OpenShift Dedicated 4

Introduction to OpenShift Dedicated

An overview of OpenShift Dedicated architecture

Last Updated: 2024-07-02

OpenShift Dedicated 4 Introduction to OpenShift Dedicated

An overview of OpenShift Dedicated architecture

Legal Notice

Abstract

This document provides an overview of the platform and application architecture in OpenShift Dedicated.

Table of Contents

CHAPTER 1. UNDERSTANDING OPENSHIFT DEDICATED

With its foundation in Kubernetes, OpenShift Dedicated is a complete OpenShift Container Platform cluster provided as a cloud service, configured for high availability, and dedicated to a single customer.

1.1. AN OVERVIEW OF OPENSHIFT DEDICATED

OpenShift Dedicated is professionally managed by Red Hat and hosted on Amazon Web Services (AWS) or Google Cloud Platform (GCP). Each OpenShift Dedicated cluster comes with a fully managed control plane (Control and Infrastructure nodes), application nodes, installation and management by Red Hat Site Reliability Engineers (SRE), premium Red Hat Support, and cluster services such as logging, metrics, monitoring, notifications portal, and a cluster portal.

OpenShift Dedicated provides enterprise-ready enhancements to Kubernetes, including the following enhancements:

- OpenShift Dedicated clusters are deployed on AWS or GCP environments and can be used as part of a hybrid approach for application management.
- Integrated Red Hat technology. Major components in OpenShift Dedicated come from Red Hat Enterprise Linux and related Red Hat technologies. OpenShift Dedicated benefits from the intense testing and certification initiatives for Red Hat's enterprise quality software.
- Open source development model. Development is completed in the open, and the source code is available from public software repositories. This open collaboration fosters rapid innovation and development.

To learn about options for assets you can create when you build and deploy containerized Kubernetes applications in OpenShift Container Platform, see Understanding OpenShift Container Platform development.

1.1.1. Custom operating system

OpenShift Dedicated uses Red Hat Enterprise Linux CoreOS (RHCOS), a container-oriented operating system that combines some of the best features and functions of the CoreOS and Red Hat Atomic Host operating systems. RHCOS is specifically designed for running containerized applications from OpenShift Dedicated and works with new tools to provide fast installation, Operator-based management, and simplified upgrades.

RHCOS includes:

- Ignition, which OpenShift Dedicated uses as a firstboot system configuration for initially bringing up and configuring machines.
- CRI-O, a Kubernetes native container runtime implementation that integrates closely with the operating system to deliver an efficient and optimized Kubernetes experience. CRI-O provides facilities for running, stopping, and restarting containers.
- Kubelet, the primary node agent for Kubernetes that is responsible for launching and monitoring containers.

1.1.2. Other key features

Operators are both the fundamental unit of the OpenShift Dedicated code base and a convenient way to deploy applications and software components for your applications to use. In OpenShift Dedicated,

Operators serve as the platform foundation and remove the need for manual upgrades of operating systems and control plane applications. OpenShift Dedicated Operators such as the Cluster Version Operator and Machine Config Operator allow simplified, cluster-wide management of those critical components.

Operator Lifecycle Manager (OLM) and the OperatorHub provide facilities for storing and distributing Operators to people developing and deploying applications.

The Red Hat Quay Container Registry is a Quay.io container registry that serves most of the container images and Operators to OpenShift Dedicated clusters. Quay.io is a public registry version of Red Hat Quay that stores millions of images and tags.

Other enhancements to Kubernetes in OpenShift Dedicated include improvements in software defined networking (SDN), authentication, log aggregation, monitoring, and routing. OpenShift Dedicated also offers a comprehensive web console and the custom OpenShift CLI (**oc**) interface.

1.1.3. Internet and Telemetry access for OpenShift Dedicated

In OpenShift Dedicated, you require access to the internet to install and upgrade your cluster.

Through the Telemetry service, information is sent to Red Hat from OpenShift Dedicated clusters to enable subscription management automation, monitor the health of clusters, assist with support, and improve customer experience.

The Telemetry service runs automatically and your cluster is registered to Red Hat OpenShift Cluster Manager. In OpenShift Dedicated, remote health reporting is always enabled and you cannot opt out. The Red Hat Site Reliability Engineering (SRE) team requires the information to provide effective support for your OpenShift Dedicated cluster.

Additional resources

- For more information about Telemetry and remote health monitoring for OpenShift Dedicated clusters, see About remote health monitoring

CHAPTER 2. POLICIES AND SERVICE DEFINITION

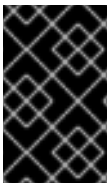
2.1. OPENSHIFT DEDICATED SERVICE DEFINITION

2.1.1. Account management

2.1.1.1. Billing options

Customers have the option to purchase annual subscriptions of OpenShift Dedicated (OSD) or consume on-demand through cloud marketplaces. Customers can decide to bring their own cloud infrastructure account, referred to as Customer Cloud Subscription (CCS), or deploy in cloud provider accounts owned by Red Hat. The table below provides additional information regarding billing, as well as the corresponding supported deployment options.

OSD Subscription-type	Cloud infrastructure account	Billed through
Annual fixed capacity subscriptions through Red Hat	Red Hat cloud account	Red Hat for consumption of both OSD subscriptions and cloud infrastructure
	Customer's own cloud account	Red Hat for consumption of the OSD subscriptions Cloud provider for consumption of cloud infrastructure
On-demand usage-based consumption through Google Cloud Marketplace	Customer's own Google Cloud account	Google Cloud for both cloud infrastructure and Red Hat OSD subscriptions
On-demand usage-based consumption through Red Hat Marketplace	Customer's own cloud account	Red Hat for consumption of the OSD subscriptions Cloud provider for consumption of cloud infrastructure



IMPORTANT

Customers that use their own cloud infrastructure account, referred to as Customer Cloud Subscription (CSS), are responsible to pre-purchase or provide Reserved Instance (RI) compute instances to ensure lower cloud infrastructure costs.

Additional resources can be purchased for an OpenShift Dedicated Cluster, including:

- Additional nodes (can be different types and sizes through the use of machine pools)
- Middleware (JBoss EAP, JBoss Fuse, and so on) - additional pricing based on specific middleware component
- Additional storage in increments of 500 GB (standard only; 100 GB included)

- Additional 12 TiB Network I/O (standard only; 12 TB included)
- Load Balancers for Services are available in bundles of 4; enables non-HTTP/SNI traffic or non-standard ports (standard only)

2.1.1.2. Cluster self-service

Customers can create, scale, and delete their clusters from OpenShift Cluster Manager, provided that they have already purchased the necessary subscriptions.

Actions available in Red Hat OpenShift Cluster Manager must not be directly performed from within the cluster as this might cause adverse affects, including having all actions automatically reverted.

2.1.1.3. Cloud providers

OpenShift Dedicated offers OpenShift Container Platform clusters as a managed service on the following cloud providers:

- Amazon Web Services (AWS)
- Google Cloud Platform (GCP)

2.1.1.4. Instance types

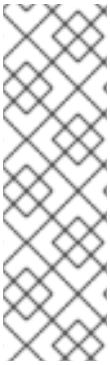
Single availability zone clusters require a minimum of 2 worker nodes for Customer Cloud Subscription (CCS) clusters deployed to a single availability zone. A minimum of 4 worker nodes is required for standard clusters. These 4 worker nodes are included in the base subscription.

Multiple availability zone clusters require a minimum of 3 worker nodes for Customer Cloud Subscription (CCS) clusters, 1 deployed to each of 3 availability zones. A minimum of 9 worker nodes are required for standard clusters. These 9 worker nodes are included in the base subscription, and additional nodes must be purchased in multiples of 3 to maintain proper node distribution.

NOTE

All worker nodes within a single OpenShift Dedicated machine pool must be of the same type and size. However, worker nodes across multiple machine pools within an OpenShift Dedicated cluster can be of different types and sizes.

Control plane and infrastructure nodes are also provided by Red Hat. There are at least 3 control plane nodes that handle etcd and API-related workloads. There are at least 2 infrastructure nodes that handle metrics, routing, the web console, and other workloads. You must not run any workloads on the control plane and infrastructure nodes. Any workloads you intend to run must be deployed on worker nodes. See the Red Hat Operator support section below for more information about Red Hat workloads that must be deployed on worker nodes.



NOTE

Approximately 1 vCPU core and 1 GiB of memory are reserved on each worker node and removed from allocatable resources. This is necessary to run processes required by the underlying platform. This includes system daemons such as udev, kubelet, container runtime, and so on, and also accounts for kernel reservations. OpenShift Container Platform core systems such as audit log aggregation, metrics collection, DNS, image registry, SDN, and so on might consume additional allocatable resources to maintain the stability and maintainability of the cluster. The additional resources consumed might vary based on usage.

IMPORTANT

As of OpenShift Dedicated 4.11, the default per-pod PID limit is **4096**. If you want to enable this PID limit, you must upgrade your OpenShift Dedicated clusters to this version or later. OpenShift Dedicated clusters running versions earlier than 4.11 use a default PID limit of **1024**.

You cannot configure the per-pod PID limit on any OpenShift Dedicated cluster.

Additional Resources

- Red Hat Operator Support

2.1.1.5. AWS instance types for Customer Cloud Subscription clusters

OpenShift Dedicated offers the following worker node instance types and sizes on AWS:

Example 2.1. General purpose

- m5.metal (96† vCPU, 384 GiB)
- m5.xlarge (4 vCPU, 16 GiB)
- m5.2xlarge (8 vCPU, 32 GiB)
- m5.4xlarge (16 vCPU, 64 GiB)
- m5.8xlarge (32 vCPU, 128 GiB)
- m5.12xlarge (48 vCPU, 192 GiB)
- m5.16xlarge (64 vCPU, 256 GiB)
- m5.24xlarge (96 vCPU, 384 GiB)
- m5a.xlarge (4 vCPU, 16 GiB)
- m5a.2xlarge (8 vCPU, 32 GiB)
- m5a.4xlarge (16 vCPU, 64 GiB)
- m5a.8xlarge (32 vCPU, 128 GiB)
- m5a.12xlarge (48 vCPU, 192 GiB)

