNEL MODULES	56
4.16. DAY 1 KERNEL MODULE LOADING	57
4.16.1. Day 1 supported use cases	58
4.16.2. OOT kernel module loading flow	58
4.16.3. The kernel module image	58
4.16.4. In-tree module replacement	58
4.16.5. MCO yaml creation	59
4.16.6. The MachineConfigPool	59
4.17. DEBUGGING AND TROUBLESHOOTING	60
4.18. KMM FIRMWARE SUPPORT	60
4.18.1. Configuring the lookup path on nodes	60
4.18.2. Building a kmod image	61
4.18.3. Tuning the Module resource	61
4.19. DAY 0 THROUGH DAY 2 KMOD INSTALLATION	62
4.19.1. Layering background	62
4.19.2. Lifecycle management	62
4.19.2.1. Treat the kmod as an in-tree driver	63
4.19.2.2. Use ordered upgrade	63
4.20. TROUBLESHOOTING KMM	63
4.20.1. Reading Operator logs	63
4.20.2. Observing events	64
Build & sign	64
Module load or unload	64
4.20.3. Using the must-gather tool	64
4.20.3.1. Gathering data for KMM	64
4.20.3.2. Gathering data for KMM-Hub	66





# CHAPTER 1. ABOUT SPECIALIZED HARDWARE AND DRIVER ENABLEMENT

The Driver Toolkit (DTK) is a container image in the OpenShift Container Platform payload which is meant to be used as a base image on which to build driver containers. The Driver Toolkit image contains the kernel packages commonly required as dependencies to build or install kernel modules as well as a few tools needed in driver containers. The version of these packages will match the kernel version running on the RHCOS nodes in the corresponding OpenShift Container Platform release.

Driver containers are container images used for building and deploying out-of-tree kernel modules and drivers on container operating systems such as Red Hat Enterprise Linux CoreOS (RHCOS). Kernel modules and drivers are software libraries running with a high level of privilege in the operating system kernel. They extend the kernel functionalities or provide the hardware-specific code required to control new devices. Examples include hardware devices like field-programmable gate arrays (FPGA) or graphics processing units (GPU), and software-defined storage solutions, which all require kernel modules on client machines. Driver containers are the first layer of the software stack used to enable these technologies on OpenShift Container Platform deployments.

## CHAPTER 2. DRIVER TOOLKIT

Learn about the Driver Toolkit and how you can use it as a base image for driver containers for enabling special software and hardware devices on OpenShift Container Platform deployments.

### 2.1. ABOUT THE DRIVER TOOLKIT

#### Background

The Driver Toolkit is a container image in the OpenShift Container Platform payload used as a base image on which you can build driver containers. The Driver Toolkit image includes the kernel packages commonly required as dependencies to build or install kernel modules, as well as a few tools needed in driver containers. The version of these packages will match the kernel version running on the Red Hat Enterprise Linux CoreOS (RHCOS) nodes in the corresponding OpenShift Container Platform release.

Driver containers are container images used for building and deploying out-of-tree kernel modules and drivers on container operating systems like RHCOS. Kernel modules and drivers are software libraries running with a high level of privilege in the operating system kernel. They extend the kernel functionalities or provide the hardware-specific code required to control new devices. Examples include hardware devices like Field Programmable Gate Arrays (FPGA) or GPUs, and software-defined storage (SDS) solutions, such as Lustre parallel file systems, which require kernel modules on client machines. Driver containers are the first layer of the software stack used to enable these technologies on Kubernetes.

The list of kernel packages in the Driver Toolkit includes the following and their dependencies:

- **kernel-core**
- **kernel-devel**
- **kernel-headers**
- **kernel-modules**
- **kernel-modules-extra**

In addition, the Driver Toolkit also includes the corresponding real-time kernel packages:

- **kernel-rt-core**
- **kernel-rt-devel**
- **kernel-rt-modules**
- **kernel-rt-modules-extra**

The Driver Toolkit also has several tools that are commonly needed to build and install kernel modules, including:

- **elfutils-libelf-devel**
- **kmod**
- **binutils-kabi-dw**
- **kernel-abi-whitelists**

- dependencies for the above

### Purpose

Prior to the Driver Toolkit's existence, users would install kernel packages in a pod or build config on OpenShift Container Platform using [entitled builds](#) or by installing from the kernel RPMs in the hosts **machine-os-content**. The Driver Toolkit simplifies the process by removing the entitlement step, and avoids the privileged operation of accessing the machine-os-content in a pod. The Driver Toolkit can also be used by partners who have access to pre-released OpenShift Container Platform versions to prebuild driver-containers for their hardware devices for future OpenShift Container Platform releases.

The Driver Toolkit is also used by the Kernel Module Management (KMM), which is currently available as a community Operator on OperatorHub. KMM supports out-of-tree and third-party kernel drivers and the support software for the underlying operating system. Users can create modules for KMM to build and deploy a driver container, as well as support software like a device plugin, or metrics. Modules can include a build config to build a driver container-based on the Driver Toolkit, or KMM can deploy a prebuilt driver container.

## 2.2. PULLING THE DRIVER TOOLKIT CONTAINER IMAGE

The **driver-toolkit** image is available from the [Container images section of the Red Hat Ecosystem Catalog](#) and in the OpenShift Container Platform release payload. The image corresponding to the most recent minor release of OpenShift Container Platform will be tagged with the version number in the catalog. The image URL for a specific release can be found using the **oc adm** CLI command.

### 2.2.1. Pulling the Driver Toolkit container image from registry.redhat.io

Instructions for pulling the **driver-toolkit** image from **registry.redhat.io** with **podman** or in OpenShift Container Platform can be found on the [Red Hat Ecosystem Catalog](#). The driver-toolkit image for the latest minor release are tagged with the minor release version on **registry.redhat.io**, for example: **registry.redhat.io/openshift4/driver-toolkit-rhel8:v4.16**.

### 2.2.2. Finding the Driver Toolkit image URL in the payload

#### Prerequisites

- You obtained the image [pull secret from Red Hat OpenShift Cluster Manager](#).
- You installed the OpenShift CLI (**oc**).

#### Procedure

1. Use the **oc adm** command to extract the image URL of the **driver-toolkit** corresponding to a certain release:

- For an x86 image, the command is as follows:

```
$ oc adm release info quay.io/openshift-release-dev/ocp-release:4.16.z-x86_64 --image-for=driver-toolkit
```

- For an ARM image, the command is as follows:

```
$ oc adm release info quay.io/openshift-release-dev/ocp-release:4.16.z-aarch64 --image-for=driver-toolkit
```

## Example output

```
quay.io/openshift-release-dev/ocp-v4.0-art-
dev@sha256:b53883ca2bac5925857148c4a1abc300ced96c222498e3bc134fe7ce3a1dd404
```

- Obtain this image using a valid pull secret, such as the pull secret required to install OpenShift Container Platform:

```
$ podman pull --authfile=path/to/pullsecret.json quay.io/openshift-release-dev/ocp-v4.0-art-
dev@sha256:<SHA>
```

## 2.3. USING THE DRIVER TOOLKIT

As an example, the Driver Toolkit can be used as the base image for building a very simple kernel module called **simple-kmod**.



### NOTE

The Driver Toolkit includes the necessary dependencies, **openssl**, **mokutil**, and **keyutils**, needed to sign a kernel module. However, in this example, the **simple-kmod** kernel module is not signed and therefore cannot be loaded on systems with **Secure Boot** enabled.

### 2.3.1. Build and run the simple-kmod driver container on a cluster

#### Prerequisites

- You have a running OpenShift Container Platform cluster.
- You set the Image Registry Operator state to **Managed** for your cluster.
- You installed the OpenShift CLI (**oc**).
- You are logged into the OpenShift CLI as a user with **cluster-admin** privileges.

#### Procedure

Create a namespace. For example:

```
$ oc new-project simple-kmod-demo
```

- The YAML defines an **ImageStream** for storing the **simple-kmod** driver container image, and a **BuildConfig** for building the container. Save this YAML as **0000-buildconfig.yaml.template**.

```
apiVersion: image.openshift.io/v1
kind: ImageStream
metadata:
  labels:
    app: simple-kmod-driver-container
    name: simple-kmod-driver-container
    namespace: simple-kmod-demo
spec: {}
---
```

```

apiVersion: build.openshift.io/v1
kind: BuildConfig
metadata:
  labels:
    app: simple-kmod-driver-build
    name: simple-kmod-driver-build
    namespace: simple-kmod-demo
spec:
  nodeSelector:
    node-role.kubernetes.io/worker: ""
  runPolicy: "Serial"
  triggers:
    - type: "ConfigChange"
    - type: "ImageChange"
  source:
    dockerfile: |
      ARG DTK
      FROM ${DTK} as builder

      ARG KVER

      WORKDIR /build/

      RUN git clone https://github.com/openshift-psap/simple-kmod.git

      WORKDIR /build/simple-kmod

      RUN make all install KVER=${KVER}

      FROM registry.redhat.io/ubi8/ubi-minimal

      ARG KVER

      # Required for installing `modprobe`
      RUN microdnf install kmod

      COPY --from=builder /lib/modules/${KVER}/simple-kmod.ko /lib/modules/${KVER}/
      COPY --from=builder /lib/modules/${KVER}/simple-procfs-kmod.ko
      /lib/modules/${KVER}/
      RUN depmod ${KVER}
    strategy:
      dockerStrategy:
        buildArgs:
          - name: KMODVER
            value: DEMO
            # $ oc adm release info quay.io/openshift-release-dev/ocp-release:<cluster version>-
            x86_64 --image-for=driver-toolkit
          - name: DTK
            value: quay.io/openshift-release-dev/ocp-v4.0-art-
            dev@sha256:34864ccd2f4b6e385705a730864c04a40908e57acede44457a783d739e377cae
          - name: KVER
            value: 4.18.0-372.26.1.el8_6.x86_64
      output:
        to:
          kind: ImageStreamTag
          name: simple-kmod-driver-container:demo

```

- Substitute the correct driver toolkit image for the OpenShift Container Platform version you are running in place of "DRIVER\_TOOLKIT\_IMAGE" with the following commands.

```
$ OCP_VERSION=$(oc get clusterversion/version -ojsonpath={.status.desired.version})
```

```
$ DRIVER_TOOLKIT_IMAGE=$(oc adm release info $OCP_VERSION --image-for=driver-toolkit)
```

```
$ sed "s#DRIVER_TOOLKIT_IMAGE#${DRIVER_TOOLKIT_IMAGE}#" 0000-buildconfig.yaml.template > 0000-buildconfig.yaml
```

- Create the image stream and build config with

```
$ oc create -f 0000-buildconfig.yaml
```

- After the builder pod completes successfully, deploy the driver container image as a **DaemonSet**.

- The driver container must run with the privileged security context in order to load the kernel modules on the host. The following YAML file contains the RBAC rules and the **DaemonSet** for running the driver container. Save this YAML as **1000-drivercontainer.yaml**.

```
apiVersion: v1
kind: ServiceAccount
metadata:
  name: simple-kmod-driver-container
---
apiVersion: rbac.authorization.k8s.io/v1
kind: Role
metadata:
  name: simple-kmod-driver-container
rules:
- apiGroups:
  - security.openshift.io
  resources:
  - securitycontextconstraints
  verbs:
  - use
  resourceNames:
  - privileged
---
apiVersion: rbac.authorization.k8s.io/v1
kind: RoleBinding
metadata:
  name: simple-kmod-driver-container
roleRef:
  apiGroup: rbac.authorization.k8s.io
  kind: Role
  name: simple-kmod-driver-container
subjects:
- kind: ServiceAccount
  name: simple-kmod-driver-container
userNames:
- system:serviceaccount:simple-kmod-demo:simple-kmod-driver-container
```

```

---
apiVersion: apps/v1
kind: DaemonSet
metadata:
  name: simple-kmod-driver-container
spec:
  selector:
    matchLabels:
      app: simple-kmod-driver-container
  template:
    metadata:
      labels:
        app: simple-kmod-driver-container
    spec:
      serviceAccount: simple-kmod-driver-container
      serviceAccountName: simple-kmod-driver-container
      containers:
        - image: image-registry.openshift-image-registry.svc:5000/simple-kmod-
          demo/simple-kmod-driver-container:demo
          name: simple-kmod-driver-container
          imagePullPolicy: Always
          command: [sleep, infinity]
          lifecycle:
            postStart:
              exec:
                command: ["modprobe", "-v", "-a", "simple-kmod", "simple-procfs-kmod"]
            preStop:
              exec:
                command: ["modprobe", "-r", "-a", "simple-kmod", "simple-procfs-kmod"]
          securityContext:
            privileged: true
          nodeSelector:
            node-role.kubernetes.io/worker: ""

```

- b. Create the RBAC rules and daemon set:

```
$ oc create -f 1000-drivercontainer.yaml
```

5. After the pods are running on the worker nodes, verify that the **simple\_kmod** kernel module is loaded successfully on the host machines with **lsmod**.

- a. Verify that the pods are running:

```
$ oc get pod -n simple-kmod-demo
```

### Example output

```

NAME                                READY STATUS   RESTARTS AGE
simple-kmod-driver-build-1-build     0/1   Completed 0      6m
simple-kmod-driver-container-b22fd  1/1   Running   0      40s
simple-kmod-driver-container-jz9vn  1/1   Running   0      40s
simple-kmod-driver-container-p45cc  1/1   Running   0      40s

```

- b. Execute the **lsmod** command in the driver container pod:

■

```
$ oc exec -it pod/simple-kmod-driver-container-p45cc -- lsmod | grep simple
```

### Example output

```
simple_procfs_kmod 16384 0
simple_kmod        16384 0
```

## 2.4. ADDITIONAL RESOURCES

- For more information about configuring registry storage for your cluster, see [Image Registry Operator in OpenShift Container Platform](#).



## CHAPTER 3. NODE FEATURE DISCOVERY OPERATOR

Learn about the Node Feature Discovery (NFD) Operator and how you can use it to expose node-level information by orchestrating Node Feature Discovery, a Kubernetes add-on for detecting hardware features and system configuration.

### 3.1. ABOUT THE NODE FEATURE DISCOVERY OPERATOR

The Node Feature Discovery Operator (NFD) manages the detection of hardware features and configuration in an OpenShift Container Platform cluster by labeling the nodes with hardware-specific information. NFD labels the host with node-specific attributes, such as PCI cards, kernel, operating system version, and so on.

The NFD Operator can be found on the Operator Hub by searching for “Node Feature Discovery”.