- m5a.16xlarge (64 vCPU, 256 GiB)
- m5a.24xlarge (96 vCPU, 384 GiB)
- m5ad.xlarge (4 vCPU, 16 GiB)
- m5ad.2xlarge (8 vCPU, 32 GiB)
- m5ad.4xlarge (16 vCPU, 64 GiB)
- m5ad.8xlarge (32 vCPU, 128 GiB)
- m5ad.12xlarge (48 vCPU, 192 GiB)
- m5ad.16xlarge (64 vCPU, 256 GiB)
- m5ad.24xlarge (96 vCPU, 384 GiB)
- m5d.metal (96† vCPU, 384 GiB)
- m5d.xlarge (4 vCPU, 16 GiB)
- m5d.2xlarge (8 vCPU, 32 GiB)
- m5d.4xlarge (16 vCPU, 64 GiB)
- m5d.8xlarge (32 vCPU, 128 GiB)
- m5d.12xlarge (48 vCPU, 192 GiB)
- m5d.16xlarge (64 vCPU, 256 GiB)
- m5d.24xlarge (96 vCPU, 384 GiB)
- m5n.metal (96 vCPU, 384 GiB)
- m5n.xlarge (4 vCPU, 16 GiB)
- m5n.2xlarge (8 vCPU, 32 GiB)
- m5n.4xlarge (16 vCPU, 64 GiB)
- m5n.8xlarge (32 vCPU, 128 GiB)
- m5n.12xlarge (48 vCPU, 192 GiB)
- m5n.16xlarge (64 vCPU, 256 GiB)
- m5n.24xlarge (96 vCPU, 384 GiB)
- m5dn.metal (96 vCPU, 384 GiB)
- m5dn.xlarge (4 vCPU, 16 GiB)
- m5dn.2xlarge (8 vCPU, 32 GiB)
- m5dn.4xlarge (16 vCPU, 64 GiB)

- m5dn.8xlarge (32 vCPU, 128 GiB)
- m5dn.12xlarge (48 vCPU, 192 GiB)
- m5dn.16xlarge (64 vCPU, 256 GiB)
- m5dn.24xlarge (96 vCPU, 384 GiB)
- m5zn.metal (48 vCPU, 192 GiB)
- m5zn.xlarge (4 vCPU, 16 GiB)
- m5zn.2xlarge (8 vCPU, 32 GiB)
- m5zn.3xlarge (12 vCPU, 48 GiB)
- m5zn.6xlarge (24 vCPU, 96 GiB)
- m5zn.12xlarge (48 vCPU, 192 GiB)
- m6a.xlarge (4 vCPU, 16 GiB)
- m6a.2xlarge (8 vCPU, 32 GiB)
- m6a.4xlarge (16 vCPU, 64 GiB)
- m6a.8xlarge (32 vCPU, 128 GiB)
- m6a.12xlarge (48 vCPU, 192 GiB)
- m6a.16xlarge (64 vCPU, 256 GiB)
- m6a.24xlarge (96 vCPU, 384 GiB)
- m6a.32xlarge (128 vCPU, 512 GiB)
- m6a.48xlarge (192 vCPU, 768 GiB)
- m6i.metal (128 vCPU, 512 GiB)
- m6i.xlarge (4 vCPU, 16 GiB)
- m6i.2xlarge (8 vCPU, 32 GiB)
- m6i.4xlarge (16 vCPU, 64 GiB)
- m6i.8xlarge (32 vCPU, 128 GiB)
- m6i.12xlarge (48 vCPU, 192 GiB)
- m6i.16xlarge (64 vCPU, 256 GiB)
- m6i.24xlarge (96 vCPU, 384 GiB)
- m6i.32xlarge (128 vCPU, 512 GiB)
- m6id.xlarge (4 vCPU, 16 GiB)

- m6id.2xlarge (8 vCPU, 32 GiB)
- m6id.4xlarge (16 vCPU, 64 GiB)
- m6id.8xlarge (32 vCPU, 128 GiB)
- m6id.12xlarge (48 vCPU, 192 GiB)
- m6id.16xlarge (64 vCPU, 256 GiB)
- m6id.24xlarge (96 vCPU, 384 GiB)
- m6id.32xlarge (128 vCPU, 512 GiB)
- m7i.xlarge (4 vCPU, 16 GiB)
- m7i.2xlarge (8 vCPU, 32 GiB)
- m7i.4xlarge (16 vCPU, 64 GiB)
- m7i.8xlarge (32 vCPU, 128 GiB)
- m7i.12xlarge (48 vCPU, 192 GiB)
- m7i.16xlarge (64 vCPU, 256 GiB)
- m7i.24xlarge (96 vCPU, 384 GiB)
- m7i.48xlarge (192 vCPU, 768 GiB)
- m7i.metal-24xl (96 vCPU, 384 GiB)
- m7i.metal-48xl (192 vCPU, 768 GiB)
- m7i-flex.xlarge (4 vCPU, 16 GiB)
- m7i-flex.2xlarge (8 vCPU, 32 GiB)
- m7i-flex.4xlarge (16 vCPU, 64 GiB)
- m7i-flex.8xlarge (32 vCPU, 128 GiB)
- m7a.xlarge (4 vCPU, 16 GiB)
- m7a.2xlarge (8 vCPU, 32 GiB)
- m7a.4xlarge (16 vCPU, 64 GiB)
- m7a.8xlarge (32 vCPU, 128 GiB)
- m7a.12xlarge (48 vCPU, 192 GiB)
- m7a.16xlarge (64 vCPU, 256 GiB)
- m7a.24xlarge (96 vCPU, 384 GiB)
- m7a.32xlarge (128 vCPU, 512 GiB)

- m7a.48xlarge (192 vCPU, 768 GiB)
- m7a.metal-48xl (192 vCPU, 768 GiB)

† These instance types provide 96 logical processors on 48 physical cores. They run on single servers with two physical Intel sockets.

Example 2.2. Burstable general purpose

- t3.xlarge (4 vCPU, 16 GiB)
- t3.2xlarge (8 vCPU, 32 GiB)
- t3a.xlarge (4 vCPU, 16 GiB)
- t3a.2xlarge (8 vCPU, 32 GiB)

Example 2.3. Memory intensive

- x1.16xlarge (64 vCPU, 976 GiB)
- x1.32xlarge (128 vCPU, 1952 GiB)
- x1e.xlarge (4 vCPU, 122 GiB)
- x1e.2xlarge (8 vCPU, 244 GiB)
- x1e.4xlarge (16 vCPU, 488 GiB)
- x1e.8xlarge (32 vCPU, 976 GiB)
- x1e.16xlarge (64 vCPU, 1,952 GiB)
- x1e.32xlarge (128 vCPU, 3,904 GiB)
- x2idn.16xlarge (64 vCPU, 1024 GiB)
- x2idn.24xlarge (96 vCPU, 1536 GiB)
- x2idn.32xlarge (128 vCPU, 2048 GiB)
- x2iedn.xlarge (4 vCPU, 128 GiB)
- x2iedn.2xlarge (8 vCPU, 256 GiB)
- x2iedn.4xlarge (16 vCPU, 512 GiB)
- x2iedn.8xlarge (32 vCPU, 1024 GiB)
- x2iedn.16xlarge (64 vCPU, 2048 GiB)
- x2iedn.24xlarge (96 vCPU, 3072 GiB)
- x2iedn.32xlarge (128 vCPU, 4096 GiB)

- x2iezn.2xlarge (8 vCPU, 256 GiB)
- x2iezn.4xlarge (16vCPU, 512 GiB)
- x2iezn.6xlarge (24vCPU, 768 GiB)
- x2iezn.8xlarge (32vCPU, 1,024 GiB)
- x2iezn.12xlarge (48vCPU, 1,536 GiB)
- x2idn.metal (128vCPU, 2,048 GiB)
- x2iedn.metal (128vCPU, 4,096 GiB)
- x2iezn.metal (48 vCPU, 1,536 GiB)

Example 2.4. Memory optimized

- r4.xlarge (4 vCPU, 30.5 GiB)
- r4.2xlarge (8 vCPU, 61 GiB)
- r4.4xlarge (16 vCPU, 122 GiB)
- r4.8xlarge (32 vCPU, 244 GiB)
- r4.16xlarge (64 vCPU, 488 GiB)
- r5.metal (96† vCPU, 768 GiB)
- r5.xlarge (4 vCPU, 32 GiB)
- r5.2xlarge (8 vCPU, 64 GiB)
- r5.4xlarge (16 vCPU, 128 GiB)
- r5.8xlarge (32 vCPU, 256 GiB)
- r5.12xlarge (48 vCPU, 384 GiB)
- r5.16xlarge (64 vCPU, 512 GiB)
- r5.24xlarge (96 vCPU, 768 GiB)
- r5a.xlarge (4 vCPU, 32 GiB)
- r5a.2xlarge (8 vCPU, 64 GiB)
- r5a.4xlarge (16 vCPU, 128 GiB)
- r5a.8xlarge (32 vCPU, 256 GiB)
- r5a.12xlarge (48 vCPU, 384 GiB)
- r5a.16xlarge (64 vCPU, 512 GiB)
- r5a.24xlarge (96 vCPU, 768 GiB)

- r5ad.xlarge (4 vCPU, 32 GiB)
- r5ad.2xlarge (8 vCPU, 64 GiB)
- r5ad.4xlarge (16 vCPU, 128 GiB)
- r5ad.8xlarge (32 vCPU, 256 GiB)
- r5ad.12xlarge (48 vCPU, 384 GiB)
- r5ad.16xlarge (64 vCPU, 512 GiB)
- r5ad.24xlarge (96 vCPU, 768 GiB)
- r5d.metal (96† vCPU, 768 GiB)
- r5d.xlarge (4 vCPU, 32 GiB)
- r5d.2xlarge (8 vCPU, 64 GiB)
- r5d.4xlarge (16 vCPU, 128 GiB)
- r5d.8xlarge (32 vCPU, 256 GiB)
- r5d.12xlarge (48 vCPU, 384 GiB)
- r5d.16xlarge (64 vCPU, 512 GiB)
- r5d.24xlarge (96 vCPU, 768 GiB)
- r5n.metal (96 vCPU, 768 GiB)
- r5n.xlarge (4 vCPU, 32 GiB)
- r5n.2xlarge (8 vCPU, 64 GiB)
- r5n.4xlarge (16 vCPU, 128 GiB)
- r5n.8xlarge (32 vCPU, 256 GiB)
- r5n.12xlarge (48 vCPU, 384 GiB)
- r5n.16xlarge (64 vCPU, 512 GiB)
- r5n.24xlarge (96 vCPU, 768 GiB)
- r5dn.metal (96 vCPU, 768 GiB)
- r5dn.xlarge (4 vCPU, 32 GiB)
- r5dn.2xlarge (8 vCPU, 64 GiB)
- r5dn.4xlarge (16 vCPU, 128 GiB)
- r5dn.8xlarge (32 vCPU, 256 GiB)
- r5dn.12xlarge (48 vCPU, 384 GiB)

- r5dn.16xlarge (64 vCPU, 512 GiB)
- r5dn.24xlarge (96 vCPU, 768 GiB)
- r6a.xlarge (4 vCPU, 32 GiB)
- r6a.2xlarge (8 vCPU, 64 GiB)
- r6a.4xlarge (16 vCPU, 128 GiB)
- r6a.8xlarge (32 vCPU, 256 GiB)
- r6a.12xlarge (48 vCPU, 384 GiB)
- r6a.16xlarge (64 vCPU, 512 GiB)
- r6a.24xlarge (96 vCPU, 768 GiB)
- r6a.32xlarge (128 vCPU, 1,024 GiB)
- r6a.48xlarge (192 vCPU, 1,536 GiB)
- r6i.metal (128 vCPU, 1,024 GiB)
- r6i.xlarge (4 vCPU, 32 GiB)
- r6i.2xlarge (8 vCPU, 64 GiB)
- r6i.4xlarge (16 vCPU, 128 GiB)
- r6i.8xlarge (32 vCPU, 256 GiB)
- r6i.12xlarge (48 vCPU, 384 GiB)
- r6i.16xlarge (64 vCPU, 512 GiB)
- r6i.24xlarge (96 vCPU, 768 GiB)
- r6i.32xlarge (128 vCPU, 1,024 GiB)
- r6id.xlarge (4 vCPU, 32 GiB)
- r6id.2xlarge (8 vCPU, 64 GiB)
- r6id.4xlarge (16 vCPU, 128 GiB)
- r6id.8xlarge (32 vCPU, 256 GiB)
- r6id.12xlarge (48 vCPU, 384 GiB)
- r6id.16xlarge (64 vCPU, 512 GiB)
- r6id.24xlarge (96 vCPU, 768 GiB)
- r6id.32xlarge (128 vCPU, 1,024 GiB)
- z1d.metal (48 vCPU, 384 GiB)

- z1d.xlarge (4 vCPU, 32 GiB)
- z1d.2xlarge (8 vCPU, 64 GiB)
- z1d.3xlarge (12 vCPU, 96 GiB)
- z1d.6xlarge (24 vCPU, 192 GiB)
- z1d.12xlarge (48 vCPU, 384 GiB)
- r7iz.xlarge (4 vCPU, 32 GiB)
- r7iz.2xlarge (8 vCPU, 64 GiB)
- r7iz.4xlarge (16 vCPU, 128 GiB)
- r7iz.8xlarge (32 vCPU, 256 GiB)
- r7iz.12xlarge (48 vCPU, 384 GiB)
- r7iz.16xlarge (64 vCPU, 512 GiB)
- r7iz.32xlarge (128 vCPU, 1024 GiB)
- r7iz.metal-16xl (64 vCPU, 512 GiB)
- r7iz.metal-32xl (128 vCPU, 1024 GiB)

† These instance types provide 96 logical processors on 48 physical cores. They run on single servers with two physical Intel sockets.

This instance type provides 48 logical processors on 24 physical cores.

Example 2.5. Accelerated computing

- p3.2xlarge (8 vCPU, 61 GiB)
- p3.8xlarge (32 vCPU, 244 GiB)
- p3.16xlarge (64 vCPU, 488 GiB)
- p3dn.24xlarge (96 vCPU, 768 GiB)
- p4d.24xlarge (96 vCPU, 1,152 GiB)
- p4de.24xlarge (96 vCPU, 1,152 GiB)
- p5.48xlarge (192 vCPU, 2,048 GiB)
- g4dn.xlarge (4 vCPU, 16 GiB)
- g4dn.2xlarge (8 vCPU, 32 GiB)
- g4dn.4xlarge (16 vCPU, 64 GiB)
- g4dn.8xlarge (32 vCPU, 128 GiB)

- g4dn.12xlarge (48 vCPU, 192 GiB)
- g4dn.16xlarge (64 vCPU, 256 GiB)
- g4dn.metal (96 vCPU, 384 GiB)
- g5.xlarge (4 vCPU, 16 GiB)
- g5.2xlarge (8 vCPU, 32 GiB)
- g5.4xlarge (16 vCPU, 64 GiB)
- g5.8xlarge (32 vCPU, 128 GiB)
- g5.16xlarge (64 vCPU, 256 GiB)
- g5.12xlarge (48 vCPU, 192 GiB)
- g5.24xlarge (96 vCPU, 384 GiB)
- g5.48xlarge (192 vCPU, 768 GiB)
- dl1.24xlarge (96 vCPU, 768 GiB)†

† Intel specific; not covered by Nvidia

Support for the GPU instance type software stack is provided by AWS. Ensure that your AWS service quotas can accommodate the desired GPU instance types.

Example 2.6. Compute optimized

- c5.metal (96 vCPU, 192 GiB)
- c5.xlarge (4 vCPU, 8 GiB)
- c5.2xlarge (8 vCPU, 16 GiB)
- c5.4xlarge (16 vCPU, 32 GiB)
- c5.9xlarge (36 vCPU, 72 GiB)
- c5.12xlarge (48 vCPU, 96 GiB)
- c5.18xlarge (72 vCPU, 144 GiB)
- c5.24xlarge (96 vCPU, 192 GiB)
- c5d.metal (96 vCPU, 192 GiB)
- c5d.xlarge (4 vCPU, 8 GiB)
- c5d.2xlarge (8 vCPU, 16 GiB)
- c5d.4xlarge (16 vCPU, 32 GiB)
- c5d.9xlarge (36 vCPU, 72 GiB)

- c5d.12xlarge (48 vCPU, 96 GiB)
- c5d.18xlarge (72 vCPU, 144 GiB)
- c5d.24xlarge (96 vCPU, 192 GiB)
- c5a.xlarge (4 vCPU, 8 GiB)
- c5a.2xlarge (8 vCPU, 16 GiB)
- c5a.4xlarge (16 vCPU, 32 GiB)
- c5a.8xlarge (32 vCPU, 64 GiB)
- c5a.12xlarge (48 vCPU, 96 GiB)
- c5a.16xlarge (64 vCPU, 128 GiB)
- c5a.24xlarge (96 vCPU, 192 GiB)
- c5ad.xlarge (4 vCPU, 8 GiB)
- c5ad.2xlarge (8 vCPU, 16 GiB)
- c5ad.4xlarge (16 vCPU, 32 GiB)
- c5ad.8xlarge (32 vCPU, 64 GiB)
- c5ad.12xlarge (48 vCPU, 96 GiB)
- c5ad.16xlarge (64 vCPU, 128 GiB)
- c5ad.24xlarge (96 vCPU, 192 GiB)
- c5n.metal (72 vCPU, 192 GiB)
- c5n.xlarge (4 vCPU, 10.5 GiB)
- c5n.2xlarge (8 vCPU, 21 GiB)
- c5n.4xlarge (16 vCPU, 42 GiB)
- c5n.9xlarge (36 vCPU, 96 GiB)
- c5n.18xlarge (72 vCPU, 192 GiB)
- c6a.xlarge (4 vCPU, 8 GiB)
- c6a.2xlarge (8 vCPU, 16 GiB)
- c6a.4xlarge (16 vCPU, 32 GiB)
- c6a.8xlarge (32 vCPU, 64 GiB)
- c6a.12xlarge (48 vCPU, 96 GiB)
- c6a.16xlarge (64 vCPU, 128 GiB)

- c6a.24xlarge (96 vCPU, 192 GiB)
- c6a.32xlarge (128 vCPU, 256 GiB)
- c6a.48xlarge (192 vCPU, 384 GiB)
- c6i.metal (128 vCPU, 256 GiB)
- c6i.xlarge (4 vCPU, 8 GiB)
- c6i.2xlarge (8 vCPU, 16 GiB)
- c6i.4xlarge (16 vCPU, 32 GiB)
- c6i.8xlarge (32 vCPU, 64 GiB)
- c6i.12xlarge (48 vCPU, 96 GiB)
- c6i.16xlarge (64 vCPU, 128 GiB)
- c6i.24xlarge (96 vCPU, 192 GiB)
- c6i.32xlarge (128 vCPU, 256 GiB)
- c6id.xlarge (4 vCPU, 8 GiB)
- c6id.2xlarge (8 vCPU, 16 GiB)
- c6id.4xlarge (16 vCPU, 32 GiB)
- c6id.8xlarge (32 vCPU, 64 GiB)
- c6id.12xlarge (48 vCPU, 96 GiB)
- c6id.16xlarge (64 vCPU, 128 GiB)
- c6id.24xlarge (96 vCPU, 192 GiB)
- c6id.32xlarge (128 vCPU, 256 GiB)

Example 2.7. Storage optimized

- i3.metal (72† vCPU, 512 GiB)
- i3.xlarge (4 vCPU, 30.5 GiB)
- i3.2xlarge (8 vCPU, 61 GiB)
- i3.4xlarge (16 vCPU, 122 GiB)
- i3.8xlarge (32 vCPU, 244 GiB)
- i3.16xlarge (64 vCPU, 488 GiB)
- i3en.metal (96 vCPU, 768 GiB)
- i3en.xlarge (4 vCPU, 32 GiB)

- i3en.2xlarge (8 vCPU, 64 GiB)
- i3en.3xlarge (12 vCPU, 96 GiB)
- i3en.6xlarge (24 vCPU, 192 GiB)
- i3en.12xlarge (48 vCPU, 384 GiB)
- i3en.24xlarge (96 vCPU, 768 GiB)
- i4i.xlarge (4 vCPU, 32 GiB)
- i4i.2xlarge (8 vCPU, 64 GiB)
- i4i.4xlarge (16 vCPU, 128 GiB)
- i4i.8xlarge (32 vCPU, 256 GiB)
- i4i.12xlarge (48 vCPU, 384 GiB)
- i4i.16xlarge (64 vCPU, 512 GiB)
- i4i.24xlarge (96 vCPU, 768 GiB)
- i4i.32xlarge (128 vCPU, 1024 GiB)
- i4i.metal (128 vCPU, 1024 GiB)

† This instance type provides 72 logical processors on 36 physical cores.

NOTE

Virtual instance types initialize faster than ".metal" instance types.

Example 2.8. High memory

- u-3tb1.56xlarge (224 vCPU, 3,072 GiB)
- u-6tb1.56xlarge (224 vCPU, 6,144 GiB)
- u-6tb1.112xlarge (448 vCPU, 6,144 GiB)
- u-6tb1.metal (448 vCPU, 6,144 GiB)
- u-9tb1.112xlarge (448 vCPU, 9,216 GiB)
- u-9tb1.metal (448 vCPU, 9,216 GiB)
- u-12tb1.112xlarge (448 vCPU, 12,288 GiB)
- u-12tb1.metal (448 vCPU, 12,288 GiB)
- u-18tb1.metal (448 vCPU, 18,432 GiB)
- u-24tb1.metal (448 vCPU, 24,576 GiB)

Additional Resources

- AWS Instance Types

2.1.1.6. AWS instance types for standard clusters

OpenShift Dedicated offers the following worker node types and sizes on AWS:

Example 2.9. General purpose

- m5.xlarge (4 vCPU, 16 GiB)
- m5.2xlarge (8 vCPU, 32 GiB)
- m5.4xlarge (16 vCPU, 64 GiB)

Example 2.10. Memory-optimized

- r5.xlarge (4 vCPU, 32 GiB)
- r5.2xlarge (8 vCPU, 64 GiB)
- r5.4xlarge (16 vCPU, 128 GiB)

Example 2.11. Compute-optimized

- c5.2xlarge (8 vCPU, 16 GiB)
- c5.4xlarge (16 vCPU, 32 GiB)

2.1.1.7. Google Cloud compute types

OpenShift Dedicated offers the following worker node types and sizes on Google Cloud that are chosen to have a common CPU and memory capacity that are the same as other cloud instance types:

NOTE

e2 and **a2** compute types are available for CCS only.