### 3.2. INSTALLING THE NODE FEATURE DISCOVERY OPERATOR

The Node Feature Discovery (NFD) Operator orchestrates all resources needed to run the NFD daemon set. As a cluster administrator, you can install the NFD Operator by using the OpenShift Container Platform CLI or the web console.

#### 3.2.1. Installing the NFD Operator using the CLI

As a cluster administrator, you can install the NFD Operator using the CLI.

##### Prerequisites

- An OpenShift Container Platform cluster
- Install the OpenShift CLI (**oc**).
- Log in as a user with **cluster-admin** privileges.

##### Procedure

1. Create a namespace for the NFD Operator.
  - a. Create the following **Namespace** custom resource (CR) that defines the **openshift-nfd** namespace, and then save the YAML in the **nfd-namespace.yaml** file:

```
apiVersion: v1
kind: Namespace
metadata:
  name: openshift-nfd
```

- b. Create the namespace by running the following command:

```
$ oc create -f nfd-namespace.yaml
```

2. Install the NFD Operator in the namespace you created in the previous step by creating the following objects:
  - a. Create the following **OperatorGroup** CR and save the YAML in the **nfd-operatorgroup.yaml** file:

```

apiVersion: operators.coreos.com/v1
kind: OperatorGroup
metadata:
  generateName: openshift-nfd-
  name: openshift-nfd
  namespace: openshift-nfd
spec:
  targetNamespaces:
    - openshift-nfd

```

- b. Create the **OperatorGroup** CR by running the following command:

```
$ oc create -f nfd-operatorgroup.yaml
```

- c. Create the following **Subscription** CR and save the YAML in the **nfd-sub.yaml** file:

### Example Subscription

```

apiVersion: operators.coreos.com/v1alpha1
kind: Subscription
metadata:
  name: nfd
  namespace: openshift-nfd
spec:
  channel: "stable"
  installPlanApproval: Automatic
  name: nfd
  source: redhat-operators
  sourceNamespace: openshift-marketplace

```

- d. Create the subscription object by running the following command:

```
$ oc create -f nfd-sub.yaml
```

- e. Change to the **openshift-nfd** project:

```
$ oc project openshift-nfd
```

### Verification

- To verify that the Operator deployment is successful, run:

```
$ oc get pods
```

### Example output

```

NAME                                READY STATUS RESTARTS AGE
nfd-controller-manager-7f86ccfb58-vgr4x 2/2   Running 0      10m

```

A successful deployment shows a **Running** status.

## 3.2.2. Installing the NFD Operator using the web console

As a cluster administrator, you can install the NFD Operator using the web console.

### Procedure

1. In the OpenShift Container Platform web console, click **Operators** → **OperatorHub**.
2. Choose **Node Feature Discovery** from the list of available Operators, and then click **Install**.
3. On the **Install Operator** page, select **A specific namespace on the cluster**, and then click **Install**. You do not need to create a namespace because it is created for you.

### Verification

To verify that the NFD Operator installed successfully:

1. Navigate to the **Operators** → **Installed Operators** page.
2. Ensure that **Node Feature Discovery** is listed in the **openshift-nfd** project with a **Status** of **InstallSucceeded**.



#### NOTE

During installation an Operator might display a **Failed** status. If the installation later succeeds with an **InstallSucceeded** message, you can ignore the **Failed** message.

### Troubleshooting

If the Operator does not appear as installed, troubleshoot further:

1. Navigate to the **Operators** → **Installed Operators** page and inspect the **Operator Subscriptions** and **Install Plans** tabs for any failure or errors under **Status**.
2. Navigate to the **Workloads** → **Pods** page and check the logs for pods in the **openshift-nfd** project.

## 3.3. USING THE NODE FEATURE DISCOVERY OPERATOR

The Node Feature Discovery (NFD) Operator orchestrates all resources needed to run the Node-Feature-Discovery daemon set by watching for a **NodeFeatureDiscovery** CR. Based on the **NodeFeatureDiscovery** CR, the Operator will create the operand (NFD) components in the desired namespace. You can edit the CR to choose another **namespace**, **image**, **imagePullPolicy**, and **nfd-worker-conf**, among other options.

As a cluster administrator, you can create a **NodeFeatureDiscovery** instance using the OpenShift Container Platform CLI or the web console.

### 3.3.1. Create a NodeFeatureDiscovery instance using the CLI

As a cluster administrator, you can create a **NodeFeatureDiscovery** CR instance using the CLI.

#### Prerequisites

- An OpenShift Container Platform cluster

- Install the OpenShift CLI (**oc**).
- Log in as a user with **cluster-admin** privileges.
- Install the NFD Operator.

## Procedure

1. Create the following **NodeFeatureDiscovery** Custom Resource (CR), and then save the YAML in the **NodeFeatureDiscovery.yaml** file:

```

apiVersion: nfd.openshift.io/v1
kind: NodeFeatureDiscovery
metadata:
  name: nfd-instance
  namespace: openshift-nfd
spec:
  instance: "" # instance is empty by default
  topologyupdater: false # False by default
  operand:
    image: registry.redhat.io/openshift4/ose-node-feature-discovery:v4.16
    imagePullPolicy: Always
  workerConfig:
    configData: |
      core:
        # labelWhiteList:
        # noPublish: false
        sleepInterval: 60s
        # sources: [all]
        # klog:
        # addDirHeader: false
        # alsologtostderr: false
        # logBacktraceAt:
        # logtostderr: true
        # skipHeaders: false
        # stderrthreshold: 2
        # v: 0
        # vmodule:
        ## NOTE: the following options are not dynamically run-time configurable
        ##       and require a nfd-worker restart to take effect after being changed
        # logDir:
        # logFile:
        # logFileMaxSize: 1800
        # skipLogHeaders: false
      sources:
        cpu:
          cpuid:
            # NOTE: whitelist has priority over blacklist
            attributeBlacklist:
              - "BMI1"
              - "BMI2"
              - "CLMUL"
              - "CMOV"
              - "CX16"
              - "ERMS"
              - "F16C"

```

```

- "HTT"
- "LZCNT"
- "MMX"
- "MMXEXT"
- "NX"
- "POPCNT"
- "RDRAND"
- "RDSEED"
- "RDTSCP"
- "SGX"
- "SSE"
- "SSE2"
- "SSE3"
- "SSE4.1"
- "SSE4.2"
- "SSSE3"
attributeWhitelist:
kernel:
  kconfigFile: "/path/to/kconfig"
  configOpts:
    - "NO_HZ"
    - "X86"
    - "DMI"
pci:
  deviceClassWhitelist:
    - "0200"
    - "03"
    - "12"
  deviceLabelFields:
    - "class"
customConfig:
  configData: |
    - name: "more.kernel.features"
  matchOn:
    - loadedKMod: ["example_kmod3"]

```

For more details on how to customize NFD workers, refer to the [Configuration file reference of nfd-worker](#).

1. Create the **NodeFeatureDiscovery** CR instance by running the following command:

```
$ oc create -f NodeFeatureDiscovery.yaml
```

## Verification

- To verify that the instance is created, run:

```
$ oc get pods
```

## Example output

```

NAME                                READY STATUS RESTARTS AGE
nfd-controller-manager-7f86ccfb58-vgr4x 2/2   Running 0    11m
nfd-master-hcn64                       1/1   Running 0    60s

```

nfd-master-lnnxx	1/1	Running	0	60s
nfd-master-mp6hr	1/1	Running	0	60s
nfd-worker-vgcz9	1/1	Running	0	60s
nfd-worker-xqbws	1/1	Running	0	60s

A successful deployment shows a **Running** status.

### 3.3.2. Create a NodeFeatureDiscovery CR using the web console

#### Procedure

1. Navigate to the **Operators** → **Installed Operators** page.
2. Find **Node Feature Discovery** and see a box under **Provided APIs**.
3. Click **Create instance**.
4. Edit the values of the **NodeFeatureDiscovery** CR.
5. Click **Create**.

## 3.4. CONFIGURING THE NODE FEATURE DISCOVERY OPERATOR

### 3.4.1. core

The **core** section contains common configuration settings that are not specific to any particular feature source.

#### core.sleepInterval

**core.sleepInterval** specifies the interval between consecutive passes of feature detection or re-detection, and thus also the interval between node re-labeling. A non-positive value implies infinite sleep interval; no re-detection or re-labeling is done.

This value is overridden by the deprecated **--sleep-interval** command line flag, if specified.

#### Example usage

```
core:
  sleepInterval: 60s 1
```

The default value is **60s**.

#### core.sources

**core.sources** specifies the list of enabled feature sources. A special value **all** enables all feature sources.

This value is overridden by the deprecated **--sources** command line flag, if specified.

Default: **[all]**

#### Example usage

```
core:
  sources:
```

- system
- custom

### core.labelWhiteList

**core.labelWhiteList** specifies a regular expression for filtering feature labels based on the label name. Non-matching labels are not published.

The regular expression is only matched against the basename part of the label, the part of the name after '/'. The label prefix, or namespace, is omitted.

This value is overridden by the deprecated **--label-whitelist** command line flag, if specified.

Default: **null**

### Example usage

```
core:
  labelWhiteList: '^cpu-cpuid'
```

### core.noPublish

Setting **core.noPublish** to **true** disables all communication with the **nfd-master**. It is effectively a dry run flag; **nfd-worker** runs feature detection normally, but no labeling requests are sent to **nfd-master**.

This value is overridden by the **--no-publish** command line flag, if specified.

Example:

### Example usage

```
core:
  noPublish: true 1
```

The default value is **false**.

### core.klog

The following options specify the logger configuration, most of which can be dynamically adjusted at run-time.

The logger options can also be specified using command line flags, which take precedence over any corresponding config file options.

#### core.klog.addDirHeader

If set to **true**, **core.klog.addDirHeader** adds the file directory to the header of the log messages.

Default: **false**

Run-time configurable: yes

#### core.klog.alsologtostderr

Log to standard error as well as files.

Default: **false**

Run-time configurable: yes

#### core.klog.logBacktraceAt

When logging hits line file:N, emit a stack trace.

Default: **empty**

Run-time configurable: yes

#### **core.klog.logDir**

If non-empty, write log files in this directory.

Default: **empty**

Run-time configurable: no

#### **core.klog.logFile**

If not empty, use this log file.

Default: **empty**

Run-time configurable: no

#### **core.klog.logFileMaxSize**

**core.klog.logFileMaxSize** defines the maximum size a log file can grow to. Unit is megabytes. If the value is **0**, the maximum file size is unlimited.

Default: **1800**

Run-time configurable: no

#### **core.klog.logtostderr**

Log to standard error instead of files

Default: **true**

Run-time configurable: yes

#### **core.klog.skipHeaders**

If **core.klog.skipHeaders** is set to **true**, avoid header prefixes in the log messages.

Default: **false**

Run-time configurable: yes

#### **core.klog.skipLogHeaders**

If **core.klog.skipLogHeaders** is set to **true**, avoid headers when opening log files.

Default: **false**

Run-time configurable: no

#### **core.klog.stderrthreshold**

Logs at or above this threshold go to stderr.

Default: **2**

Run-time configurable: yes

#### **core.klog.v**

**core.klog.v** is the number for the log level verbosity.



Default: **0**

Run-time configurable: yes

#### **core.klog.vmodule**

**core.klog.vmodule** is a comma-separated list of **pattern=N** settings for file-filtered logging.

Default: **empty**

Run-time configurable: yes

### 3.4.2. sources

The **sources** section contains feature source specific configuration parameters.

#### **sources.cpu.cpuid.attributeBlacklist**

Prevent publishing **cpuid** features listed in this option.

This value is overridden by **sources.cpu.cpuid.attributeWhitelist**, if specified.

Default: **[BMI1, BMI2, CLMUL, CMOV, CX16, ERMS, F16C, HTT, LZCNT, MMX, MMXEXT, NX, POPCNT, RDRAND, RDSEED, RDTSCP, SGX, SGXLC, SSE, SSE2, SSE3, SSE4.1, SSE4.2, SSSE3]**

#### Example usage

```
sources:
  cpu:
    cpuid:
      attributeBlacklist: [MMX, MMXEXT]
```

#### **sources.cpu.cpuid.attributeWhitelist**

Only publish the **cpuid** features listed in this option.

**sources.cpu.cpuid.attributeWhitelist** takes precedence over **sources.cpu.cpuid.attributeBlacklist**.

Default: **empty**

#### Example usage

```
sources:
  cpu:
    cpuid:
      attributeWhitelist: [AVX512BW, AVX512CD, AVX512DQ, AVX512F, AVX512VL]
```

#### **sources.kernel.kconfigFile**

**sources.kernel.kconfigFile** is the path of the kernel config file. If empty, NFD runs a search in the well-known standard locations.

Default: **empty**

#### Example usage

```
sources:
  kernel:
    kconfigFile: "/path/to/kconfig"
```

**sources.kernel.configOpts**

**sources.kernel.configOpts** represents kernel configuration options to publish as feature labels.

Default: **[NO\_HZ, NO\_HZ\_IDLE, NO\_HZ\_FULL, PREEMPT]**

**Example usage**

```
sources:
  kernel:
    configOpts: [NO_HZ, X86, DMI]
```

**sources.pci.deviceClassWhitelist**

**sources.pci.deviceClassWhitelist** is a list of [PCI device class IDs](#) for which to publish a label. It can be specified as a main class only (for example, **03**) or full class-subclass combination (for example **0300**). The former implies that all subclasses are accepted. The format of the labels can be further configured with **deviceLabelFields**.

Default: **["03", "0b40", "12"]**

**Example usage**

```
sources:
  pci:
    deviceClassWhitelist: ["0200", "03"]
```

**sources.pci.deviceLabelFields**

**sources.pci.deviceLabelFields** is the set of PCI ID fields to use when constructing the name of the feature label. Valid fields are **class**, **vendor**, **device**, **subsystem\_vendor** and **subsystem\_device**.

Default: **[class, vendor]**

**Example usage**

```
sources:
  pci:
    deviceLabelFields: [class, vendor, device]
```

With the example config above, NFD would publish labels such as **feature.node.kubernetes.io/pci-<class-id>\_<vendor-id>\_<device-id>.present=true**

**sources.usb.deviceClassWhitelist**

**sources.usb.deviceClassWhitelist** is a list of USB [device class](#) IDs for which to publish a feature label. The format of the labels can be further configured with **deviceLabelFields**.

Default: **["0e", "ef", "fe", "ff"]**

**Example usage**

```
sources:
  usb:
    deviceClassWhitelist: ["ef", "ff"]
```

**sources.usb.deviceLabelFields**

**sources.usb.deviceLabelFields** is the set of USB ID fields from which to compose the name of the feature label. Valid fields are **class**, **vendor**, and **device**.