Example 2.12. General purpose

- custom-4-16384 (4 vCPU, 16 GiB)
- custom-8-32768 (8 vCPU, 32 GiB)
- custom-16-65536 (16 vCPU, 64 GiB)
- custom-32-131072 (32 vCPU, 128 GiB)
- custom-48-199608 (48 vCPU, 192 GiB)

- custom-64-262144 (64 vCPU, 256 GiB)
- custom-96-393216 (96 vCPU, 384 GiB)
- e2-standard-4 (4 vCPU, 16 GiB)
- n2-standard-4 (4 vCPU, 16 GiB)
- e2-standard-8 (8 vCPU, 32 GiB)
- n2-standard-8 (8 vCPU, 32 GiB)
- e2-standard-16 (16 vCPU, 64 GiB)
- n2-standard-16 (16 vCPU, 64 GiB)
- e2-standard-32 (32 vCPU, 128 GiB)
- n2-standard-32 (32 vCPU, 128 GiB)
- n2-standard-48 (48 vCPU, 192 GiB)
- n2-standard-64 (64 vCPU, 256 GiB)
- n2-standard-80 (80 vCPU, 320 GiB)
- n2-standard-96 (96 vCPU, 384 GiB)
- n2-standard-128 (128 vCPU, 512 GiB)

Example 2.13. Memory-optimized

- custom-4-32768-ext (4 vCPU, 32 GiB)
- custom-8-65536-ext (8 vCPU, 64 GiB)
- custom-16-131072-ext (16 vCPU, 128 GiB)
- e2-highmem-4 (4 vCPU, 32 GiB)
- e2-highmem-8 (8 vCPU, 64 GiB)
- e2-highmem-16 (16 vCPU, 128 GiB)
- n2-highmem-4 (4 vCPU, 32 GiB)
- n2-highmem-8 (8 vCPU, 64 GiB)
- n2-highmem-16 (16 vCPU, 128 GiB)
- n2-highmem-32 (32 vCPU, 256 GiB)
- n2-highmem-48 (48 vCPU, 384 GiB)
- n2-highmem-64 (64 vCPU, 512 GiB)
- n2-highmem-80 (80 vCPU, 640 GiB)

- n2-highmem-96 (96 vCPU, 768 GiB)
- n2-highmem-128 (128 vCPU, 864 GiB)

Example 2.14. Compute-optimized

- custom-8-16384 (8 vCPU, 16 GiB)
- custom-16-32768 (16 vCPU, 32 GiB)
- custom-36-73728 (36 vCPU, 72 GiB)
- custom-48-98304 (48 vCPU, 96 GiB)
- custom-72-147456 (72 vCPU, 144 GiB)
- custom-96-196608 (96 vCPU, 192 GiB)
- c2-standard-4 (4 vCPU, 16 GiB)
- c2-standard-8 (8 vCPU, 32 GiB)
- c2-standard-16 (16 vCPU, 64 GiB)
- c2-standard-30 (30 vCPU, 120 GiB)
- c2-standard-60 (60 vCPU, 240 GiB)
- e2-highcpu-8 (8 vCPU, 8 GiB)
- e2-highcpu-16 (16 vCPU, 16 GiB)
- e2-highcpu-32 (32 vCPU, 32 GiB)
- n2-highcpu-8 (8 vCPU, 8 GiB)
- n2-highcpu-16 (16 vCPU, 16 GiB)
- n2-highcpu-32 (32 vCPU, 32 GiB)
- n2-highcpu-48 (48 vCPU, 48 GiB)
- n2-highcpu-64 (64 vCPU, 64 GiB)
- n2-highcpu-80 (80 vCPU, 80 GiB)
- n2-highcpu-96 (96 vCPU, 96 GiB)

Example 2.15. Accelerated computing

- a2-highgpu-1g (12 vCPU, 85 GiB)
- a2-highgpu-2g (24 vCPU, 170 GiB)
- a2-highgpu-4g (48 vCPU, 340 GiB)

- a2-highgpu-8g (96 vCPU, 680 GiB)
- a2-megagpu-16g (96 vCPU, 1.33 TiB)

2.1.1.8. Regions and availability zones

The following AWS regions are supported by OpenShift Container Platform 4 and are supported for OpenShift Dedicated:

- af-south-1 (Cape Town, AWS opt-in required)
- ap-east-1 (Hong Kong, AWS opt-in required)
- ap-northeast-1 (Tokyo)
- ap-northeast-2 (Seoul)
- ap-northeast-3 (Osaka)
- ap-south-1 (Mumbai)
- ap-south-2 (Hyderabad, AWS opt-in required)
- ap-southeast-1 (Singapore)
- ap-southeast-2 (Sydney)
- ap-southeast-3 (Jakarta, AWS opt-in required)
- ap-southeast-4 (Melbourne, AWS opt-in required)
- ca-central-1 (Central Canada)
- eu-central-1 (Frankfurt)
- eu-central-2 (Zurich, AWS opt-in required)
- eu-north-1 (Stockholm)
- eu-south-1 (Milan, AWS opt-in required)
- eu-south-2 (Spain, AWS opt-in required)
- eu-west-1 (Ireland)
- eu-west-2 (London)
- eu-west-3 (Paris)
- me-central-1 (UAE, AWS opt-in required)
- me-south-1 (Bahrain, AWS opt-in required)
- sa-east-1 (São Paulo)
- us-east-1 (N. Virginia)

- us-east-2 (Ohio)
- us-west-1 (N. California)
- us-west-2 (Oregon)

The following Google Cloud regions are currently supported:

- asia-east1, Changhua County, Taiwan
- asia-east2, Hong Kong
- asia-northeast1, Tokyo, Japan
- asia-northeast2, Osaka, Japan
- asia-south1, Mumbai, India
- asia-south2, Delhi, India
- asia-southeast1, Jurong West, Singapore
- australia-southeast1, Sydney, Australia
- australia-southeast2, Melbourne, Australia
- europe-north1, Hamina, Finland
- europe-west1, St. Ghislain, Belgium
- europe-west2, London, England, UK
- europe-west3, Frankfurt, Germany
- europe-west4, Eemshaven, Netherlands
- europe-west6, Zürich, Switzerland
- europe-west8, Milan, Italy
- europe-west12, Turin, Italy
- europe-southwest1, Madrid, Spain
- northamerica-northeast1, Montréal, Québec, Canada
- southamerica-east1, Osasco (São Paulo), Brazil
- southamerica-west1, Santiago, Chile
- us-central1, Council Bluffs, Iowa, USA
- us-east1, Moncks Corner, South Carolina, USA
- us-east4, Ashburn, Northern Virginia, USA
- us-west1, The Dalles, Oregon, USA

- us-west2, Los Angeles, California, USA
- me-central1, Doha, Qatar
- me-central2, Dammam, Saudi Arabia

Multi-AZ clusters can only be deployed in regions with at least 3 availability zones (see AWS and Google Cloud).

Each new OpenShift Dedicated cluster is installed within a dedicated Virtual Private Cloud (VPC) in a single Region, with the option to deploy into a single Availability Zone (Single-AZ) or across multiple Availability Zones (Multi-AZ). This provides cluster-level network and resource isolation, and enables cloud-provider VPC settings, such as VPN connections and VPC Peering. Persistent volumes are backed by cloud block storage and are specific to the availability zone in which they are provisioned. Persistent volumes do not bind to a volume until the associated pod resource is assigned into a specific availability zone in order to prevent unschedulable pods. Availability zone-specific resources are only usable by resources in the same availability zone.

WARNING

The region and the choice of single or multi availability zone cannot be changed once a cluster has been deployed.

2.1.1.9. Service level agreement (SLA)

Any SLAs for the service itself are defined in Appendix 4 of the Red Hat Enterprise Agreement Appendix 4 (Online Subscription Services).

2.1.1.10. Limited support status

When a cluster transitions to a *Limited Support* status, Red Hat no longer proactively monitors the cluster, the SLA is no longer applicable, and credits requested against the SLA are denied. It does not mean that you no longer have product support. In some cases, the cluster can return to a fully-supported status if you remediate the violating factors. However, in other cases, you might have to delete and recreate the cluster.

A cluster might transition to a Limited Support status for many reasons, including the following scenarios:

If you do not upgrade a cluster to a supported version before the end-of-life date

Red Hat does not make any runtime or SLA guarantees for versions after their end-of-life date. To receive continued support, upgrade the cluster to a supported version prior to the end-of-life date. If you do not upgrade the cluster prior to the end-of-life date, the cluster transitions to a Limited Support status until it is upgraded to a supported version.

Red Hat provides commercially reasonable support to upgrade from an unsupported version to a supported version. However, if a supported upgrade path is no longer available, you might have to create a new cluster and migrate your workloads.

If you remove or replace any native OpenShift Dedicated components or any other component that is installed and managed by Red Hat

If cluster administrator permissions were used, Red Hat is not responsible for any of your or your authorized users' actions, including those that affect infrastructure services, service availability, or data loss. If Red Hat detects any such actions, the cluster might transition to a Limited Support status. Red Hat notifies you of the status change and you should either revert the action or create a support case to explore remediation steps that might require you to delete and recreate the cluster.

If you have questions about a specific action that might cause a cluster to transition to a Limited Support status or need further assistance, open a support ticket.

2.1.1.11. Support

OpenShift Dedicated includes Red Hat Premium Support, which can be accessed by using the Red Hat Customer Portal.

See the Scope of Coverage Page for more details on what is covered with included support for OpenShift Dedicated.

See OpenShift Dedicated SLAs for support response times.

2.1.2. Logging

OpenShift Dedicated provides optional integrated log forwarding to Amazon CloudWatch (on AWS) or Google Cloud Logging (on GCP).

For more information, see About log collection and forwarding.

2.1.2.1. Cluster audit logging

Cluster audit logs are available through Amazon CloudWatch (on AWS) or Google Cloud Logging (on GCP), if the integration is enabled. If the integration is not enabled, you can request the audit logs by opening a support case. Audit log requests must specify a date and time range not to exceed 21 days. When requesting audit logs, customers should be aware that audit logs are many GB per day in size.

2.1.2.2. Application logging

Application logs sent to **STDOUT** are forwarded to Amazon CloudWatch (on AWS) or Google Cloud Logging (on GCP) through the cluster logging stack, if it is installed.

2.1.3. Monitoring

2.1.3.1. Cluster metrics

OpenShift Dedicated clusters come with an integrated Prometheus/Grafana stack for cluster monitoring including CPU, memory, and network-based metrics. This is accessible through the web console and can also be used to view cluster-level status and capacity/usage through a Grafana dashboard. These metrics also allow for horizontal pod autoscaling based on CPU or memory metrics provided by an OpenShift Dedicated user.

2.1.3.2. Cluster notifications

Cluster notifications are messages about the status, health, or performance of your cluster.

Cluster notifications are the primary way that Red Hat Site Reliability Engineering (SRE) communicates with you about the health of your managed cluster. SRE may also use cluster notifications to prompt you to perform an action in order to resolve or prevent an issue with your cluster.

Cluster owners and administrators must regularly review and action cluster notifications to ensure clusters remain healthy and supported.

You can view cluster notifications in the Red Hat Hybrid Cloud Console, in the **Cluster history** tab for your cluster. By default, only the cluster owner receives cluster notifications as emails. If other users need to receive cluster notification emails, add each user as a notification contact for your cluster.

2.1.4. Networking

2.1.4.1. Custom domains for applications

WARNING

Starting with OpenShift Dedicated 4.14, the Custom Domain Operator is deprecated. To manage Ingress in OpenShift Dedicated 4.14 or later, use the Ingress Operator. The functionality is unchanged for OpenShift Dedicated 4.13 and earlier versions.

To use a custom hostname for a route, you must update your DNS provider by creating a canonical name (CNAME) record. Your CNAME record should map the OpenShift canonical router hostname to your custom domain. The OpenShift canonical router hostname is shown on the **Route Details** page after a Route is created. Alternatively, a wildcard CNAME record can be created once to route all subdomains for a given hostname to the cluster's router.

2.1.4.2. Custom domains for cluster services

Custom domains and subdomains are not available for the platform service routes, for example, the API or web console routes, or for the default application routes.

2.1.4.3. Domain validated certificates

OpenShift Dedicated includes TLS security certificates needed for both internal and external services on the cluster. For external routes, there are two, separate TLS wildcard certificates that are provided and installed on each cluster, one for the web console and route default hostnames and the second for the API endpoint. *Let's Encrypt* is the certificate authority used for certificates. Routes within the cluster, for example, the internal API endpoint, use TLS certificates signed by the cluster's built-in certificate authority and require the CA bundle available in every pod for trusting the TLS certificate.

2.1.4.4. Custom certificate authorities for builds

OpenShift Dedicated supports the use of custom certificate authorities to be trusted by builds when pulling images from an image registry.

2.1.4.5. Load balancers

OpenShift Dedicated uses up to 5 different load balancers:

- Internal control plane load balancer that is internal to the cluster and used to balance traffic for internal cluster communications.
- External control plane load balancer that is used for accessing the OpenShift Container Platform and Kubernetes APIs. This load balancer can be disabled in Red Hat OpenShift Cluster Manager. If this load balancer is disabled, Red Hat reconfigures the API DNS to point to the internal control load balancer.
- External control plane load balancer for Red Hat that is reserved for cluster management by Red Hat. Access is strictly controlled, and communication is only possible from allowlisted bastion hosts.
- Default router/ingress load balancer that is the default application load balancer, denoted by **apps** in the URL. The default load balancer can be configured in OpenShift Cluster Manager to be either publicly accessible over the internet, or only privately accessible over a pre-existing private connection. All application routes on the cluster are exposed on this default router load balancer, including cluster services such as the logging UI, metrics API, and registry.
- Optional: Secondary router/ingress load balancer that is a secondary application load balancer, denoted by **apps2** in the URL. The secondary load balancer can be configured in OpenShift Cluster Manager to be either publicly accessible over the internet, or only privately accessible over a pre-existing private connection. If a 'Label match' is configured for this router load balancer, then only application routes matching this label will be exposed on this router load balancer, otherwise all application routes are also exposed on this router load balancer.
- Optional: Load balancers for services that can be mapped to a service running on OpenShift Dedicated to enable advanced ingress features, such as non-HTTP/SNI traffic or the use of non-standard ports. These can be purchased in groups of 4 for standard clusters, or they can be provisioned without charge in Customer Cloud Subscription (CCS) clusters; however, each AWS account has a quota that limits the number of Classic Load Balancers that can be used within each cluster.

2.1.4.6. Network usage

For standard OpenShift Dedicated clusters, network usage is measured based on data transfer between inbound, VPC peering, VPN, and AZ traffic. On a standard OpenShift Dedicated base cluster, 12 TB of network I/O is provided. Additional network I/O can be purchased in 12 TB increments. For CCS OpenShift Dedicated clusters, network usage is not monitored, and is billed directly by the cloud provider.

2.1.4.7. Cluster ingress

Project administrators can add route annotations for many different purposes, including ingress control through IP allowlisting.

Ingress policies can also be changed by using **NetworkPolicy** objects, which leverage the **ovs-networkpolicy** plugin. This allows for full control over the ingress network policy down to the pod level, including between pods on the same cluster and even in the same namespace.

All cluster ingress traffic goes through the defined load balancers. Direct access to all nodes is blocked by cloud configuration.

2.1.4.8. Cluster egress

Pod egress traffic control through **EgressNetworkPolicy** objects can be used to prevent or limit outbound traffic in OpenShift Dedicated.

Public outbound traffic from the control plane and infrastructure nodes is required and necessary to maintain cluster image security and cluster monitoring. This requires the **0.0.0.0/0** route to belong only to the internet gateway; it is not possible to route this range over private connections.

OpenShift Dedicated clusters use NAT Gateways to present a public, static IP for any public outbound traffic leaving the cluster. Each subnet a cluster is deployed into receives a distinct NAT Gateway. For clusters deployed on AWS with multiple availability zones, up to 3 unique static IP addresses can exist for cluster egress traffic. For clusters deployed on Google Cloud, regardless of availability zone topology, there will by 1 static IP address for worker node egress traffic. Any traffic that remains inside the cluster or does not go out to the public internet will not pass through the NAT Gateway and will have a source IP address belonging to the node that the traffic originated from. Node IP addresses are dynamic, and therefore a customer should not rely on allowlisting individual IP address when accessing private resources.

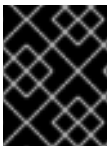
Customers can determine their public static IP addresses by running a pod on the cluster and then querying an external service. For example:

```
$ oc run ip-lookup --image=busybox -i -t --restart=Never --rm -- /bin/sh -c "/bin/nslookup -type=a
myip.opendns.com resolver1.opendns.com | grep -E 'Address: [0-9.]+'"
```

2.1.4.9. Cloud network configuration

OpenShift Dedicated allows for the configuration of a private network connection through several cloud provider managed technologies:

- VPN connections
- AWS VPC peering
- AWS Transit Gateway
- AWS Direct Connect
- Google Cloud VPC Network peering
- Google Cloud Classic VPN
- Google Cloud HA VPN



IMPORTANT

Red Hat SREs do not monitor private network connections. Monitoring these connections is the responsibility of the customer.

2.1.4.10. DNS forwarding

For OpenShift Dedicated clusters that have a private cloud network configuration, a customer can specify internal DNS servers available on that private connection that should be queried for explicitly provided domains.

2.1.4.11. Network verification

Network verification checks run automatically when you deploy an OpenShift Dedicated cluster into an existing Virtual Private Cloud (VPC) or create an additional machine pool with a subnet that is new to your cluster. The checks validate your network configuration and highlight errors, enabling you to resolve configuration issues prior to deployment.

You can also run the network verification checks manually to validate the configuration for an existing cluster.

Additional resources

- For more information about the network verification checks, see Network verification.

2.1.5. Storage

2.1.5.1. Encrypted-at-rest OS/node storage

Control plane nodes use encrypted-at-rest-EBS storage.

2.1.5.2. Encrypted-at-rest PV

EBS volumes used for persistent volumes (PVs) are encrypted-at-rest by default.

2.1.5.3. Block storage (RWO)

Persistent volumes (PVs) are backed by AWS EBS and Google Cloud persistent disk block storage, which uses the ReadWriteOnce (RWO) access mode. On a standard OpenShift Dedicated base cluster, 100 GB of block storage is provided for PVs, which is dynamically provisioned and recycled based on application requests. Additional persistent storage can be purchased in 500 GB increments.

PVs can only be attached to a single node at a time and are specific to the availability zone in which they were provisioned, but they can be attached to any node in the availability zone.

Each cloud provider has its own limits for how many PVs can be attached to a single node. See AWS instance type limits or Google Cloud Platform custom machine types for details.

2.1.5.4. Shared storage (RWX)

The AWS CSI Driver can be used to provide RWX support for OpenShift Dedicated on AWS. A community Operator is provided to simplify setup. See AWS EFS Setup for OpenShift Dedicated and Red Hat OpenShift Service on AWS for details.

2.1.6. Platform

2.1.6.1. Cluster backup policy

IMPORTANT

It is critical that customers have a backup plan for their applications and application data.

Application and application data backups are not a part of the OpenShift Dedicated service. All Kubernetes objects in each OpenShift Dedicated cluster are backed up to facilitate a prompt recovery in the unlikely event that a cluster becomes irreparably inoperable.

The backups are stored in a secure object storage (Multi-AZ) bucket in the same account as the cluster. Node root volumes are not backed up because Red Hat Enterprise Linux CoreOS is fully managed by the OpenShift Container Platform cluster and no stateful data should be stored on the root volume of a node.

The following table shows the frequency of backups:

Component	Snapshot Frequency	Retention	Notes
Full object store backup	Daily at 0100 UTC	7 days	This is a full backup of all Kubernetes objects. No persistent volumes (PVs) are backed up in this backup schedule.
Full object store backup	Weekly on Mondays at 0200 UTC	30 days	This is a full backup of all Kubernetes objects. No PVs are backed up in this backup schedule.
Full object store backup	Hourly at 17 minutes past the hour	24 hours	This is a full backup of all Kubernetes objects. No PVs are backed up in this backup schedule.

2.1.6.2. Autoscaling

Node autoscaling is available on OpenShift Dedicated. See About autoscaling nodes on a cluster for more information on autoscaling nodes on a cluster.

2.1.6.3. Daemon sets

Customers may create and run DaemonSets on OpenShift Dedicated. In order to restrict DaemonSets to only running on worker nodes, use the following nodeSelector:

```
...
spec:
  nodeSelector:
    role: worker
...
```

2.1.6.4. Multiple availability zone

In a multiple availability zone cluster, control nodes are distributed across availability zones and at least three worker nodes are required in each availability zone.

2.1.6.5. Node labels

Custom node labels are created by Red Hat during node creation and cannot be changed on OpenShift Dedicated clusters at this time.

2.1.6.6. OpenShift version

OpenShift Dedicated is run as a service and is kept up to date with the latest OpenShift Container Platform version.

2.1.6.7. Upgrades

Refer to OpenShift Dedicated Life Cycle for more information on the upgrade policy and procedures.

2.1.6.8. Windows containers

Windows containers are not available on OpenShift Dedicated at this time.

2.1.6.9. Container engine

OpenShift Dedicated runs on OpenShift 4 and uses CRI-O as the only available container engine.

2.1.6.10. Operating system

OpenShift Dedicated runs on OpenShift 4 and uses Red Hat Enterprise Linux CoreOS as the operating system for all control plane and worker nodes.