Default: **[class, vendor, device]**

### Example usage

```
sources:
  pci:
    deviceLabelFields: [class, vendor]
```

With the example config above, NFD would publish labels like: **feature.node.kubernetes.io/usb-<class-id>\_<vendor-id>.present=true**.

### sources.custom

**sources.custom** is the list of rules to process in the custom feature source to create user-specific labels.

Default: **empty**

### Example usage

```
source:
  custom:
    - name: "my.custom.feature"
      matchOn:
        - loadedKMod: ["e1000e"]
        - pcid:
            class: ["0200"]
            vendor: ["8086"]
```

## 3.5. ABOUT THE NODEFEATURERULE CUSTOM RESOURCE

**NodeFeatureRule** objects are a **NodeFeatureDiscovery** custom resource designed for rule-based custom labeling of nodes. Some use cases include application-specific labeling or distribution by hardware vendors to create specific labels for their devices.

**NodeFeatureRule** objects provide a method to create vendor- or application-specific labels and taints. It uses a flexible rule-based mechanism for creating labels and optionally taints based on node features.

## 3.6. USING THE NODEFEATURERULE CUSTOM RESOURCE

Create a **NodeFeatureRule** object to label nodes if a set of rules match the conditions.

### Procedure

1. Create a custom resource file named **nodefeaturerule.yaml** that contains the following text:

```
apiVersion: nfd.openshift.io/v1
kind: NodeFeatureRule
metadata:
  name: example-rule
spec:
  rules:
```

```

- name: "example rule"
  labels:
    "example-custom-feature": "true"
  # Label is created if all of the rules below match
  matchFeatures:
    # Match if "veth" kernel module is loaded
    - feature: kernel.loadedmodule
      matchExpressions:
        veth: {op: Exists}
    # Match if any PCI device with vendor 8086 exists in the system
    - feature: pci.device
      matchExpressions:
        vendor: {op: In, value: ["8086"]}

```

This custom resource specifies that labelling occurs when the **veth** module is loaded and any PCI device with vendor code **8086** exists in the cluster.

2. Apply the **nodefeaturerule.yaml** file to your cluster by running the following command:

```
$ oc apply -f https://raw.githubusercontent.com/kubernetes-sigs/node-feature-discovery/v0.13.6/examples/nodefeaturerule.yaml
```

The example applies the feature label on nodes with the **veth** module loaded and any PCI device with vendor code **8086** exists.



#### NOTE

A relabeling delay of up to 1 minute might occur.

## 3.7. USING THE NFD TOPOLOGY UPDATER

The Node Feature Discovery (NFD) Topology Updater is a daemon responsible for examining allocated resources on a worker node. It accounts for resources that are available to be allocated to new pod on a per-zone basis, where a zone can be a Non-Uniform Memory Access (NUMA) node. The NFD Topology Updater communicates the information to nfd-master, which creates a **NodeResourceTopology** custom resource (CR) corresponding to all of the worker nodes in the cluster. One instance of the NFD Topology Updater runs on each node of the cluster.

To enable the Topology Updater workers in NFD, set the **topologyupdater** variable to **true** in the **NodeFeatureDiscovery** CR, as described in the section **Using the Node Feature Discovery Operator**.

### 3.7.1. NodeResourceTopology CR

When run with NFD Topology Updater, NFD creates custom resource instances corresponding to the node resource hardware topology, such as:

```

apiVersion: topology.node.k8s.io/v1alpha1
kind: NodeResourceTopology
metadata:
  name: node1
topologyPolicies: ["SingleNUMANodeContainerLevel"]
zones:
  - name: node-0
    type: Node

```

```

resources:
  - name: cpu
    capacity: 20
    allocatable: 16
    available: 10
  - name: vendor/nic1
    capacity: 3
    allocatable: 3
    available: 3
- name: node-1
  type: Node
  resources:
    - name: cpu
      capacity: 30
      allocatable: 30
      available: 15
    - name: vendor/nic2
      capacity: 6
      allocatable: 6
      available: 6
- name: node-2
  type: Node
  resources:
    - name: cpu
      capacity: 30
      allocatable: 30
      available: 15
    - name: vendor/nic1
      capacity: 3
      allocatable: 3
      available: 3

```

### 3.7.2. NFD Topology Updater command line flags

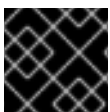
To view available command line flags, run the **nfd-topology-updater -help** command. For example, in a podman container, run the following command:

```
$ podman run gcr.io/k8s-staging-nfd/node-feature-discovery:master nfd-topology-updater -help
```

#### **-ca-file**

The **-ca-file** flag is one of the three flags, together with the **-cert-file** and **-key-file** flags, that controls the mutual TLS authentication on the NFD Topology Updater. This flag specifies the TLS root certificate that is used for verifying the authenticity of nfd-master.

Default: empty



#### IMPORTANT

The **-ca-file** flag must be specified together with the **-cert-file** and **-key-file** flags.

#### Example

```
$ nfd-topology-updater -ca-file=/opt/nfd/ca.crt -cert-file=/opt/nfd/updater.crt -key-file=/opt/nfd/updater.key
```

**-cert-file**

The **-cert-file** flag is one of the three flags, together with the **-ca-file** and **-key-file flags**, that controls mutual TLS authentication on the NFD Topology Updater. This flag specifies the TLS certificate presented for authenticating outgoing requests.

Default: empty

**IMPORTANT**

The **-cert-file** flag must be specified together with the **-ca-file** and **-key-file** flags.

**Example**

```
$ nfd-topology-updater -cert-file=/opt/nfd/updater.crt -key-file=/opt/nfd/updater.key -ca-file=/opt/nfd/ca.crt
```

**-h, -help**

Print usage and exit.

**-key-file**

The **-key-file** flag is one of the three flags, together with the **-ca-file** and **-cert-file** flags, that controls the mutual TLS authentication on the NFD Topology Updater. This flag specifies the private key corresponding the given certificate file, or **-cert-file**, that is used for authenticating outgoing requests.

Default: empty

**IMPORTANT**

The **-key-file** flag must be specified together with the **-ca-file** and **-cert-file** flags.

**Example**

```
$ nfd-topology-updater -key-file=/opt/nfd/updater.key -cert-file=/opt/nfd/updater.crt -ca-file=/opt/nfd/ca.crt
```

**-kubelet-config-file**

The **-kubelet-config-file** specifies the path to the Kubelet's configuration file.

Default: **/host-var/lib/kubelet/config.yaml**

**Example**

```
$ nfd-topology-updater -kubelet-config-file=/var/lib/kubelet/config.yaml
```

**-no-publish**

The **-no-publish** flag disables all communication with the nfd-master, making it a dry run flag for nfd-topology-updater. NFD Topology Updater runs resource hardware topology detection normally, but no CR requests are sent to nfd-master.

Default: **false**

**Example**

```
$ nfd-topology-updater -no-publish
```

### 3.7.2.1. -oneshot

The **-oneshot** flag causes the NFD Topology Updater to exit after one pass of resource hardware topology detection.

Default: **false**

#### Example

```
$ nfd-topology-updater -oneshot -no-publish
```

#### -podresources-socket

The **-podresources-socket** flag specifies the path to the Unix socket where kubelet exports a gRPC service to enable discovery of in-use CPUs and devices, and to provide metadata for them.

Default: **/host-var/liblib/kubelet/pod-resources/kubelet.sock**

#### Example

```
$ nfd-topology-updater -podresources-socket=/var/lib/kubelet/pod-resources/kubelet.sock
```

#### -server

The **-server** flag specifies the address of the nfd-master endpoint to connect to.

Default: **localhost:8080**

#### Example

```
$ nfd-topology-updater -server=nfd-master.nfd.svc.cluster.local:443
```

#### -server-name-override

The **-server-name-override** flag specifies the common name (CN) which to expect from the nfd-master TLS certificate. This flag is mostly intended for development and debugging purposes.

Default: empty

#### Example

```
$ nfd-topology-updater -server-name-override=localhost
```

#### -sleep-interval

The **-sleep-interval** flag specifies the interval between resource hardware topology re-examination and custom resource updates. A non-positive value implies infinite sleep interval and no re-detection is done.

Default: **60s**

#### Example

```
$ nfd-topology-updater -sleep-interval=1h
```

**-version**

Print version and exit.

**-watch-namespace**

The **-watch-namespace** flag specifies the namespace to ensure that resource hardware topology examination only happens for the pods running in the specified namespace. Pods that are not running in the specified namespace are not considered during resource accounting. This is particularly useful for testing and debugging purposes. A \* value means that all of the pods across all namespaces are considered during the accounting process.

Default: \*

**Example**

```
$ nfd-topology-updater -watch-namespace=rte
```



## CHAPTER 4. KERNEL MODULE MANAGEMENT OPERATOR

Learn about the Kernel Module Management (KMM) Operator and how you can use it to deploy out-of-tree kernel modules and device plugins on OpenShift Container Platform clusters.

### 4.1. ABOUT THE KERNEL MODULE MANAGEMENT OPERATOR

The Kernel Module Management (KMM) Operator manages, builds, signs, and deploys out-of-tree kernel modules and device plugins on OpenShift Container Platform clusters.

KMM adds a new **Module** CRD which describes an out-of-tree kernel module and its associated device plugin. You can use **Module** resources to configure how to load the module, define **ModuleLoader** images for kernel versions, and include instructions for building and signing modules for specific kernel versions.

KMM is designed to accommodate multiple kernel versions at once for any kernel module, allowing for seamless node upgrades and reduced application downtime.

### 4.2. INSTALLING THE KERNEL MODULE MANAGEMENT OPERATOR

As a cluster administrator, you can install the Kernel Module Management (KMM) Operator by using the OpenShift CLI or the web console.

The KMM Operator is supported on OpenShift Container Platform 4.12 and later. Installing KMM on version 4.11 does not require specific additional steps. For details on installing KMM on version 4.10 and earlier, see the section "Installing the Kernel Module Management Operator on earlier versions of OpenShift Container Platform".

#### 4.2.1. Installing the Kernel Module Management Operator using the web console

As a cluster administrator, you can install the Kernel Module Management (KMM) Operator using the OpenShift Container Platform web console.

##### Procedure

1. Log in to the OpenShift Container Platform web console.
2. Install the Kernel Module Management Operator:
  - a. In the OpenShift Container Platform web console, click **Operators** → **OperatorHub**.
  - b. Select **Kernel Module Management Operator** from the list of available Operators, and then click **Install**.
  - c. From the **Installed Namespace** list, select the **openshift-kmm** namespace.
  - d. Click **Install**.

##### Verification

To verify that KMM Operator installed successfully:

1. Navigate to the **Operators** → **Installed Operators** page.

2. Ensure that **Kernel Module Management Operator** is listed in the **openshift-kmm** project with a **Status** of **InstallSucceeded**.



#### NOTE

During installation, an Operator might display a **Failed** status. If the installation later succeeds with an **InstallSucceeded** message, you can ignore the **Failed** message.

### Troubleshooting

1. To troubleshoot issues with Operator installation:
  - a. Navigate to the **Operators** → **Installed Operators** page and inspect the **Operator Subscriptions** and **Install Plans** tabs for any failure or errors under **Status**.
  - b. Navigate to the **Workloads** → **Pods** page and check the logs for pods in the **openshift-kmm** project.

### 4.2.2. Installing the Kernel Module Management Operator by using the CLI

As a cluster administrator, you can install the Kernel Module Management (KMM) Operator by using the OpenShift CLI.

#### Prerequisites

- You have a running OpenShift Container Platform cluster.
- You installed the OpenShift CLI (**oc**).
- You are logged into the OpenShift CLI as a user with **cluster-admin** privileges.

#### Procedure

1. Install KMM in the **openshift-kmm** namespace:
  - a. Create the following **Namespace** CR and save the YAML file, for example, **kmm-namespace.yaml**:
 

```
apiVersion: v1
kind: Namespace
metadata:
  name: openshift-kmm
```
  - b. Create the following **OperatorGroup** CR and save the YAML file, for example, **kmm-op-group.yaml**:
 

```
apiVersion: operators.coreos.com/v1
kind: OperatorGroup
metadata:
  name: kernel-module-management
  namespace: openshift-kmm
```

- c. Create the following **Subscription** CR and save the YAML file, for example, **kmm-sub.yaml**:

```

apiVersion: operators.coreos.com/v1alpha1
kind: Subscription
metadata:
  name: kernel-module-management
  namespace: openshift-kmm
spec:
  channel: release-1.0
  installPlanApproval: Automatic
  name: kernel-module-management
  source: redhat-operators
  sourceNamespace: openshift-marketplace
  startingCSV: kernel-module-management.v1.0.0

```

- d. Create the subscription object by running the following command:

```
$ oc create -f kmm-sub.yaml
```

### Verification

- To verify that the Operator deployment is successful, run the following command:

```
$ oc get -n openshift-kmm deployments.apps kmm-operator-controller
```

### Example output

NAME	READY	UP-TO-DATE	AVAILABLE	AGE
kmm-operator-controller	1/1	1	1	97s

The Operator is available.

### 4.2.3. Installing the Kernel Module Management Operator on earlier versions of OpenShift Container Platform

The KMM Operator is supported on OpenShift Container Platform 4.12 and later. For version 4.10 and earlier, you must create a new **SecurityContextConstraint** object and bind it to the Operator's **ServiceAccount**. As a cluster administrator, you can install the Kernel Module Management (KMM) Operator by using the OpenShift CLI.

#### Prerequisites

- You have a running OpenShift Container Platform cluster.
- You installed the OpenShift CLI (**oc**).
- You are logged into the OpenShift CLI as a user with **cluster-admin** privileges.