2.1.6.11. Red Hat Operator support

Red Hat workloads typically refer to Red Hat-provided Operators made available through Operator Hub. Red Hat workloads are not managed by the Red Hat SRE team, and must be deployed on worker nodes. These Operators may require additional Red Hat subscriptions, and may incur additional cloud infrastructure costs. Examples of these Red Hat-provided Operators are:

- Red Hat Quay
- Red Hat Advanced Cluster Management
- Red Hat Advanced Cluster Security
- Red Hat OpenShift Service Mesh
- OpenShift Serverless
- Red Hat OpenShift Logging
- Red Hat OpenShift Pipelines

2.1.6.12. Kubernetes Operator support

All Operators listed in the OperatorHub marketplace should be available for installation. Operators installed from OperatorHub, including Red Hat Operators, are not SRE managed as part of the OpenShift Dedicated service. Refer to the Red Hat Customer Portal for more information on the supportability of a given Operator.

2.1.7. Security

This section provides information about the service definition for OpenShift Dedicated security.

2.1.7.1. Authentication provider

Authentication for the cluster is configured as part of Red Hat OpenShift Cluster Manager cluster creation process. OpenShift is not an identity provider, and all access to the cluster must be managed by the customer as part of their integrated solution. Provisioning multiple identity providers provisioned at the same time is supported. The following identity providers are supported:

- GitHub or GitHub Enterprise OAuth
- GitLab OAuth
- Google OAuth
- LDAP
- OpenID connect

2.1.7.2. Privileged containers

Privileged containers are not available by default on OpenShift Dedicated. The **anyuid** and **nonroot** Security Context Constraints are available for members of the **dedicated-admins** group, and should address many use cases. Privileged containers are only available for **cluster-admin** users.

2.1.7.3. Customer administrator user

In addition to normal users, OpenShift Dedicated provides access to an OpenShift Dedicated-specific group called **dedicated-admin**. Any users on the cluster that are members of the **dedicated-admin** group:

- Have administrator access to all customer-created projects on the cluster.
- Can manage resource quotas and limits on the cluster.
- Can add and manage **NetworkPolicy** objects.
- Are able to view information about specific nodes and PVs in the cluster, including scheduler information.
- Can access the reserved **dedicated-admin** project on the cluster, which allows for the creation of service accounts with elevated privileges and also gives the ability to update default limits and quotas for projects on the cluster.
- Can install Operators from OperatorHub (***** verbs in all ***.operators.coreos.com** API groups).

2.1.7.4. Cluster administration role

As an administrator of OpenShift Dedicated with Customer Cloud Subscriptions (CCS), you have access to the **cluster-admin** role. While logged in to an account with the **cluster-admin** role, users have mostly unrestricted access to control and configure the cluster. There are some configurations that are blocked with webhooks to prevent destablizing the cluster, or because they are managed in OpenShift Cluster Manager and any in-cluster changes would be overwritten.

2.1.7.5. Project self-service

All users, by default, have the ability to create, update, and delete their projects. This can be restricted if a member of the **dedicated-admin** group removes the self-provisioner role from authenticated users:

```
$ oc adm policy remove-cluster-role-from-group self-provisioner system:authenticated:oauth
```

Restrictions can be reverted by applying:

```
$ oc adm policy add-cluster-role-to-group self-provisioner system:authenticated:oauth
```

2.1.7.6. Regulatory compliance

OpenShift Dedicated follows common industry best practices for security and controls. The certifications are outlined in the following table.

Table 2.1. Security and control certifications for OpenShift Dedicated

Compliance	OpenShift Dedicated on AWS	OpenShift Dedicated on GCP
HIPAA Qualified	Yes (Only Customer Cloud Subscriptions)	Yes (Only Customer Cloud Subscriptions)
ISO 27001	Yes	Yes
PCI DSS	Yes	Yes
SOC 2 Type 2	Yes	Yes

2.1.7.7. Network security

Each OpenShift Dedicated cluster is protected by a secure network configuration at the cloud infrastructure level using firewall rules (AWS Security Groups or Google Cloud Compute Engine firewall rules). OpenShift Dedicated customers on AWS are also protected against DDoS attacks with AWS Shield Standard. Similarly, all GCP load balancers and public IP addresses used by OpenShift Dedicated on GCP are protected against DDoS attacks with Google Cloud Armor Standard .

2.1.7.8. etcd encryption

In OpenShift Dedicated, the control plane storage is encrypted at rest by default and this includes encryption of the etcd volumes. This storage-level encryption is provided through the storage layer of the cloud provider.

You can also enable etcd encryption, which encrypts the key values in etcd, but not the keys. If you enable etcd encryption, the following Kubernetes API server and OpenShift API server resources are encrypted:

- Secrets
- Config maps
- Routes
- OAuth access tokens
- OAuth authorize tokens

The etcd encryption feature is not enabled by default and it can be enabled only at cluster installation time. Even with etcd encryption enabled, the etcd key values are accessible to anyone with access to the control plane nodes or **cluster-admin** privileges.

IMPORTANT

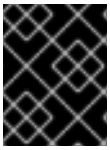
By enabling etcd encryption for the key values in etcd, you will incur a performance overhead of approximately 20%. The overhead is a result of introducing this second layer of encryption, in addition to the default control plane storage encryption that encrypts the etcd volumes. Red Hat recommends that you enable etcd encryption only if you specifically require it for your use case.

2.2. RESPONSIBILITY ASSIGNMENT MATRIX

Understanding the Red Hat, cloud provider, and customer responsibilities for the OpenShift Dedicated managed service.

2.2.1. Overview of responsibilities for OpenShift Dedicated

While Red Hat manages the OpenShift Dedicated service, the customer shares responsibility with respect to certain aspects. The OpenShift Dedicated services are accessed remotely, hosted on public cloud resources, created in either Red Hat or customer-owned cloud service provider accounts, and have underlying platform and data security that is owned by Red Hat.



IMPORTANT

If the **cluster-admin** role is enabled on a cluster, see the responsibilities and exclusion notes in the Red Hat Enterprise Agreement Appendix 4 (Online Subscription Services) .

Resource	Incident and operations management	Change management	Identity and access management	Security and regulation compliance	Disaster recovery
Customer data	Customer	Customer	Customer	Customer	Customer
Customer applications	Customer	Customer	Customer	Customer	Customer
Developer services	Customer	Customer	Customer	Customer	Customer
Platform monitoring	Red Hat	Red Hat	Red Hat	Red Hat	Red Hat
Logging	Red Hat	Shared	Shared	Shared	Red Hat
Application networking	Shared	Shared	Shared	Red Hat	Red Hat
Cluster networking	Red Hat	Shared	Shared	Red Hat	Red Hat
Virtual networking	Shared	Shared	Shared	Shared	Shared

Resource	Incident and operations management	Change management	Identity and access management	Security and regulation compliance	Disaster recovery
Control plane and infrastructure nodes	Red Hat	Red Hat	Red Hat	Red Hat	Red Hat
Worker nodes	Red Hat	Red Hat	Red Hat	Red Hat	Red Hat
Cluster version	Red Hat	Shared	Red Hat	Red Hat	Red Hat
Capacity management	Red Hat	Shared	Red Hat	Red Hat	Red Hat
Virtual storage	Red Hat and cloud provider	Red Hat and cloud provider	Red Hat and cloud provider	Red Hat and cloud provider	Red Hat and cloud provider
Physical infrastructure and security	Cloud provider	Cloud provider	Cloud provider	Cloud provider	Cloud provider

2.2.2. Shared responsibility matrix

The customer and Red Hat share responsibility for the monitoring and maintenance of an OpenShift Dedicated cluster. This documentation illustrates the delineation of responsibilities by area and task.

2.2.2.1. Incident and operations management

The customer is responsible for incident and operations management of customer application data and any custom networking the customer has configured for the cluster network or virtual network.

Resource	Red Hat responsibilities	Customer responsibilities
Application networking	Monitor cloud load balancers and native OpenShift router service, and respond to alerts.	• Monitor health of service load balancer endpoints • Monitor health of application routes, and the endpoints behind them. • Report outages to Red Hat.
Virtual networking	Monitor cloud load balancers, subnets, and public cloud components necessary for default platform networking, and respond to alerts.	Monitor network traffic that is optionally configured through VPC to VPC connection, VPN connection, or Direct connection for potential issues or security threats.

2.2.2.2. Change management

Red Hat is responsible for enabling changes to the cluster infrastructure and services that the customer will control, as well as maintaining versions for the control plane nodes, infrastructure nodes and services, and worker nodes. The customer is responsible for initiating infrastructure change requests and installing and maintaining optional services and networking configurations on the cluster, as well as all changes to customer data and customer applications.

Resource	Red Hat responsibilities	Customer responsibilities
Logging	• Centrally aggregate and monitor platform audit logs. • Provide and maintain a logging operator to enable the customer to deploy a logging stack for default application logging. • Provide audit logs upon customer request.	• Install the optional default application logging operator on the cluster. • Install, configure, and maintain any optional app logging solutions, such as logging sidecar containers or third-party logging applications. • Tune size and frequency of application logs being produced by customer applications if they are affecting the stability of the logging stack or the cluster. • Request platform audit logs through a support case for researching specific incidents.
Application networking	• Set up public cloud load balancers. Provide the ability to set up private load balancers and up to one additional load balancer when required. • Set up native OpenShift router service. Provide the ability to set the router as private and add up to one additional router shard. • Install, configure, and maintain OpenShift SDN components for default internal pod traffic. • Provide the ability for the customer to manage **NetworkPolicy** and **EgressNetworkPolicy** (firewall) objects.	• Configure non-default pod network permissions for project and pod networks, pod ingress, and pod egress using **NetworkPolicy** objects. • Use Red Hat OpenShift Cluster Manager to request a private load balancer for default application routes. • Use OpenShift Cluster Manager to configure up to one additional public or private router shard and corresponding load balancer. • Request and configure any additional service load balancers for specific services. • Configure any necessary DNS forwarding rules.

Resource	Red Hat responsibilities	Customer responsibilities
Cluster networking	• Set up cluster management components, such as public or private service endpoints and necessary integration with virtual networking components. • Set up internal networking components required for internal cluster communication between worker, infrastructure, and control plane nodes.	• Provide optional non-default IP address ranges for machine CIDR, service CIDR, and pod CIDR if needed through OpenShift Cluster Manager when the cluster is provisioned. • Request that the API service endpoint be made public or private on cluster creation or after cluster creation through OpenShift Cluster Manager.
Virtual networking	• Set up and configure virtual networking components required to provision the cluster, including virtual private cloud, subnets, load balancers, internet gateways, NAT gateways, etc. • Provide the ability for the customer to manage VPN connectivity with on-premises resources, VPC to VPC connectivity, and Direct connectivity as required through OpenShift Cluster Manager. • Enable customers to create and deploy public cloud load balancers for use with service load balancers.	• Set up and maintain optional public cloud networking components, such as VPC to VPC connection, VPN connection, or Direct connection. • Request and configure any additional service load balancers for specific services.
Cluster version	• Enable upgrade scheduling process. • Monitor upgrade progress and remedy any issues encountered. • Publish changelogs and release notes for minor and maintenance upgrades.	• Schedule maintenance version upgrades either immediately, for the future, or have automatic upgrades. • Acknowledge and schedule minor version upgrades. • Ensure the cluster version stays on a supported minor version. • Test customer applications on minor and maintenance versions to ensure compatibility.