#### Procedure

- Install KMM in the **openshift-kmm** namespace:
  - Create the following **Namespace** CR and save the YAML file, for example, **kmm-namespace.yaml** file:

```

apiVersion: v1
kind: Namespace
metadata:
  name: openshift-kmm

```

- b. Create the following **SecurityContextConstraint** object and save the YAML file, for example, **kmm-security-constraint.yaml**:

```

allowHostDirVolumePlugin: false
allowHostIPC: false
allowHostNetwork: false
allowHostPID: false
allowHostPorts: false
allowPrivilegeEscalation: false
allowPrivilegedContainer: false
allowedCapabilities:
  - NET_BIND_SERVICE
apiVersion: security.openshift.io/v1
defaultAddCapabilities: null
fsGroup:
  type: MustRunAs
groups: []
kind: SecurityContextConstraints
metadata:
  name: restricted-v2
priority: null
readOnlyRootFilesystem: false
requiredDropCapabilities:
  - ALL
runAsUser:
  type: MustRunAsRange
seLinuxContext:
  type: MustRunAs
seccompProfiles:
  - runtime/default
supplementalGroups:
  type: RunAsAny
users: []
volumes:
  - configMap
  - downwardAPI
  - emptyDir
  - persistentVolumeClaim
  - projected
  - secret

```

- c. Bind the **SecurityContextConstraint** object to the Operator's **ServiceAccount** by running the following commands:

```
$ oc apply -f kmm-security-constraint.yaml
```

```
$ oc adm policy add-scc-to-user kmm-security-constraint -z kmm-operator-controller -n openshift-kmm
```

- d. Create the following **OperatorGroup** CR and save the YAML file, for example, **kmm-op-group.yaml**:

```
apiVersion: operators.coreos.com/v1
kind: OperatorGroup
metadata:
  name: kernel-module-management
  namespace: openshift-kmm
```

- e. Create the following **Subscription** CR and save the YAML file, for example, **kmm-sub.yaml**:

```
apiVersion: operators.coreos.com/v1alpha1
kind: Subscription
metadata:
  name: kernel-module-management
  namespace: openshift-kmm
spec:
  channel: release-1.0
  installPlanApproval: Automatic
  name: kernel-module-management
  source: redhat-operators
  sourceNamespace: openshift-marketplace
  startingCSV: kernel-module-management.v1.0.0
```

- f. Create the subscription object by running the following command:

```
$ oc create -f kmm-sub.yaml
```

## Verification

- To verify that the Operator deployment is successful, run the following command:

```
$ oc get -n openshift-kmm deployments.apps kmm-operator-controller
```

## Example output

```
NAME                                READY UP-TO-DATE AVAILABLE AGE
kmm-operator-controller             1/1 1 1 97s
```

The Operator is available.

## 4.3. CONFIGURING THE KERNEL MODULE MANAGEMENT OPERATOR

In most cases, the default configuration for the Kernel Module Management (KMM) Operator does not need to be modified. However, you can modify the Operator settings to suit your environment using the following procedure.

The Operator configuration is set in the **kmm-operator-manager-config ConfigMap** in the Operator namespace.

### Procedure

- To modify the settings, edit the **ConfigMap** data by entering the following command:

```
$ oc edit configmap -n "$namespace" kmm-operator-manager-config
```

### Example output

```
healthProbeBindAddress: :8081
job:
  gcDelay: 1h
leaderElection:
  enabled: true
  resourceID: kmm.sigs.x-k8s.io
webhook:
  disableHTTP2: true # CVE-2023-44487
  port: 9443
metrics:
  enableAuthnAuthz: true
  disableHTTP2: true # CVE-2023-44487
  bindAddress: 0.0.0.0:8443
  secureServing: true
worker:
  runAsUser: 0
  seLinuxType: spc_t
  setFirmwareClassPath: /var/lib/firmware
```

Table 4.1. Operator configuration parameters

Parameter	Description
<b>healthProbeBindAddress</b>	Defines the address on which the Operator monitors for kubelet health probes. The recommended value is <b>:8081</b> .
<b>job.gcDelay</b>	Defines the duration that successful build pods should be preserved for before they are deleted. There is no recommended value for this setting. For information about the valid values for this setting, see <a href="#">ParseDuration</a> .
<b>leaderElection.enabled</b>	Determines whether leader election is used to ensure that only one replica of the KMM Operator is running at any time. For more information, see <a href="#">Leases</a> . The recommended value is <b>true</b> .
<b>leaderElection.resourceID</b>	Determines the name of the resource that leader election uses for holding the leader lock. The recommended value is <b>kmm.sigs.x-k8s.io</b> .
<b>webhook.disableHTTP2</b>	If <b>true</b> , disables HTTP/2 for the webhook server, as a mitigation for <a href="#">cve-2023-44487</a> . The recommended value is <b>true</b> .
<b>webhook.port</b>	Defines the port on which the Operator monitors webhook requests. The recommended value is <b>9443</b> .

Parameter	Description
<b>metrics.enableAuthnAuthz</b>	<p>Determines if metrics are authenticated using <b>TokenReviews</b> and authorized using <b>SubjectAccessReviews</b> with the kube-apiserver.</p> <p>For authentication and authorization, the controller needs a <b>ClusterRole</b> with the following rules:</p> <ul style="list-style-type: none"> <li>• <b>apiGroups: authentication.k8s.io, resources: tokenreviews, verbs: create</b></li> <li>• <b>apiGroups: authorization.k8s.io, resources: subjectaccessreviews, verbs: create</b></li> </ul> <p>To scrape metrics, for example, using Prometheus, the client needs a <b>ClusterRole</b> with the following rule:</p> <ul style="list-style-type: none"> <li>• <b>nonResourceURLs: "/metrics", verbs: get</b></li> </ul> <p>The recommended value is <b>true</b>.</p>
<b>metrics.disableHTTP2</b>	<p>If <b>true</b>, disables HTTP/2 for the metrics server as a mitigation for <a href="#">CVE-2023-44487</a>. The recommended value is <b>true</b>.</p>
<b>metrics.bindAddress</b>	<p>Determines the bind address for the metrics server. If unspecified, the default is <b>:8080</b>. To disable the metrics server, set to <b>0</b>. The recommended value is <b>0.0.0.0:8443</b>.</p>
<b>metrics.secureServing</b>	<p>Determines whether the metrics are served over HTTPS instead of HTTP. The recommended value is <b>true</b>.</p>
<b>worker.runAsUser</b>	<p>Determines the value of the <b>runAsUser</b> field of the worker container's security context. For more information, see <a href="#">SecurityContext</a>. The recommended value is <b>9443</b>.</p>
<b>worker.seLinuxContextType</b>	<p>Determines the value of the <b>seLinuxOptions.type</b> field of the worker container's security context. For more information, see <a href="#">SecurityContext</a>. The recommended value is <b>spc_t</b>.</p>
<b>worker.setFirmwareClassPath</b>	<p>Sets the kernel's firmware search path into the <b>/sys/module/firmware_class/parameters/path</b> file on the node. The recommended value is <b>/var/lib/firmware</b> if you need to set that value through the worker app. Otherwise, unset.</p>

2. After modifying the settings, restart the controller with the following command:

```
$ oc delete pod -n "<namespace>" -l app.kubernetes.io/component=kmm
```



#### NOTE

The value of <namespace> depends on your original installation method.

## Additional resources

- For more information, see [Installing the Kernel Module Management Operator](#).

### 4.3.1. Unloading the kernel module

You must unload the kernel modules when moving to a newer version or if they introduce some undesirable side effect on the node.

#### Procedure

- To unload a module loaded with KMM from nodes, delete the corresponding **Module** resource. KMM then creates worker pods, where required, to run **modprobe -r** and unload the kernel module from the nodes.



#### WARNING

When unloading worker pods, KMM needs all the resources it uses when loading the kernel module. This includes the **ServiceAccount** referenced in the **Module** as well as any RBAC defined to allow privileged KMM worker Pods to run. It also includes any pull secret referenced in **.spec.imageRepoSecret**.

To avoid situations where KMM is unable to unload the kernel module from nodes, make sure those resources are not deleted while the **Module** resource is still present in the cluster in any state, including **Terminating**. KMM includes a validating admission webhook that rejects the deletion of namespaces that contain at least one **Module** resource.

### 4.3.2. Setting the kernel firmware search path

The Linux kernel accepts the **firmware\_class.path** parameter as a search path for firmware, as explained in [Firmware search paths](#).

KMM worker pods can set this value on nodes by writing to sysfs before attempting to load kmods.

#### Procedure

- To define a firmware search path, set **worker.setFirmwareClassPath** to **/var/lib/firmware** in the Operator configuration.

## Additional resources

- For more information about the **worker.setFirmwareClassPath** path, see [Configuring the Kernel Module Management Operator](#).

## 4.4. UNINSTALLING THE KERNEL MODULE MANAGEMENT OPERATOR



Use one of the following procedures to uninstall the Kernel Module Management (KMM) Operator, depending on how the KMM Operator was installed.

#### 4.4.1. Uninstalling a Red Hat catalog installation

Use this procedure if KMM was installed from the Red Hat catalog.

##### Procedure

Use the following method to uninstall the KMM Operator:

- Use the OpenShift console under **Operators** → **Installed Operators** to locate and uninstall the Operator.



##### NOTE

Alternatively, you can delete the **Subscription** resource in the KMM namespace.

#### 4.4.2. Uninstalling a CLI installation

Use this command if the KMM Operator was installed using the OpenShift CLI.

##### Procedure

- Run the following command to uninstall the KMM Operator:

```
$ oc delete -k https://github.com/rh-ecosystem-edge/kernel-module-management/config/default
```



##### NOTE

Using this command deletes the **Module** CRD and all **Module** instances in the cluster.

## 4.5. KERNEL MODULE DEPLOYMENT

Kernel Module Management (KMM) monitors **Node** and **Module** resources in the cluster to determine if a kernel module should be loaded on or unloaded from a node.

To be eligible for a module, a node must contain the following:

- Labels that match the module's **.spec.selector** field.
- A kernel version matching one of the items in the module's **.spec.moduleLoader.container.kernelMappings** field.
- If ordered upgrade (**ordered\_upgrade.md**) is configured in the module, a label that matches its **.spec.moduleLoader.container.version** field.

When KMM reconciles nodes with the desired state as configured in the **Module** resource, it creates worker pods on the target nodes to run the necessary action. The KMM Operator monitors the outcome of the pods and records the information. The Operator uses this information to label the **Node** objects when the module is successfully loaded, and to run the device plugin, if configured.

Worker pods run the KMM **worker** binary that performs the following tasks:

- Pulls the kmod image configured in the **Module** resource. Kmod images are standard OCI images that contain **.ko** files.
- Extracts the image in the pod's filesystem.
- Runs **modprobe** with the specified arguments to perform the necessary action.

#### 4.5.1. The Module custom resource definition

The **Module** custom resource definition (CRD) represents a kernel module that can be loaded on all or select nodes in the cluster, through a kmod image. A **Module** custom resource (CR) specifies one or more kernel versions with which it is compatible, and a node selector.

The compatible versions for a **Module** resource are listed under **.spec.moduleLoader.container.kernelMappings**. A kernel mapping can either match a **literal** version, or use **regexp** to match many of them at the same time.

The reconciliation loop for the **Module** resource runs the following steps:

1. List all nodes matching **.spec.selector**.
2. Build a set of all kernel versions running on those nodes.
3. For each kernel version:
  - a. Go through **.spec.moduleLoader.container.kernelMappings** and find the appropriate container image name. If the kernel mapping has **build** or **sign** defined and the container image does not already exist, run the build, the signing pod, or both, as needed.
  - b. Create a worker pod to pull the container image determined in the previous step and run **modprobe**.
  - c. If **.spec.devicePlugin** is defined, create a device plugin daemon set using the configuration specified under **.spec.devicePlugin.container**.
4. Run **garbage-collect** on:
  - a. Obsolete device plugin **DaemonSets** that do not target any node.
  - b. Successful build pods.
  - c. Successful signing pods.

#### 4.5.2. Set soft dependencies between kernel modules

Some configurations require that several kernel modules be loaded in a specific order to work properly, even though the modules do not directly depend on each other through symbols. These are called soft dependencies. **depmod** is usually not aware of these dependencies, and they do not appear in the files it produces. For example, if **mod\_a** has a soft dependency on **mod\_b**, **modprobe mod\_a** will not load **mod\_b**.

You can resolve these situations by declaring soft dependencies in the Module custom resource definition (CRD) using the **modulesLoadingOrder** field.

```
spec:
  moduleLoader:
    container:
      modprobe:
        moduleName: mod_a
        dirName: /opt
        firmwarePath: /firmware
        parameters:
          - param=1
      modulesLoadingOrder:
        - mod_a
        - mod_b
```

In the configuration above, the worker pod will first try to unload the in-tree **mod\_b** before loading **mod\_a** from the kmod image. When the worker pod is terminated and **mod\_a** is unloaded, **mod\_b** will not be loaded again.



#### NOTE

The first value in the list, to be loaded last, must be equivalent to the **moduleName**.

## 4.6. SECURITY AND PERMISSIONS



#### IMPORTANT

Loading kernel modules is a highly sensitive operation. After they are loaded, kernel modules have all possible permissions to do any kind of operation on the node.

### 4.6.1. ServiceAccounts and SecurityContextConstraints

Kernel Module Management (KMM) creates a privileged workload to load the kernel modules on nodes. That workload needs **ServiceAccounts** allowed to use the **privileged SecurityContextConstraint** (SCC) resource.

The authorization model for that workload depends on the namespace of the **Module** resource, as well as its spec.

- If the **.spec.moduleLoader.serviceAccountName** or **.spec.devicePlugin.serviceAccountName** fields are set, they are always used.
- If those fields are not set, then:
  - If the **Module** resource is created in the Operator's namespace ( **openshift-kmm** by default), then KMM uses its default, powerful **ServiceAccounts** to run the worker and device plugin pods.
  - If the **Module** resource is created in any other namespace, then KMM runs the pods with the namespace's **default ServiceAccount**. The **Module** resource cannot run a privileged workload unless you manually enable it to use the **privileged SCC**.



## IMPORTANT

**openshift-kmm** is a trusted namespace.

When setting up RBAC permissions, remember that any user or **ServiceAccount** creating a **Module** resource in the **openshift-kmm** namespace results in KMM automatically running privileged workloads on potentially all nodes in the cluster.

To allow any **ServiceAccount** to use the **privileged** SCC and run worker or device plugin pods, you can use the **oc adm policy** command, as in the following example:

```
$ oc adm policy add-scc-to-user privileged -z "${serviceAccountName}" [ -n "${namespace}" ]
```

### 4.6.2. Pod security standards

OpenShift runs a synchronization mechanism that sets the namespace Pod Security level automatically based on the security contexts in use. No action is needed.

#### Additional resources

- [Understanding and managing pod security admission](#)

## 4.7. REPLACING IN-TREE MODULES WITH OUT-OF-TREE MODULES

You can use Kernel Module Management (KMM) to build kernel modules that can be loaded or unloaded into the kernel on demand. These modules extend the functionality of the kernel without the need to reboot the system. Modules can be configured as built-in or dynamically loaded.

Dynamically loaded modules include in-tree modules and out-of-tree (OOT) modules. In-tree modules are internal to the Linux kernel tree, that is, they are already part of the kernel. Out-of-tree modules are external to the Linux kernel tree. They are generally written for development and testing purposes, such as testing the new version of a kernel module that is shipped in-tree, or to deal with incompatibilities.