Resource	Red Hat responsibilities	Customer responsibilities
Capacity management	• Monitor utilization of control plane (control plane nodes and infrastructure nodes). • Scale or resize control plane nodes to maintain quality of service. • Monitor utilization of customer resources including Network, Storage and Compute capacity. Where autoscaling features are not enabled alert customer for any changes required to cluster resources (for example, new compute nodes to scale, additional storage, etc).	• Use the provided OpenShift Cluster Manager controls to add or remove additional worker nodes as required. • Respond to Red Hat notifications regarding cluster resource requirements.

2.2.2.3. Access and identity authorization

The access and identity authorization matrix includes responsibilities for managing authorized access to clusters, applications, and infrastructure resources. This includes tasks such as providing access control mechanisms, authentication, authorization, and managing access to resources.

Resource	Red Hat responsibilities	Customer responsibilities
Logging	• Adhere to an industry standards-based tiered internal access process for platform audit logs. • Provide native OpenShift RBAC capabilities.	• Configure OpenShift RBAC to control access to projects and by extension a project's application logs. • For third-party or custom application logging solutions, the customer is responsible for access management.
Application networking	Provide native OpenShift RBAC and **dedicated-admin** capabilities.	• Configure OpenShift dedicated-admins and RBAC to control access to route configuration as required. • Manage Org Admins for Red Hat organization to grant access to OpenShift Cluster Manager. OpenShift Cluster Manager is used to configure router options and provide service load balancer quota.

Resource	Red Hat responsibilities	Customer responsibilities
Cluster networking	• Provide customer access controls through OpenShift Cluster Manager. • Provide native OpenShift RBAC and **dedicated-admin** capabilities.	• Manage Red Hat organization membership of Red Hat accounts. • Manage Org Admins for Red Hat organization to grant access to OpenShift Cluster Manager. • Configure OpenShift dedicated-admins and RBAC to control access to route configuration as required.
Virtual networking	Provide customer access controls through OpenShift Cluster Manager.	Manage optional user access to public cloud components through OpenShift Cluster Manager.

2.2.2.4. Security and regulation compliance

The following are the responsibilities and controls related to compliance:

Resource	Red Hat responsibilities	Customer responsibilities
Logging	Send cluster audit logs to a Red Hat SIEM to analyze for security events. Retain audit logs for a defined period of time to support forensic analysis.	Analyze application logs for security events. Send application logs to an external endpoint through logging sidecar containers or third-party logging applications if longer retention is required than is offered by the default logging stack.
Virtual networking	• Monitor virtual networking components for potential issues and security threats. • Leverage additional public cloud provider tools for additional monitoring and protection.	• Monitor optionally-configured virtual networking components for potential issues and security threats. • Configure any necessary firewall rules or data center protections as required.

2.2.2.5. Disaster recovery

Disaster recovery includes data and configuration backup, replicating data and configuration to the disaster recovery environment, and failover on disaster events.

Resource	Red Hat responsibilities	Customer responsibilities
Virtual networking	Restore or recreate affected virtual network components that are necessary for the platform to function.	• Configure virtual networking connections with more than one tunnel where possible for protection against outages as recommended by the public cloud provider. • Maintain failover DNS and load balancing if using a global load balancer with multiple clusters.

2.2.3. Customer responsibilities for data and applications

The customer is responsible for the applications, workloads, and data that they deploy to OpenShift Dedicated. However, Red Hat provides various tools to help the customer manage data and applications on the platform.

Resource	Red Hat responsibilities	Customer responsibilities
Customer data	• Maintain platform-level standards for data encryption. • Provide OpenShift components to help manage application data, such as secrets. • Enable integration with third-party data services (such as AWS RDS or Google Cloud SQL) to store and manage data outside of the cluster and/or cloud provider.	Maintain responsibility for all customer data stored on the platform and how customer applications consume and expose this data.

Resource	Red Hat responsibilities	Customer responsibilities
Customer applications	• Provision clusters with OpenShift components installed so that customers can access the OpenShift and Kubernetes APIs to deploy and manage containerized applications. • Create clusters with image pull secrets so that customer deployments can pull images from the Red Hat Container Catalog registry. • Provide access to OpenShift APIs that a customer can use to set up Operators to add community, third-party, and Red Hat services to the cluster. • Provide storage classes and plugins to support persistent volumes for use with customer applications. • Provide a container image registry so customers can securely store application container images on the cluster to deploy and manage applications.	• Maintain responsibility for customer and third-party applications, data, and their complete lifecycle. • If a customer adds Red Hat, community, third-party, their own, or other services to the cluster by using Operators or external images, the customer is responsible for these services and for working with the appropriate provider (including Red Hat) to troubleshoot any issues. • Use the provided tools and features to configure and deploy; keep up-to-date; set up resource requests and limits; size the cluster to have enough resources to run apps; set up permissions; integrate with other services; manage any image streams or templates that the customer deploys; externally serve; save, back up, and restore data; and otherwise manage their highly available and resilient workloads. • Maintain responsibility for monitoring the applications run on OpenShift Dedicated; including installing and operating software to gather metrics and create alerts.

2.3. UNDERSTANDING PROCESS AND SECURITY FOR OPENSHIFT DEDICATED

2.3.1. Review and action cluster notifications

Cluster notifications are messages about the status, health, or performance of your cluster.

Cluster notifications are the primary way that Red Hat Site Reliability Engineering (SRE) communicates with you about the health of your managed cluster. SRE may also use cluster notifications to prompt you to perform an action in order to resolve or prevent an issue with your cluster.

Cluster owners and administrators must regularly review and action cluster notifications to ensure clusters remain healthy and supported.

You can view cluster notifications in the Red Hat Hybrid Cloud Console, in the **Cluster history** tab for your cluster. By default, only the cluster owner receives cluster notifications as emails. If other users need to receive cluster notification emails, add each user as a notification contact for your cluster.

2.3.1.1. Cluster notification policy

Cluster notifications are designed to keep you informed about the health of your cluster and high impact events that affect it.

Most cluster notifications are generated and sent automatically to ensure that you are immediately informed of problems or important changes to the state of your cluster.

In certain situations, Red Hat Site Reliability Engineering (SRE) creates and sends cluster notifications to provide additional context and guidance for a complex issue.

Cluster notifications are not sent for low-impact events, low-risk security updates, routine operations and maintenance, or minor, transient issues that are quickly resolved by SRE.

Red Hat services automatically send notifications when:

- Remote health monitoring or environment verification checks detect an issue in your cluster, for example, when a worker node has low disk space.
- Significant cluster life cycle events occur, for example, when scheduled maintenance or upgrades begin, or cluster operations are impacted by an event, but do not require customer intervention.
- Significant cluster management changes occur, for example, when cluster ownership or administrative control is transferred from one user to another.
- Your cluster subscription is changed or updated, for example, when Red Hat makes updates to subscription terms or features available to your cluster.

SRE creates and sends notifications when:

- An incident results in a degradation or outage that impacts your cluster's availability or performance, for example, your cloud provider has a regional outage. SRE sends subsequent notifications to inform you of incident resolution progress, and when the incident is resolved.
- A security vulnerability, security breach, or unusual activity is detected on your cluster.
- Red Hat detects that changes you have made are creating or may result in cluster instability.
- Red Hat detects that your workloads are causing performance degradation or instability in your cluster.

2.3.2. Incident and operations management

This documentation details the Red Hat responsibilities for the OpenShift Dedicated managed service. The cloud provider is responsible for protecting the hardware infrastructure that runs the services offered by the cloud provider. The customer is responsible for incident and operations management of customer application data and any custom networking the customer has configured for the cluster network or virtual network.

2.3.2.1. Platform monitoring

A Red Hat Site Reliability Engineer (SRE) maintains a centralized monitoring and alerting system for all OpenShift Dedicated cluster components, SRE services, and underlying cloud provider accounts. Platform audit logs are securely forwarded to a centralized SIEM (Security Information and Event Monitoring) system, where they might trigger configured alerts to the SRE team and are also subject to manual review. Audit logs are retained in the SIEM for one year. Audit logs for a given cluster are not deleted at the time the cluster is deleted.

2.3.2.2. Incident management

An incident is an event that results in a degradation or outage of one or more Red Hat services. An incident can be raised by a customer or Customer Experience and Engagement (CEE) member through a support case, directly by the centralized monitoring and alerting system, or directly by a member of the SRE team.

Depending on the impact on the service and customer, the incident is categorized in terms of severity.

The general workflow of how a new incident is managed by Red Hat:

1. An SRE first responder is alerted to a new incident, and begins an initial investigation.
2. After the initial investigation, the incident is assigned an incident lead, who coordinates the recovery efforts.
3. The incident lead manages all communication and coordination around recovery, including any relevant notifications or support case updates.
4. The incident is recovered.
5. The incident is documented and a root cause analysis is performed within 5 business days of the incident.
6. A root cause analysis (RCA) draft document is shared with the customer within 7 business days of the incident.

2.3.2.3. Backup and recovery

All OpenShift Dedicated clusters are backed up using cloud provider snapshots. Notably, this does not include customer data stored on persistent volumes (PVs). All snapshots are taken using the appropriate cloud provider snapshot APIs and are uploaded to a secure object storage bucket (S3 in AWS, and GCS in Google Cloud) in the same account as the cluster.

Component	Snapshot frequency	Retention	Notes
Full object store backup	Daily	7 days	This is a full backup of all Kubernetes objects like etcd. No PVs are backed up in this backup schedule.
	Weekly	30 days	
Full object store backup	Hourly	24 hour	This is a full backup of all Kubernetes objects like etcd. No PVs are backed up in this backup schedule.

Component	Snapshot frequency	Retention	Notes
Node root volume	Never	N/A	Nodes are considered to be short-term. Nothing critical should be stored on a node's root volume.

- Red Hat does not commit to any Recovery Point Objective (RPO) or Recovery Time Objective (RTO).
- Customers are responsible for taking regular backups of their data
- Customers should deploy multi-AZ clusters with workloads that follow Kubernetes best practices to ensure high availability within a region.
- If an entire cloud region is unavailable, customers must install a new cluster in a different region and restore their apps using their backup data.

2.3.2.4. Cluster capacity

Evaluating and managing cluster capacity is a responsibility that is shared between Red Hat and the customer. Red Hat SRE is responsible for the capacity of all control plane and infrastructure nodes on the cluster.

Red Hat SRE also evaluates cluster capacity during upgrades and in response to cluster alerts. The impact of a cluster upgrade on capacity is evaluated as part of the upgrade testing process to ensure that capacity is not negatively impacted by new additions to the cluster. During a cluster upgrade, additional worker nodes are added to make sure that total cluster capacity is maintained during the upgrade process.

Capacity evaluations by SRE staff also happen in response to alerts from the cluster, once usage thresholds are exceeded for a certain period of time. Such alerts can also result in a notification to the customer.

2.3.3. Change management

This section describes the policies about how cluster and configuration changes, patches, and releases are managed.