Some modules that are loaded by KMM could replace in-tree modules that are already loaded on the node. To unload in-tree modules before loading your module, set the value of the **.spec.moduleLoader.container.inTreeModulesToRemove** field to the modules that you want to unload. The following example demonstrates module replacement for all kernel mappings:

```
# ...
spec:
  moduleLoader:
    container:
      modprobe:
        moduleName: mod_a

    inTreeModulesToRemove: [mod_a, mod_b]
```

In this example, the **moduleLoader** pod uses **inTreeModulesToRemove** to unload the in-tree **mod\_a** and **mod\_b** before loading **mod\_a** from the **moduleLoader** image. When the **moduleLoader** pod is terminated and **mod\_a** is unloaded, **mod\_b** is not loaded again.

The following is an example for module replacement for specific kernel mappings:

```
# ...
```

```
spec:
  moduleLoader:
    container:
      kernelMappings:
        - literal: 6.0.15-300.fc37.x86_64
          containerImage: "some.registry/org/my-kmod:${KERNEL_FULL_VERSION}"
          inTreeModulesToRemove: [<module_name>, <module_name>]
```

## Additional resources

- [Building a linux kernel module](#)

### 4.7.1. Example Module CR

The following is an annotated **Module** example:

```
apiVersion: kmm.sigs.x-k8s.io/v1beta1
kind: Module
metadata:
  name: <my_kmod>
spec:
  moduleLoader:
    container:
      modprobe:
        moduleName: <my_kmod> 1
        dirName: /opt 2
        firmwarePath: /firmware 3
        parameters: 4
        - param=1
      kernelMappings: 5
        - literal: 6.0.15-300.fc37.x86_64
          containerImage: some.registry/org/my-kmod:6.0.15-300.fc37.x86_64
        - regexp: '^.+\\fc37\\.x86_64$' 6
          containerImage: "some.other.registry/org/<my_kmod>:${KERNEL_FULL_VERSION}"
        - regexp: '^.+$$' 7
          containerImage: "some.registry/org/<my_kmod>:${KERNEL_FULL_VERSION}"
      build:
        buildArgs: 8
        - name: ARG_NAME
          value: <some_value>
        secrets:
        - name: <some_kubernetes_secret> 9
      baseImageRegistryTLS: 10
        insecure: false
        insecureSkipTLSVerify: false 11
      dockerfileConfigMap: 12
        name: <my_kmod_dockerfile>
      sign:
        certSecret:
        name: <cert_secret> 13
        keySecret:
        name: <key_secret> 14
      filesToSign:
```

```

- /opt/lib/modules/${KERNEL_FULL_VERSION}/<my_kmod>.ko
registryTLS: 15
  insecure: false 16
  insecureSkipTLSVerify: false
serviceAccountName: <sa_module_loader> 17
devicePlugin: 18
container:
  image: some.registry/org/device-plugin:latest 19
  env:
  - name: MY_DEVICE_PLUGIN_ENV_VAR
    value: SOME_VALUE
  volumeMounts: 20
  - mountPath: /some/mountPath
    name: <device_plugin_volume>
  volumes: 21
  - name: <device_plugin_volume>
  configMap:
    name: <some_configmap>
  serviceAccountName: <sa_device_plugin> 22
imageRepoSecret: 23
  name: <secret_name>
selector:
  node-role.kubernetes.io/worker: ""

```

1 1 1 Required.

2 Optional.

3 Optional: Copies **/firmware/\*** into **/var/lib/firmware/** on the node.

4 Optional.

5 At least one kernel item is required.

6 For each node running a kernel matching the regular expression, KMM creates a **DaemonSet** resource running the image specified in **containerImage** with **\${KERNEL\_FULL\_VERSION}** replaced with the kernel version.

7 For any other kernel, build the image using the Dockerfile in the **my-kmod** ConfigMap.

8 Optional.

9 Optional: A value for **some-kubernetes-secret** can be obtained from the build environment at **/run/secrets/some-kubernetes-secret**.

10 This field has no effect. When building kmod images or signing kmods within a kmod image, you might sometimes need to pull base images from a registry that serves a certificate signed by an untrusted Certificate Authority (CA). In order for KMM to trust that CA, it must also trust the new CA by replacing the cluster's CA bundle.

See "Additional resources" to learn how to replace the cluster's CA bundle.

11 Optional: Avoid using this parameter. If set to **true**, the build will skip any TLS server certificate validation when pulling the image in the Dockerfile **FROM** instruction using plain HTTP.

- 12 Required.
- 13 Required: A secret holding the public secureboot key with the key 'cert'.
- 14 Required: A secret holding the private secureboot key with the key 'key'.
- 15 Optional: Avoid using this parameter. If set to **true**, KMM will be allowed to check if the container image already exists using plain HTTP.
- 16 Optional: Avoid using this parameter. If set to **true**, KMM will skip any TLS server certificate validation when checking if the container image already exists.
- 17 Optional.
- 18 Optional.
- 19 Required: If the device plugin section is present.
- 20 Optional.
- 21 Optional.
- 22 Optional.
- 23 Optional: Used to pull module loader and device plugin images.

#### Additional resources

- [Replacing the CA Bundle certificate](#)

## 4.8. SYMBOLIC LINKS FOR IN-TREE DEPENDENCIES

Some kernel modules depend on other kernel modules that are shipped with the node's operating system. To avoid copying those dependencies into the kmod image, Kernel Module Management (KMM) mounts **/usr/lib/modules** into both the build and the worker pod's filesystems.

By creating a symlink from **/opt/usr/lib/modules/<kernel\_version>/<symlink\_name>** to **/usr/lib/modules/<kernel\_version>**, **depmod** can use the in-tree kmods on the building node's filesystem to resolve dependencies.

At runtime, the worker pod extracts the entire image, including the **<symlink\_name>** symbolic link. That symbolic link points to **/usr/lib/modules/<kernel\_version>** in the worker pod, which is mounted from the node's filesystem. **modprobe** can then follow that link and load the in-tree dependencies as needed.

In the following example, **host** is the symbolic link name under **/opt/usr/lib/modules/<kernel\_version>**:

```
ARG DTK_AUTO
FROM ${DTK_AUTO} as builder
#
# Build steps
#
FROM ubi9/ubi
```

```
ARG KERNEL_FULL_VERSION
```

```
RUN dnf update && dnf install -y kmod
```

```
COPY --from=builder /usr/src/kernel-module-management/ci/kmm-kmod/kmm_ci_a.ko  
/opt/lib/modules/${KERNEL_FULL_VERSION}/
```

```
COPY --from=builder /usr/src/kernel-module-management/ci/kmm-kmod/kmm_ci_b.ko  
/opt/lib/modules/${KERNEL_FULL_VERSION}/
```

```
# Create the symbolic link
```

```
RUN ln -s /lib/modules/${KERNEL_FULL_VERSION}  
/opt/lib/modules/${KERNEL_FULL_VERSION}/host
```

```
RUN depmod -b /opt ${KERNEL_FULL_VERSION}
```



## NOTE

**depmod** generates dependency files based on the kernel modules present on the node that runs the kmod image build.

On the node on which KMM loads the kernel modules, **modprobe** expects the files to be present under `/usr/lib/modules/<kernel_version>`, and the same filesystem layout. It is highly recommended that the build and the target nodes share the same operating system and release.

## 4.9. CREATING A KMOD IMAGE

Kernel Module Management (KMM) works with purpose-built kmod images, which are standard OCI images that contain `.ko` files. The location of the `.ko` files must match the following pattern: `<prefix>/lib/modules/[kernel-version]/`.

Keep the following in mind when working with the `.ko` files:

- In most cases, `<prefix>` should be equal to `/opt`. This is the **Module** CRD's default value.
- **kernel-version** must not be empty and must be equal to the kernel version the kernel modules were built for.

### 4.9.1. Running depmod

It is recommended to run **depmod** at the end of the build process to generate `modules.dep` and `.map` files. This is especially useful if your kmod image contains several kernel modules and if one of the modules depends on another module.



## NOTE

You must have a Red Hat subscription to download the **kernel-devel** package.

### Procedure

- Generate `modules.dep` and `.map` files for a specific kernel version by running the following command:



```
$ depmod -b /opt ${KERNEL_FULL_VERSION}+`.
```

#### 4.9.1.1. Example Dockerfile

If you are building your image on OpenShift Container Platform, consider using the Driver Tool Kit (DTK).

For further information, see [using an entitled build](#).

```
apiVersion: v1
kind: ConfigMap
metadata:
  name: kmm-ci-dockerfile
data:
  dockerfile: |
    ARG DTK_AUTO
    FROM ${DTK_AUTO} as builder
    ARG KERNEL_FULL_VERSION
    WORKDIR /usr/src
    RUN ["git", "clone", "https://github.com/rh-ecosystem-edge/kernel-module-management.git"]
    WORKDIR /usr/src/kernel-module-management/ci/kmm-kmod
    RUN KERNEL_SRC_DIR=/lib/modules/${KERNEL_FULL_VERSION}/build make all
    FROM registry.redhat.io/ubi9/ubi-minimal
    ARG KERNEL_FULL_VERSION
    RUN microdnf install kmod
    COPY --from=builder /usr/src/kernel-module-management/ci/kmm-kmod/kmm_ci_a.ko
    /opt/lib/modules/${KERNEL_FULL_VERSION}/
    COPY --from=builder /usr/src/kernel-module-management/ci/kmm-kmod/kmm_ci_b.ko
    /opt/lib/modules/${KERNEL_FULL_VERSION}/
    RUN depmod -b /opt ${KERNEL_FULL_VERSION}
```

#### Additional resources

- [Driver Toolkit](#)

#### 4.9.2. Building in the cluster

KMM can build kmod images in the cluster. Follow these guidelines:

- Provide build instructions using the **build** section of a kernel mapping.
- Copy the **Dockerfile** for your container image into a **ConfigMap** resource, under the **dockerfile** key.
- Ensure that the **ConfigMap** is located in the same namespace as the **Module**.

KMM checks if the image name specified in the **containerImage** field exists. If it does, the build is skipped.

Otherwise, KMM creates a **Build** resource to build your image. After the image is built, KMM proceeds with the **Module** reconciliation. See the following example.

```
# ...
- regexp: '^.+${'
```

```

containerImage: "some.registry/org/<my_kmod>:${KERNEL_FULL_VERSION}"
build:
  buildArgs: ❶
    - name: ARG_NAME
      value: <some_value>
  secrets: ❷
    - name: <some_kubernetes_secret> ❸
  baseImageRegistryTLS:
    insecure: false ❹
    insecureSkipTLSVerify: false ❺
  dockerfileConfigMap: ❻
    name: <my_kmod_dockerfile>
  registryTLS:
    insecure: false ❼
    insecureSkipTLSVerify: false ❽

```

- ❶ Optional.
- ❷ Optional.
- ❸ Will be mounted in the build pod as **/run/secrets/some-kubernetes-secret**.
- ❹ Optional: Avoid using this parameter. If set to **true**, the build will be allowed to pull the image in the Dockerfile **FROM** instruction using plain HTTP.
- ❺ Optional: Avoid using this parameter. If set to **true**, the build will skip any TLS server certificate validation when pulling the image in the Dockerfile **FROM** instruction using plain HTTP.
- ❻ Required.
- ❼ Optional: Avoid using this parameter. If set to **true**, KMM will be allowed to check if the container image already exists using plain HTTP.
- ❽ Optional: Avoid using this parameter. If set to **true**, KMM will skip any TLS server certificate validation when checking if the container image already exists.

Successful build pods are garbage collected immediately, unless the **job.gcDelay** parameter is set in the Operator configuration. Failed build pods are always preserved and must be deleted manually by the administrator for the build to be restarted.

#### Additional resources

- [Build configuration resources](#)
- [Preflight validation for Kernel Module Management \(KMM\) Modules](#)

### 4.9.3. Using the Driver Toolkit

The Driver Toolkit (DTK) is a convenient base image for building build kmod loader images. It contains tools and libraries for the OpenShift version currently running in the cluster.

#### Procedure

Use DTK as the first stage of a multi-stage **Dockerfile**.

1. Build the kernel modules.
2. Copy the **.ko** files into a smaller end-user image such as [ubi-minimal](#).
3. To leverage DTK in your in-cluster build, use the **DTK\_AUTO** build argument. The value is automatically set by KMM when creating the **Build** resource. See the following example.

```
ARG DTK_AUTO
FROM ${DTK_AUTO} as builder
ARG KERNEL_FULL_VERSION
WORKDIR /usr/src
RUN ["git", "clone", "https://github.com/rh-ecosystem-edge/kernel-module-management.git"]
WORKDIR /usr/src/kernel-module-management/ci/kmm-kmod
RUN KERNEL_SRC_DIR=/lib/modules/${KERNEL_FULL_VERSION}/build make all
FROM ubi9/ubi-minimal
ARG KERNEL_FULL_VERSION
RUN microdnf install kmod
COPY --from=builder /usr/src/kernel-module-management/ci/kmm-kmod/kmm_ci_a.ko
/opt/lib/modules/${KERNEL_FULL_VERSION}/
COPY --from=builder /usr/src/kernel-module-management/ci/kmm-kmod/kmm_ci_b.ko
/opt/lib/modules/${KERNEL_FULL_VERSION}/
RUN depmod -b /opt ${KERNEL_FULL_VERSION}
```

#### Additional resources

- [Driver Toolkit](#)

## 4.10. USING SIGNING WITH KERNEL MODULE MANAGEMENT (KMM)

On a Secure Boot enabled system, all kernel modules (kmods) must be signed with a public/private key-pair enrolled into the Machine Owner's Key (MOK) database. Drivers distributed as part of a distribution should already be signed by the distribution's private key, but for kernel modules build out-of-tree, KMM supports signing kernel modules using the **sign** section of the kernel mapping.

For more details on using Secure Boot, see [Generating a public and private key pair](#)

#### Prerequisites

- A public private key pair in the correct (DER) format.
- At least one secure-boot enabled node with the public key enrolled in its MOK database.
- Either a pre-built driver container image, or the source code and **Dockerfile** needed to build one in-cluster.

## 4.11. ADDING THE KEYS FOR SECUREBOOT

To use KMM Kernel Module Management (KMM) to sign kernel modules, a certificate and private key are required. For details on how to create these, see [Generating a public and private key pair](#).

For details on how to extract the public and private key pair, see [Signing kernel modules with the private key](#). Use steps 1 through 4 to extract the keys into files.

#### Procedure

1. Create the **sb\_cert.cer** file that contains the certificate and the **sb\_cert.priv** file that contains the private key:

```
$ openssl req -x509 -new -nodes -utf8 -sha256 -days 36500 -batch -config
configuration_file.config -outform DER -out my_signing_key_pub.der -keyout
my_signing_key.priv
```

2. Add the files by using one of the following methods:

- Add the files as [secrets](#) directly:

```
$ oc create secret generic my-signing-key --from-file=key=<my_signing_key.priv>
```

```
$ oc create secret generic my-signing-key-pub --from-file=cert=
<my_signing_key_pub.der>
```

- Add the files by base64 encoding them:

```
$ cat sb_cert.priv | base64 -w 0 > my_signing_key2.base64
```

```
$ cat sb_cert.cer | base64 -w 0 > my_signing_key_pub.base64
```

3. Add the encoded text to a YAML file:

```
apiVersion: v1
kind: Secret
metadata:
  name: my-signing-key-pub
  namespace: default 1
type: Opaque
data:
  cert: <base64_encoded_secureboot_public_key>

---
apiVersion: v1
kind: Secret
metadata:
  name: my-signing-key
  namespace: default 2
type: Opaque
data:
  key: <base64_encoded_secureboot_private_key>
```

**1 2 namespace** - Replace **default** with a valid namespace.

4. Apply the YAML file:

```
$ oc apply -f <yaml_filename>
```

#### 4.11.1. Checking the keys

After you have added the keys, you must check them to ensure they are set correctly.

## Procedure

1. Check to ensure the public key secret is set correctly:

```
$ oc get secret -o yaml <certificate secret name> | awk '/cert/{print $2; exit}' | base64 -d |
openssl x509 -inform der -text
```

This should display a certificate with a Serial Number, Issuer, Subject, and more.

2. Check to ensure the private key secret is set correctly:

```
$ oc get secret -o yaml <private key secret name> | awk '/key/{print $2; exit}' | base64 -d
```

This should display the key enclosed in the **-----BEGIN PRIVATE KEY-----** and **-----END PRIVATE KEY-----** lines.

## 4.12. SIGNING KMODS IN A PRE-BUILT IMAGE

Use this procedure if you have a pre-built image, such as an image either distributed by a hardware vendor or built elsewhere.

The following YAML file adds the public/private key-pair as secrets with the required key names - **key** for the private key, **cert** for the public key. The cluster then pulls down the **unsignedImage** image, opens it, signs the kernel modules listed in **filesToSign**, adds them back, and pushes the resulting image as **containerImage**.

KMM then loads the signed kmods onto all the nodes with that match the selector. The kmods are successfully loaded on any nodes that have the public key in their MOK database, and any nodes that are not secure-boot enabled, which will ignore the signature.

### Prerequisites

- The **keySecret** and **certSecret** secrets have been created in the same namespace as the rest of the resources.

### Procedure

- Apply the YAML file:

```
---
apiVersion: kmm.sigs.x-k8s.io/v1beta1
kind: Module
metadata:
  name: example-module
spec:
  moduleLoader:
    serviceAccountName: default
  container:
    modprobe: 1
      moduleName: '<module_name>'
  kernelMappings:
    # the kmods will be deployed on all nodes in the cluster with a kernel that matches the
    # regexp
    - regexp: '^.*\x86_64$'
      # the container to produce containing the signed kmods
```

```

    containerImage: <image_name> ❷
    sign:
      # the image containing the unsigned kmods (we need this because we are not
      # building the kmods within the cluster)
      unsignedImage: <image_name> ❸
      keySecret: # a secret holding the private secureboot key with the key 'key'
        name: <private_key_secret_name>
      certSecret: # a secret holding the public secureboot key with the key 'cert'
        name: <certificate_secret_name>
      filesToSign: # full path within the unsignedImage container to the kmod(s) to sign
        - /opt/lib/modules/4.18.0-348.2.1.el8_5.x86_64/kmm_ci_a.ko
    imageRepoSecret:
      # the name of a secret containing credentials to pull unsignedImage and push
      # containerImage to the registry
      name: repo-pull-secret
      selector:
        kubernetes.io/arch: amd64

```

- ❶ The name of the kmod to load.
- ❷ The name of the container image. For example, **quay.io/myuser/my-driver:<kernelversion>**.
- ❸ The name of the unsigned image. For example, **quay.io/myuser/my-driver:<kernelversion>**.

## 4.13. BUILDING AND SIGNING A KMOD IMAGE

Use this procedure if you have source code and must build your image first.

The following YAML file builds a new container image using the source code from the repository. The image produced is saved back in the registry with a temporary name, and this temporary image is then signed using the parameters in the **sign** section.

The temporary image name is based on the final image name and is set to be **<containerImage>:<tag>-<namespace>\_<module name>\_kmm\_unsigned**.

For example, using the following YAML file, Kernel Module Management (KMM) builds an image named **example.org/repository/minimal-driver:final-default\_example-module\_kmm\_unsigned** containing the build with unsigned kmods and pushes it to the registry. Then it creates a second image named **example.org/repository/minimal-driver:final** that contains the signed kmods. It is this second image that is pulled by the worker pods and contains the kmods to be loaded on the cluster nodes.

After it is signed, you can safely delete the temporary image from the registry. It will be rebuilt, if needed.

### Prerequisites

- The **keySecret** and **certSecret** secrets have been created in the same namespace as the rest of the resources.

### Procedure

- Apply the YAML file:

```

---
apiVersion: v1

```

```

kind: ConfigMap
metadata:
  name: example-module-dockerfile
  namespace: <namespace> 1
data:
  Dockerfile: |
    ARG DTK_AUTO
    ARG KERNEL_VERSION
    FROM ${DTK_AUTO} as builder
    WORKDIR /build/
    RUN git clone -b main --single-branch https://github.com/rh-ecosystem-edge/kernel-
module-management.git
    WORKDIR kernel-module-management/ci/kmm-kmod/
    RUN make
    FROM registry.access.redhat.com/ubi9/ubi:latest
    ARG KERNEL_VERSION
    RUN yum -y install kmod && yum clean all
    RUN mkdir -p /opt/lib/modules/${KERNEL_VERSION}
    COPY --from=builder /build/kernel-module-management/ci/kmm-kmod/*.ko
/opt/lib/modules/${KERNEL_VERSION}/
    RUN /usr/sbin/depmod -b /opt
---
apiVersion: kmm.sigs.x-k8s.io/v1beta1
kind: Module
metadata:
  name: example-module
  namespace: <namespace> 2
spec:
  moduleLoader:
    serviceAccountName: default 3
  container:
    modprobe:
      moduleName: simple_kmod
    kernelMappings:
      - regexp: '^.*\x86_64$'
        containerImage: <final_driver_container_name>
        build:
          dockerfileConfigMap:
            name: example-module-dockerfile
      sign:
        keySecret:
          name: <private_key_secret_name>
        certSecret:
          name: <certificate_secret_name>
        filesToSign:
          - /opt/lib/modules/4.18.0-348.2.1.el8_5.x86_64/kmm_ci_a.ko
  imageRepoSecret: 4
    name: repo-pull-secret
    selector: # top-level selector
    kubernetes.io/arch: amd64

```

**1** **2** Replace **default** with a valid namespace.

**3**

The default **serviceAccountName** does not have the required permissions to run a module that is privileged. For information on creating a service account, see "Creating service accounts" in the

- Used as **imagePullSecrets** in the **DaemonSet** object and to pull and push for the build and sign features.

#### Additional resources

- [Creating service accounts](#).

## 4.14. KMM HUB AND SPOKE

In hub and spoke scenarios, many spoke clusters are connected to a central, powerful hub cluster. Kernel Module Management (KMM) depends on Red Hat Advanced Cluster Management (RHACM) to operate in hub and spoke environments.

KMM is compatible with hub and spoke environments through decoupling KMM features. A **ManagedClusterModule** custom resource definition (CRD) is provided to wrap the existing **Module** CRD and extend it to select Spoke clusters. Also provided is KMM-Hub, a new standalone controller that builds images and signs modules on the hub cluster.

In hub and spoke setups, spokes are focused, resource-constrained clusters that are centrally managed by a hub cluster. Spokes run the single-cluster edition of KMM, with those resource-intensive features disabled. To adapt KMM to this environment, you should reduce the workload running on the spokes to the minimum, while the hub takes care of the expensive tasks.

Building kernel module images and signing the **.ko** files, should run on the hub. The scheduling of the Module Loader and Device Plugin **DaemonSets** can only happen on the spokes.

#### Additional resources

- [Red Hat Advanced Cluster Management \(RHACM\)](#)

### 4.14.1. KMM-Hub

The KMM project provides KMM-Hub, an edition of KMM dedicated to hub clusters. KMM-Hub monitors all kernel versions running on the spokes and determines the nodes on the cluster that should receive a kernel module.

KMM-Hub runs all compute-intensive tasks such as image builds and kmod signing, and prepares the trimmed-down **Module** to be transferred to the spokes through RHACM.



#### NOTE

KMM-Hub cannot be used to load kernel modules on the hub cluster. Install the regular edition of KMM to load kernel modules.

#### Additional resources

- [Installing KMM](#)

### 4.14.2. Installing KMM-Hub



You can use one of the following methods to install KMM-Hub:

- With the Operator Lifecycle Manager (OLM)
- Creating KMM resources

#### Additional resources

- [KMM Operator bundle](#)

#### 4.14.2.1. Installing KMM-Hub using the Operator Lifecycle Manager

Use the **Operators** section of the OpenShift console to install KMM-Hub.

#### 4.14.2.2. Installing KMM-Hub by creating KMM resources

##### Procedure

- If you want to install KMM-Hub programmatically, you can use the following resources to create the **Namespace**, **OperatorGroup** and **Subscription** resources:

```
---
apiVersion: v1
kind: Namespace
metadata:
  name: openshift-kmm-hub
---
apiVersion: operators.coreos.com/v1
kind: OperatorGroup
metadata:
  name: kernel-module-management-hub
  namespace: openshift-kmm-hub
---
apiVersion: operators.coreos.com/v1alpha1
kind: Subscription
metadata:
  name: kernel-module-management-hub
  namespace: openshift-kmm-hub
spec:
  channel: stable
  installPlanApproval: Automatic
  name: kernel-module-management-hub
  source: redhat-operators
  sourceNamespace: openshift-marketplace
```

#### 4.14.3. Using the ManagedClusterModule CRD

Use the **ManagedClusterModule** Custom Resource Definition (CRD) to configure the deployment of kernel modules on spoke clusters. This CRD is cluster-scoped, wraps a **Module** spec and adds the following additional fields:

```
apiVersion: hub.kmm.sigs.x-k8s.io/v1beta1
kind: ManagedClusterModule
metadata:
```

```

name: <my-mcm>
# No namespace, because this resource is cluster-scoped.
spec:
  moduleSpec: ❶
  selector: ❷
    node-wants-my-mcm: 'true'

  spokeNamespace: <some-namespace> ❸

  selector: ❹
    wants-my-mcm: 'true'

```

- ❶ **moduleSpec**: Contains **moduleLoader** and **devicePlugin** sections, similar to a **Module** resource.
- ❷ Selects nodes within the **ManagedCluster**.
- ❸ Specifies in which namespace the **Module** should be created.
- ❹ Selects **ManagedCluster** objects.

If build or signing instructions are present in **.spec.moduleSpec**, those pods are run on the hub cluster in the operator's namespace.

When the **.spec.selector matches** one or more **ManagedCluster** resources, then KMM-Hub creates a **ManifestWork** resource in the corresponding namespace(s). **ManifestWork** contains a trimmed-down **Module** resource, with kernel mappings preserved but all **build** and **sign** subsections are removed. **containerImage** fields that contain image names ending with a tag are replaced with their digest equivalent.

#### 4.14.4. Running KMM on the spoke

After installing Kernel Module Management (KMM) on the spoke, no further action is required. Create a **ManagedClusterModule** object from the hub to deploy kernel modules on spoke clusters.

##### Procedure

You can install KMM on the spokes cluster through a RHACM **Policy** object. In addition to installing KMM from the OperatorHub and running it in a lightweight spoke mode, the **Policy** configures additional RBAC required for the RHACM agent to be able to manage **Module** resources.

- Use the following RHACM policy to install KMM on spoke clusters:

```

---
apiVersion: policy.open-cluster-management.io/v1
kind: Policy
metadata:
  name: install-kmm
spec:
  remediationAction: enforce
  disabled: false
  policy-templates:
  - objectDefinition:
      apiVersion: policy.open-cluster-management.io/v1
      kind: ConfigurationPolicy
      metadata:

```

```

name: install-kmm
spec:
severity: high
object-templates:
- complianceType: mustonlyhave
  objectDefinition:
    apiVersion: v1
    kind: Namespace
    metadata:
      name: openshift-kmm
- complianceType: mustonlyhave
  objectDefinition:
    apiVersion: operators.coreos.com/v1
    kind: OperatorGroup
    metadata:
      name: kmm
      namespace: openshift-kmm
    spec:
      upgradeStrategy: Default
- complianceType: mustonlyhave
  objectDefinition:
    apiVersion: operators.coreos.com/v1alpha1
    kind: Subscription
    metadata:
      name: kernel-module-management
      namespace: openshift-kmm
    spec:
      channel: stable
      config:
        env:
          - name: KMM_MANAGED 1
            value: "1"
        installPlanApproval: Automatic
        name: kernel-module-management
        source: redhat-operators
        sourceNamespace: openshift-marketplace
- complianceType: mustonlyhave
  objectDefinition:
    apiVersion: rbac.authorization.k8s.io/v1
    kind: ClusterRole
    metadata:
      name: kmm-module-manager
    rules:
      - apiGroups: [kmm.sigs.x-k8s.io]
        resources: [modules]
        verbs: [create, delete, get, list, patch, update, watch]
- complianceType: mustonlyhave
  objectDefinition:
    apiVersion: rbac.authorization.k8s.io/v1
    kind: ClusterRoleBinding
    metadata:
      name: klusterlet-kmm
    subjects:
      - kind: ServiceAccount
        name: klusterlet-work-sa
        namespace: open-cluster-management-agent

```

```

    roleRef:
      kind: ClusterRole
      name: kmm-module-manager
      apiGroup: rbac.authorization.k8s.io
  ---
  apiVersion: apps.open-cluster-management.io/v1
  kind: PlacementRule
  metadata:
    name: all-managed-clusters
  spec:
    clusterSelector: ❷
    matchExpressions: []
  ---
  apiVersion: policy.open-cluster-management.io/v1
  kind: PlacementBinding
  metadata:
    name: install-kmm
  placementRef:
    apiGroup: apps.open-cluster-management.io
    kind: PlacementRule
    name: all-managed-clusters
  subjects:
  - apiGroup: policy.open-cluster-management.io
    kind: Policy
    name: install-kmm

```

- ❶ This environment variable is required when running KMM on a spoke cluster.
- ❷ The **spec.clusterSelector** field can be customized to target select clusters only.

## 4.15. CUSTOMIZING UPGRADES FOR KERNEL MODULES

Use this procedure to upgrade the kernel module while running maintenance operations on the node, including rebooting the node, if needed. To minimize the impact on the workloads running in the cluster, run the kernel upgrade process sequentially, one node at a time.



### NOTE

This procedure requires knowledge of the workload utilizing the kernel module and must be managed by the cluster administrator.

### Prerequisites

- Before upgrading, set the **kmm.node.kubernetes.io/version-module.<module\_namespace>.<module\_name>=\$moduleVersion** label on all the nodes that are used by the kernel module.
- Terminate all user application workloads on the node or move them to another node.
- Unload the currently loaded kernel module.
- Ensure that the user workload (the application running in the cluster that is accessing kernel module) is not running on the node prior to kernel module unloading and that the workload is back running on the node after the new kernel module version has been loaded.

## Procedure

1. Ensure that the device plugin managed by KMM on the node is unloaded.
2. Update the following fields in the **Module** custom resource (CR):
  - **containerImage** (to the appropriate kernel version)
  - **version**  
The update should be atomic; that is, both the **containerImage** and **version** fields must be updated simultaneously.
3. Terminate any workload using the kernel module on the node being upgraded.
4. Remove the **kmm.node.kubernetes.io/version-module.<module\_namespace>.<module\_name>** label on the node. Run the following command to unload the kernel module from the node:

```
$ oc label node/<node_name> kmm.node.kubernetes.io/version-module.<module_namespace>.<module_name>-
```

5. If required, as the cluster administrator, perform any additional maintenance required on the node for the kernel module upgrade.  
If no additional upgrading is needed, you can skip Steps 3 through 6 by updating the **kmm.node.kubernetes.io/version-module.<module\_namespace>.<module\_name>** label value to the new **\$moduleVersion** as set in the **Module**.
6. Run the following command to add the **kmm.node.kubernetes.io/version-module.<module\_namespace>.<module\_name>=\$moduleVersion** label to the node. The **\$moduleVersion** must be equal to the new value of the **version** field in the **Module** CR.

```
$ oc label node/<node_name> kmm.node.kubernetes.io/version-module.<module_namespace>.<module_name>=<desired_version>
```



### NOTE

Because of Kubernetes limitations in label names, the combined length of **Module** name and namespace must not exceed 39 characters.

7. Restore any workload that leverages the kernel module on the node.
8. Reload the device plugin managed by KMM on the node.

## 4.16. DAY 1 KERNEL MODULE LOADING

Kernel Module Management (KMM) is typically a Day 2 Operator. Kernel modules are loaded only after the complete initialization of a Linux (RHCOS) server. However, in some scenarios the kernel module must be loaded at an earlier stage. Day 1 functionality allows you to use the Machine Config Operator (MCO) to load kernel modules during the Linux **systemd** initialization stage.

### Additional resources

- [Machine Config Operator](#)

### 4.16.1. Day 1 supported use cases

The Day 1 functionality supports a limited number of use cases. The main use case is to allow loading out-of-tree (OOT) kernel modules prior to NetworkManager service initialization. It does not support loading kernel module at the **initramfs** stage.

The following are the conditions needed for Day 1 functionality:

- The kernel module is not loaded in the kernel.
- The in-tree kernel module is loaded into the kernel, but can be unloaded and replaced by the OOT kernel module. This means that the in-tree module is not referenced by any other kernel modules.
- In order for Day 1 functionality to work, the node must have a functional network interface, that is, an in-tree kernel driver for that interface. The OOT kernel module can be a network driver that will replace the functional network driver.

### 4.16.2. OOT kernel module loading flow

The loading of the out-of-tree (OOT) kernel module leverages the Machine Config Operator (MCO). The flow sequence is as follows:

#### Procedure

1. Apply a **MachineConfig** resource to the existing running cluster. In order to identify the necessary nodes that need to be updated, you must create an appropriate **MachineConfigPool** resource.
2. MCO applies the reboots node by node. On any rebooted node, two new **systemd** services are deployed: **pull** service and **load** service.
3. The **load** service is configured to run prior to the **NetworkConfiguration** service. The service tries to pull a predefined kernel module image and then, using that image, to unload an in-tree module and load an OOT kernel module.
4. The **pull** service is configured to run after NetworkManager service. The service checks if the preconfigured kernel module image is located on the node's filesystem. If it is, the service exists normally, and the server continues with the boot process. If not, it pulls the image onto the node and reboots the node afterwards.

### 4.16.3. The kernel module image

The Day 1 functionality uses the same DTK based image leveraged by Day 2 KMM builds. The out-of-tree kernel module should be located under **/opt/lib/modules/\${kernelVersion}**.

#### Additional resources

- [Driver Toolkit](#)

### 4.16.4. In-tree module replacement

The Day 1 functionality always tries to replace the in-tree kernel module with the OOT version. If the in-tree kernel module is not loaded, the flow is not affected; the service proceeds and loads the OOT kernel module.

### 4.16.5. MCO yaml creation

KMM provides an API to create an MCO YAML manifest for the Day 1 functionality:

```
ProduceMachineConfig(machineConfigName, machineConfigPoolRef, kernelModuleImage,
kernelModuleName string) (string, error)
```

The returned output is a string representation of the MCO YAML manifest to be applied. It is up to the customer to apply this YAML.

The parameters are:

#### machineConfigName

The name of the MCO YAML manifest. This parameter is set as the **name** parameter of the metadata of the MCO YAML manifest.

#### machineConfigPoolRef

The **MachineConfigPool** name used to identify the targeted nodes.

#### kernelModuleImage

The name of the container image that includes the OOT kernel module.

#### kernelModuleName

The name of the OOT kernel module. This parameter is used both to unload the in-tree kernel module (if loaded into the kernel) and to load the OOT kernel module.

The API is located under **pkg/mcproducer** package of the KMM source code. The KMM operator does not need to be running to use the Day 1 functionality. You only need to import the **pkg/mcproducer** package into their operator/utility code, call the API, and apply the produced MCO YAML to the cluster.

### 4.16.6. The MachineConfigPool

The **MachineConfigPool** identifies a collection of nodes that are affected by the applied MCO.

```
kind: MachineConfigPool
metadata:
  name: sfc
spec:
  machineConfigSelector: ❶
  matchExpressions:
    - {key: machineconfiguration.openshift.io/role, operator: In, values: [worker, sfc]}
  nodeSelector: ❷
  matchLabels:
    node-role.kubernetes.io/sfc: ""
  paused: false
  maxUnavailable: 1
```

❶ Matches the labels in the MachineConfig.

❷ Matches the labels on the node.

There are predefined **MachineConfigPools** in the OCP cluster:

- **worker**: Targets all worker nodes in the cluster

- **master**: Targets all master nodes in the cluster

Define the following **MachineConfig** to target the master **MachineConfigPool**:

```
metadata:
  labels:
    machineconfiguration.openshift.io/role: master
```

Define the following **MachineConfig** to target the worker **MachineConfigPool**:

```
metadata:
  labels:
    machineconfiguration.openshift.io/role: worker
```

#### Additional resources

- [About MachineConfigPool](#)

## 4.17. DEBUGGING AND TROUBLESHOOTING

If the kmods in your driver container are not signed or are signed with the wrong key, then the container can enter a **PostStartHookError** or **CrashLoopBackOff** status. You can verify by running the **oc describe** command on your container, which displays the following message in this scenario:

```
modprobe: ERROR: could not insert '<your_kmod_name>': Required key not available
```

## 4.18. KMM FIRMWARE SUPPORT

Kernel modules sometimes need to load firmware files from the file system. KMM supports copying firmware files from the kmod image to the node's file system.

The contents of **.spec.moduleLoader.container.modprobe.firmwarePath** are copied into the **/var/lib/firmware** path on the node before running the **modprobe** command to insert the kernel module.

All files and empty directories are removed from that location before running the **modprobe -r** command to unload the kernel module, when the pod is terminated.

### 4.18.1. Configuring the lookup path on nodes

On OpenShift Container Platform nodes, the set of default lookup paths for firmwares does not include the **/var/lib/firmware** path.

#### Procedure

1. Use the Machine Config Operator to create a **MachineConfig** custom resource (CR) that contains the **/var/lib/firmware** path:

```
apiVersion: machineconfiguration.openshift.io/v1
kind: MachineConfig
metadata:
  labels:
```



```

machineconfiguration.openshift.io/role: worker 1
name: 99-worker-kernel-args-firmware-path
spec:
  kernelArguments:
    - 'firmware_class.path=/var/lib/firmware'

```

- 1** You can configure the label based on your needs. In the case of single-node OpenShift, use either **control-plane** or **master** objects.

2. By applying the **MachineConfig** CR, the nodes are automatically rebooted.

### Additional resources

- [Machine Config Operator](#).

## 4.18.2. Building a kmod image

### Procedure

- In addition to building the kernel module itself, include the binary firmware in the builder image:

```

FROM registry.redhat.io/ubi9/ubi-minimal as builder

# Build the kmod

RUN ["mkdir", "/firmware"]
RUN ["curl", "-o", "/firmware/firmware.bin", "https://artifacts.example.com/firmware.bin"]

FROM registry.redhat.io/ubi9/ubi-minimal

# Copy the kmod, install modprobe, run depmod

COPY --from=builder /firmware /firmware

```

## 4.18.3. Tuning the Module resource

### Procedure

- Set **.spec.moduleLoader.container.modprobe.firmwarePath** in the **Module** custom resource (CR):

```

apiVersion: kmm.sigs.x-k8s.io/v1beta1
kind: Module
metadata:
  name: my-kmod
spec:
  moduleLoader:
    container:
      modprobe:
        moduleName: my-kmod # Required

    firmwarePath: /firmware 1

```

- 1 Optional: Copies `/firmware/*` into `/var/lib/firmware/` on the node.

## 4.19. DAY 0 THROUGH DAY 2 KMOD INSTALLATION

You can install some kernel modules (kmods) during Day 0 through Day 2 operations without Kernel Module Management (KMM). This could assist in the transition of the kmods to KMM.

Use the following criteria to determine suitable kmod installations.

### Day 0

The most basic kmods that are required for a node to become **Ready** in the cluster. Examples of these types of kmods include:

- A storage driver that is required to mount the rootFS as part of the boot process
- A network driver that is required for the machine to access **machine-config-server** on the bootstrap node to pull the ignition and join the cluster

### Day 1

Kmods that are not required for a node to become **Ready** in the cluster but cannot be unloaded when the node is **Ready**.

An example of this type of kmod is an out-of-tree (OOT) network driver that replaces an outdated in-tree driver to exploit the full potential of the NIC while **NetworkManager** depends on it. When the node is **Ready**, you cannot unload the driver because of the **NetworkManager** dependency.

### Day 2

Kmods that can be dynamically loaded to the kernel or removed from it without interfering with the cluster infrastructure, for example, connectivity.

Examples of these types of kmods include:

- GPU operators
- Secondary network adapters
- field-programmable gate arrays (FPGAs)

### 4.19.1. Layering background

When a Day 0 kmod is installed in the cluster, layering is applied through the Machine Config Operator (MCO) and OpenShift Container Platform upgrades do not trigger node upgrades.

You only need to recompile the driver if you add new features to it, because the node's operating system will remain the same.

### 4.19.2. Lifecycle management

You can leverage KMM to manage the Day 0 through Day 2 lifecycle of kmods without a reboot when the driver allows it.

**NOTE**

This will not work if the upgrade requires a node reboot, for example, when rebuilding **initramfs** files is needed.

Use one of the following options for lifecycle management.

**4.19.2.1. Treat the kmod as an in-tree driver**

Use this method when you want to upgrade the kmods. In this case, treat the kmod as an in-tree driver and create a **Module** in the cluster with the **inTreeRemoval** field to unload the old version of the driver.

Note the following characteristics of treating the kmod as an in-tree driver:

- Downtime might occur as KMM tries to unload and load the kmod on all the selected nodes simultaneously.
- This works if removing the driver makes the node lose connectivity because KMM uses a single pod to unload and load the driver.

**4.19.2.2. Use ordered upgrade**

You can use ordered upgrade (`ordered_upgrade.md`) to create a versioned **Module** in the cluster representing the kmods with no effect, because the kmods are already loaded.

Note the following characteristics of using ordered upgrade:

- There is no cluster downtime because you control the pace of the upgrade and how many nodes are upgraded at the same time; therefore, an upgrade with no downtime is possible.
- This method will not work if unloading the driver results in losing connection to the node, because KMM creates two different worker pods for unloading and another for loading. These pods will not be scheduled.

**4.20. TROUBLESHOOTING KMM**

When troubleshooting KMM installation issues, you can monitor logs to determine at which stage issues occur. Then, retrieve diagnostic data relevant to that stage.

**4.20.1. Reading Operator logs**

You can use the **oc logs** command to read Operator logs, as in the following examples.

**Example command for KMM controller**

```
$ oc logs -fn openshift-kmm deployments/kmm-operator-controller
```

**Example command for KMM webhook server**

```
$ oc logs -fn openshift-kmm deployments/kmm-operator-webhook-server
```

**Example command for KMM-Hub controller**

```
$ oc logs -fn openshift-kmm-hub deployments/kmm-operator-hub-controller
```

### Example command for KMM-Hub webhook server

```
$ oc logs -fn openshift-kmm deployments/kmm-operator-hub-webhook-server
```

## 4.20.2. Observing events

Use the following methods to view KMM events.

### Build & sign

KMM publishes events whenever it starts a kmod image build or observes its outcome. These events are attached to **Module** objects and are available at the end of the output of **oc describe module** command, as in the following example:

```
$ oc describe modules.kmm.sigs.x-k8s.io kmm-ci-a
[...]
Events:
  Type    Reason          Age           From Message
  ----    -
Normal   BuildCreated    2m29s        kmm   Build created for kernel 6.6.2-201.fc39.x86_64
Normal   BuildSucceeded  63s          kmm   Build job succeeded for kernel 6.6.2-201.fc39.x86_64
Normal   SignCreated     64s (x2 over 64s) kmm   Sign created for kernel 6.6.2-201.fc39.x86_64
Normal   SignSucceeded  57s          kmm   Sign job succeeded for kernel 6.6.2-201.fc39.x86_64
```

### Module load or unload

KMM publishes events whenever it successfully loads or unloads a kernel module on a node. These events are attached to **Node** objects and are available at the end of the output of **oc describe node** command, as in the following example:

```
$ oc describe node my-node
[...]
Events:
  Type    Reason          Age    From Message
  ----    -
[...]
Normal   ModuleLoaded    4m17s kmm   Module default/kmm-ci-a loaded into the kernel
Normal   ModuleUnloaded  2s     kmm   Module default/kmm-ci-a unloaded from the kernel
```

## 4.20.3. Using the must-gather tool

The **oc adm must-gather** command is the preferred way to collect a support bundle and provide debugging information to Red Hat Support. Collect specific information by running the command with the appropriate arguments as described in the following sections.

### Additional resources

- [About the must-gather tool](#)

### 4.20.3.1. Gathering data for KMM

#### Procedure

## 1. Gather the data for the KMM Operator controller manager:

- a. Set the
- MUST\_GATHER\_IMAGE**
- variable:

```
$ export MUST_GATHER_IMAGE=$(oc get deployment -n openshift-kmm kmm-
operator-controller -ojsonpath='{.spec.template.spec.containers[?
(@.name=="manager")].env[?
(@.name=="RELATED_IMAGE_MUST_GATHER")].value}')
$ oc adm must-gather --image="${MUST_GATHER_IMAGE}" -- /usr/bin/gather
```

**NOTE**

Use the **-n <namespace>** switch to specify a namespace if you installed KMM in a custom namespace.

- b. Run the
- must-gather**
- tool:

```
$ oc adm must-gather --image="${MUST_GATHER_IMAGE}" -- /usr/bin/gather
```

## 2. View the Operator logs:

```
$ oc logs -fn openshift-kmm deployments/kmm-operator-controller
```

**Example 4.1. Example output**

```
10228 09:36:37.352405      1 request.go:682] Waited for 1.001998746s due to client-side
throttling, not priority and fairness, request:
GET:https://172.30.0.1:443/apis/machine.openshift.io/v1beta1?timeout=32s
10228 09:36:40.767060      1 listener.go:44] kmm/controller-runtime/metrics
"msg"="Metrics server is starting to listen" "addr"="127.0.0.1:8080"
10228 09:36:40.769483      1 main.go:234] kmm/setup "msg"="starting manager"
10228 09:36:40.769907      1 internal.go:366] kmm "msg"="Starting server" "addr"=
{"IP":"127.0.0.1","Port":8080,"Zone":""} "kind"="metrics" "path"="/metrics"
10228 09:36:40.770025      1 internal.go:366] kmm "msg"="Starting server" "addr"=
{"IP":"","Port":8081,"Zone":""} "kind"="health probe"
10228 09:36:40.770128      1 leaderelection.go:248] attempting to acquire leader lease
openshift-kmm/kmm.sigs.x-k8s.io...
10228 09:36:40.784396      1 leaderelection.go:258] successfully acquired lease
openshift-kmm/kmm.sigs.x-k8s.io
10228 09:36:40.784876      1 controller.go:185] kmm "msg"="Starting EventSource"
"controller"="Module" "controllerGroup"="kmm.sigs.x-k8s.io" "controllerKind"="Module"
"source"="kind source: *v1beta1.Module"
10228 09:36:40.784925      1 controller.go:185] kmm "msg"="Starting EventSource"
"controller"="Module" "controllerGroup"="kmm.sigs.x-k8s.io" "controllerKind"="Module"
"source"="kind source: *v1.DaemonSet"
10228 09:36:40.784968      1 controller.go:185] kmm "msg"="Starting EventSource"
"controller"="Module" "controllerGroup"="kmm.sigs.x-k8s.io" "controllerKind"="Module"
"source"="kind source: *v1.Build"
10228 09:36:40.785001      1 controller.go:185] kmm "msg"="Starting EventSource"
"controller"="Module" "controllerGroup"="kmm.sigs.x-k8s.io" "controllerKind"="Module"
"source"="kind source: *v1.Job"
10228 09:36:40.785025      1 controller.go:185] kmm "msg"="Starting EventSource"
"controller"="Module" "controllerGroup"="kmm.sigs.x-k8s.io" "controllerKind"="Module"
"source"="kind source: *v1.Node"
```

```

10228 09:36:40.785039      1 controller.go:193] kmm "msg"="Starting Controller"
"controller"="Module" "controllerGroup"="kmm.sigs.x-k8s.io" "controllerKind"="Module"
10228 09:36:40.785458      1 controller.go:185] kmm "msg"="Starting EventSource"
"controller"="PodNodeModule" "controllerGroup"="" "controllerKind"="Pod" "source"="kind
source: *v1.Pod"
10228 09:36:40.786947      1 controller.go:185] kmm "msg"="Starting EventSource"
"controller"="PreflightValidation" "controllerGroup"="kmm.sigs.x-k8s.io"
"controllerKind"="PreflightValidation" "source"="kind source: *v1beta1.PreflightValidation"
10228 09:36:40.787406      1 controller.go:185] kmm "msg"="Starting EventSource"
"controller"="PreflightValidation" "controllerGroup"="kmm.sigs.x-k8s.io"
"controllerKind"="PreflightValidation" "source"="kind source: *v1.Build"
10228 09:36:40.787474      1 controller.go:185] kmm "msg"="Starting EventSource"
"controller"="PreflightValidation" "controllerGroup"="kmm.sigs.x-k8s.io"
"controllerKind"="PreflightValidation" "source"="kind source: *v1.Job"
10228 09:36:40.787488      1 controller.go:185] kmm "msg"="Starting EventSource"
"controller"="PreflightValidation" "controllerGroup"="kmm.sigs.x-k8s.io"
"controllerKind"="PreflightValidation" "source"="kind source: *v1beta1.Module"
10228 09:36:40.787603      1 controller.go:185] kmm "msg"="Starting EventSource"
"controller"="NodeKernel" "controllerGroup"="" "controllerKind"="Node" "source"="kind
source: *v1.Node"
10228 09:36:40.787634      1 controller.go:193] kmm "msg"="Starting Controller"
"controller"="NodeKernel" "controllerGroup"="" "controllerKind"="Node"
10228 09:36:40.787680      1 controller.go:193] kmm "msg"="Starting Controller"
"controller"="PreflightValidation" "controllerGroup"="kmm.sigs.x-k8s.io"
"controllerKind"="PreflightValidation"
10228 09:36:40.785607      1 controller.go:185] kmm "msg"="Starting EventSource"
"controller"="imagestream" "controllerGroup"="image.openshift.io"
"controllerKind"="ImageStream" "source"="kind source: *v1.ImageStream"
10228 09:36:40.787822      1 controller.go:185] kmm "msg"="Starting EventSource"
"controller"="preflightvalidationocp" "controllerGroup"="kmm.sigs.x-k8s.io"
"controllerKind"="PreflightValidationOCP" "source"="kind source:
*v1beta1.PreflightValidationOCP"
10228 09:36:40.787853      1 controller.go:193] kmm "msg"="Starting Controller"
"controller"="imagestream" "controllerGroup"="image.openshift.io"
"controllerKind"="ImageStream"
10228 09:36:40.787879      1 controller.go:185] kmm "msg"="Starting EventSource"
"controller"="preflightvalidationocp" "controllerGroup"="kmm.sigs.x-k8s.io"
"controllerKind"="PreflightValidationOCP" "source"="kind source:
*v1beta1.PreflightValidation"
10228 09:36:40.787905      1 controller.go:193] kmm "msg"="Starting Controller"
"controller"="preflightvalidationocp" "controllerGroup"="kmm.sigs.x-k8s.io"
"controllerKind"="PreflightValidationOCP"
10228 09:36:40.786489      1 controller.go:193] kmm "msg"="Starting Controller"
"controller"="PodNodeModule" "controllerGroup"="" "controllerKind"="Pod"

```

#### 4.20.3.2. Gathering data for KMM-Hub

##### Procedure

1. Gather the data for the KMM Operator hub controller manager:
  - a. Set the **MUST\_GATHER\_IMAGE** variable:

```
$ export MUST_GATHER_IMAGE=$(oc get deployment -n openshift-kmm-hub kmm-
```

```
operator-hub-controller -ojsonpath='{.spec.template.spec.containers[?
(@.name=="manager")].env[?
(@.name=="RELATED_IMAGE_MUST_GATHER")].value}'
$ oc adm must-gather --image="{MUST_GATHER_IMAGE}" -- /usr/bin/gather -u
```



## NOTE

Use the **-n <namespace>** switch to specify a namespace if you installed KMM in a custom namespace.

- b. Run the **must-gather** tool:

```
$ oc adm must-gather --image="{MUST_GATHER_IMAGE}" -- /usr/bin/gather -u
```

2. View the Operator logs:

```
$ oc logs -fn openshift-kmm-hub deployments/kmm-operator-hub-controller
```

### Example 4.2. Example output

```
10417 11:34:08.807472    1 request.go:682] Waited for 1.023403273s due to client-side
throttling, not priority and fairness, request:
GET:https://172.30.0.1:443/apis/tuned.openshift.io/v1?timeout=32s
10417 11:34:12.373413    1 listener.go:44] kmm-hub/controller-runtime/metrics
"msg"="Metrics server is starting to listen" "addr"="127.0.0.1:8080"
10417 11:34:12.376253    1 main.go:150] kmm-hub/setup "msg"="Adding controller"
"name"="ManagedClusterModule"
10417 11:34:12.376621    1 main.go:186] kmm-hub/setup "msg"="starting manager"
10417 11:34:12.377690    1 leaderelection.go:248] attempting to acquire leader lease
openshift-kmm-hub/kmm-hub.sigs.x-k8s.io...
10417 11:34:12.378078    1 internal.go:366] kmm-hub "msg"="Starting server" "addr"=
{"IP":"127.0.0.1","Port":8080,"Zone":""} "kind"="metrics" "path"="/metrics"
10417 11:34:12.378222    1 internal.go:366] kmm-hub "msg"="Starting server" "addr"=
{"IP":"","Port":8081,"Zone":""} "kind"="health probe"
10417 11:34:12.395703    1 leaderelection.go:258] successfully acquired lease
openshift-kmm-hub/kmm-hub.sigs.x-k8s.io
10417 11:34:12.396334    1 controller.go:185] kmm-hub "msg"="Starting EventSource"
"controller"="ManagedClusterModule" "controllerGroup"="hub.kmm.sigs.x-k8s.io"
"controllerKind"="ManagedClusterModule" "source"="kind source:
*v1beta1.ManagedClusterModule"
10417 11:34:12.396403    1 controller.go:185] kmm-hub "msg"="Starting EventSource"
"controller"="ManagedClusterModule" "controllerGroup"="hub.kmm.sigs.x-k8s.io"
"controllerKind"="ManagedClusterModule" "source"="kind source: *v1.ManifestWork"
10417 11:34:12.396430    1 controller.go:185] kmm-hub "msg"="Starting EventSource"
"controller"="ManagedClusterModule" "controllerGroup"="hub.kmm.sigs.x-k8s.io"
"controllerKind"="ManagedClusterModule" "source"="kind source: *v1.Build"
10417 11:34:12.396469    1 controller.go:185] kmm-hub "msg"="Starting EventSource"
"controller"="ManagedClusterModule" "controllerGroup"="hub.kmm.sigs.x-k8s.io"
"controllerKind"="ManagedClusterModule" "source"="kind source: *v1.Job"
10417 11:34:12.396522    1 controller.go:185] kmm-hub "msg"="Starting EventSource"
"controller"="ManagedClusterModule" "controllerGroup"="hub.kmm.sigs.x-k8s.io"
"controllerKind"="ManagedClusterModule" "source"="kind source: *v1.ManagedCluster"
10417 11:34:12.396543    1 controller.go:193] kmm-hub "msg"="Starting Controller"
"controller"="ManagedClusterModule" "controllerGroup"="hub.kmm.sigs.x-k8s.io"
```

```
"controllerKind"="ManagedClusterModule"
I0417 11:34:12.397175    1 controller.go:185] kmm-hub "msg"="Starting EventSource"
"controller"="imagestream" "controllerGroup"="image.openshift.io"
"controllerKind"="ImageStream" "source"="kind source: *v1.ImageStream"
I0417 11:34:12.397221    1 controller.go:193] kmm-hub "msg"="Starting Controller"
"controller"="imagestream" "controllerGroup"="image.openshift.io"
"controllerKind"="ImageStream"
I0417 11:34:12.498335    1 filter.go:196] kmm-hub "msg"="Listing all
ManagedClusterModules" "managedcluster"="local-cluster"
I0417 11:34:12.498570    1 filter.go:205] kmm-hub "msg"="Listed
ManagedClusterModules" "count"=0 "managedcluster"="local-cluster"
I0417 11:34:12.498629    1 filter.go:238] kmm-hub "msg"="Adding reconciliation
requests" "count"=0 "managedcluster"="local-cluster"
I0417 11:34:12.498687    1 filter.go:196] kmm-hub "msg"="Listing all
ManagedClusterModules" "managedcluster"="sno1-0"
I0417 11:34:12.498750    1 filter.go:205] kmm-hub "msg"="Listed
ManagedClusterModules" "count"=0 "managedcluster"="sno1-0"
I0417 11:34:12.498801    1 filter.go:238] kmm-hub "msg"="Adding reconciliation
requests" "count"=0 "managedcluster"="sno1-0"
I0417 11:34:12.501947    1 controller.go:227] kmm-hub "msg"="Starting workers"
"controller"="imagestream" "controllerGroup"="image.openshift.io"
"controllerKind"="ImageStream" "worker count"=1
I0417 11:34:12.501948    1 controller.go:227] kmm-hub "msg"="Starting workers"
"controller"="ManagedClusterModule" "controllerGroup"="hub.kmm.sigs.x-k8s.io"
"controllerKind"="ManagedClusterModule" "worker count"=1
I0417 11:34:12.502285    1 imagestream_reconciler.go:50] kmm-hub "msg"="registered
imagestream info mapping" "ImageStream"={"name":"driver-
toolkit","namespace":"openshift"} "controller"="imagestream"
"controllerGroup"="image.openshift.io" "controllerKind"="ImageStream"
"dtkImage"="quay.io/openshift-release-dev/ocp-v4.0-art-
dev@sha256:df42b4785a7a662b30da53bdb0d206120cf4d24b45674227b16051ba4b7c393
4" "name"="driver-toolkit" "namespace"="openshift"
"osImageVersion"="412.86.202302211547-0" "reconcileID"="e709ff0a-5664-4007-8270-
49b5dff8bae9"
```