2.3.3.1. Customer-initiated changes

You can initiate changes using self-service capabilities such as cluster deployment, worker node scaling, or cluster deletion.

Change history is captured in the **Cluster History** section in the OpenShift Cluster Manager **Overview tab**, and is available for you to view. The change history includes, but is not limited to, logs from the following changes:

- Adding or removing identity providers

- Adding or removing users to or from the **dedicated-admins** group
- Scaling the cluster compute nodes
- Scaling the cluster load balancer
- Scaling the cluster persistent storage
- Upgrading the cluster

You can implement a maintenance exclusion by avoiding changes in OpenShift Cluster Manager for the following components:

- Deleting a cluster
- Adding, modifying, or removing identity providers
- Adding, modifying, or removing a user from an elevated group
- Installing or removing add-ons
- Modifying cluster networking configurations
- Adding, modifying, or removing machine pools
- Enabling or disabling user workload monitoring
- Initiating an upgrade

IMPORTANT

To enforce the maintenance exclusion, ensure machine pool autoscaling or automatic upgrade policies have been disabled. After the maintenance exclusion has been lifted, proceed with enabling machine pool autoscaling or automatic upgrade policies as desired.

2.3.3.2. Red Hat-initiated changes

Red Hat site reliability engineering (SRE) manages the infrastructure, code, and configuration of OpenShift Dedicated using a GitOps workflow and fully automated CI/CD pipelines. This process ensures that Red Hat can safely introduce service improvements on a continuous basis without negatively impacting customers.

Every proposed change undergoes a series of automated verifications immediately upon check-in. Changes are then deployed to a staging environment where they undergo automated integration testing. Finally, changes are deployed to the production environment. Each step is fully automated.

An authorized SRE reviewer must approve advancement to each step. The reviewer cannot be the same individual who proposed the change. All changes and approvals are fully auditable as part of the GitOps workflow.

Some changes are released to production incrementally, using feature flags to control availability of new features to specified clusters or customers.

2.3.3.3. Patch management

OpenShift Container Platform software and the underlying immutable Red Hat Enterprise Linux CoreOS (RHCOS) operating system image are patched for bugs and vulnerabilities in regular z-stream upgrades. Read more about RHCOS architecture in the OpenShift Container Platform documentation.

2.3.3.4. Release management

Red Hat does not automatically upgrade your clusters. You can schedule to upgrade the clusters at regular intervals (recurring upgrade) or just once (individual upgrade) using the OpenShift Cluster Manager web console. Red Hat might forcefully upgrade a cluster to a new z-stream version only if the cluster is affected by a critical impact CVE. You can review the history of all cluster upgrade events in the OpenShift Cluster Manager web console. For more information about releases, see the Life Cycle policy.

2.3.4. Security and regulation compliance

Security and regulation compliance includes tasks, such as the implementation of security controls and compliance certification.

2.3.4.1. Data classification

Red Hat defines and follows a data classification standard to determine the sensitivity of data and highlight inherent risk to the confidentiality and integrity of that data while it is collected, used, transmitted stored, and processed. Customer-owned data is classified at the highest level of sensitivity and handling requirements.

2.3.4.2. Data management

OpenShift Dedicated uses cloud provider services such as AWS Key Management Service (KMS) and Google Cloud KMS to help securely manage encryption keys for persistent data. These keys are used for encrypting all control plane, infrastructure, and worker node root volumes. Customers can specify their own KMS key for encrypting root volumes at installation time. Persistent volumes (PVs) also use KMS for key management. Customers can specify their own KMS key for encrypting PVs by creating a new **StorageClass** referencing the KMS key Amazon Resource Name (ARN) or ID.

When a customer deletes their OpenShift Dedicated cluster, all cluster data is permanently deleted, including control plane data volumes and customer application data volumes, such a persistent volumes (PV).

2.3.4.3. Vulnerability management

Red Hat performs periodic vulnerability scanning of OpenShift Dedicated using industry standard tools. Identified vulnerabilities are tracked to their remediation according to timelines based on severity. Vulnerability scanning and remediation activities are documented for verification by third-party assessors in the course of compliance certification audits.

2.3.4.4. Network security

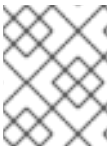
2.3.4.4.1. Firewall and DDoS protection

Each OpenShift Dedicated cluster is protected by a secure network configuration at the cloud infrastructure level using firewall rules (AWS Security Groups or Google Cloud Compute Engine firewall rules). OpenShift Dedicated customers on AWS are also protected against DDoS attacks with AWS Shield Standard. Similarly, all GCP load balancers and public IP addresses used by OpenShift Dedicated on GCP are protected against DDoS attacks with Google Cloud Armor Standard.

2.3.4.4.2. Private clusters and network connectivity

Customers can optionally configure their OpenShift Dedicated cluster endpoints (web console, API, and application router) to be made private so that the cluster control plane or applications are not accessible from the Internet.

For AWS, customers can configure a private network connection to their OpenShift Dedicated cluster through AWS VPC peering, AWS VPN, or AWS Direct Connect.



NOTE

At this time, private clusters are not supported for OpenShift Dedicated clusters on Google Cloud.

2.3.4.4.3. Cluster network access controls

Fine-grained network access control rules can be configured by customers per project by using **NetworkPolicy** objects and the OpenShift SDN.

2.3.4.5. Penetration testing

Red Hat performs periodic penetration tests against OpenShift Dedicated. Tests are performed by an independent internal team using industry standard tools and best practices.

Any issues that are discovered are prioritized based on severity. Any issues found belonging to open source projects are shared with the community for resolution.

2.3.4.6. Compliance

OpenShift Dedicated follows common industry best practices for security and controls. The certifications are outlined in the following table.

Table 2.2. Security and control certifications for OpenShift Dedicated

Compliance	OpenShift Dedicated on AWS	OpenShift Dedicated on GCP
HIPAA Qualified	Yes (Only Customer Cloud Subscriptions)	Yes (Only Customer Cloud Subscriptions)
ISO 27001	Yes	Yes
PCI DSS	Yes	Yes
SOC 2 Type 2	Yes	Yes

Additional resources

- See Red Hat Subprocessor List for information on SRE residency.

2.3.5. Disaster recovery

OpenShift Dedicated provides disaster recovery for failures that occur at the pod, worker node, infrastructure node, control plane node, and availability zone levels.

All disaster recovery requires that the customer use best practices for deploying highly available applications, storage, and cluster architecture (for example, single-zone deployment vs. multi-zone deployment) to account for the level of desired availability.

One single-zone cluster will not provide disaster avoidance or recovery in the event of an availability zone or region outage. Multiple single-zone clusters with customer-maintained failover can account for outages at the zone or region levels.

One multi-zone cluster will not provide disaster avoidance or recovery in the event of a full region outage. Multiple multi-zone clusters with customer-maintained failover can account for outages at the region level.

2.3.6. Additional resources

- For more information about Red Hat site reliability engineering (SRE) teams access, see Identity and access management.

2.4. SRE AND SERVICE ACCOUNT ACCESS

2.4.1. Identity and access management

Most access by Red Hat site reliability engineering (SRE) teams is done by using cluster Operators through automated configuration management.

2.4.1.1. Subprocessors

For a list of the available subprocessors, see the Red Hat Subprocessor List on the Red Hat Customer Portal.

2.4.1.2. SRE access to all OpenShift Dedicated clusters

SREs access OpenShift Dedicated clusters through a proxy. The proxy mints a service account in an OpenShift Dedicated cluster for the SREs when they log in. As no identity provider is configured for OpenShift Dedicated clusters, SREs access the proxy by running a local web console container. SREs do not access the cluster web console directly. SREs must authenticate as individual users to ensure auditability. All authentication attempts are logged to a Security Information and Event Management (SIEM) system.

2.4.1.3. Privileged access controls in OpenShift Dedicated

Red Hat SRE adheres to the principle of least privilege when accessing OpenShift Dedicated and public cloud provider components. There are four basic categories of manual SRE access:

- SRE admin access through the Red Hat Customer Portal with normal two-factor authentication and no privileged elevation.
- SRE admin access through the Red Hat corporate SSO with normal two-factor authentication and no privileged elevation.
- OpenShift elevation, which is a manual elevation using Red Hat SSO. It is fully audited and management approval is required for every operation SREs make.

- Cloud provider access or elevation, which is a manual elevation for cloud provider console or CLI access. Access is limited to 60 minutes and is fully audited.

Each of these access types has different levels of access to components:

Component	Typical SRE admin access (Red Hat Customer Portal)	Typical SRE admin access (Red Hat SSO)	OpenShift elevation	Cloud provider access
OpenShift Cluster Manager	R/W	No access	No access	No access
OpenShift web console	No access	R/W	R/W	No access
Node operating system	No access	A specific list of elevated OS and network permissions.	A specific list of elevated OS and network permissions.	No access
AWS Console	No access	No access, but this is the account used to request cloud provider access.	No access	All cloud provider permissions using the SRE identity.

2.4.1.4. SRE access to cloud infrastructure accounts

Red Hat personnel do not access cloud infrastructure accounts in the course of routine OpenShift Dedicated operations. For emergency troubleshooting purposes, Red Hat SRE have well-defined and auditable procedures to access cloud infrastructure accounts.

In AWS, SREs generate a short-lived AWS access token for the **BYOCAdminAccess** user using the AWS Security Token Service (STS). Access to the STS token is audit logged and traceable back to individual users. The **BYOCAdminAccess** has the **AdministratorAccess** IAM policy attached.

In Google Cloud, SREs access resources after being authenticated against a Red Hat SAML identity provider (IDP). The IDP authorizes tokens that have time-to-live expirations. The issuance of the token is auditable by corporate Red Hat IT and linked back to an individual user.

2.4.1.5. Red Hat support access

Members of the Red Hat CEE team typically have read-only access to parts of the cluster. Specifically, CEE has limited access to the core and product namespaces and does not have access to the customer namespaces.

Role	Core namespace	Layered product namespace	Customer namespace	Cloud infrastructure account*
OpenShift SRE	Read: All Write: Very Limited [1]	Read: All Write: None	Read: None[2] Write: None	Read: All [3] Write: All [3]
CEE	Read: All Write: None	Read: All Write: None	Read: None[2] Write: None	Read: None Write: None
Customer administrator	Read: None Write: None	Read: None Write: None	Read: All Write: All	Read: Limited[4] Write: Limited[4]
Customer user	Read: None Write: None	Read: None Write: None	Read: Limited[5] Write: Limited[5]	Read: None Write: None
Everybody else	Read: None Write: None	Read: None Write: None	Read: None Write: None	Read: None Write: None

Cloud Infrastructure Account refers to the underlying AWS or Google Cloud account

1. Limited to addressing common use cases such as failing deployments, upgrading a cluster, and replacing bad worker nodes.
2. Red Hat associates have no access to customer data by default.
3. SRE access to the cloud infrastructure account is a "break-glass" procedure for exceptional troubleshooting during a documented incident.
4. Customer administrator has limited access to the cloud infrastructure account console through Cloud Infrastructure Access.
5. Limited to what is granted through RBAC by the customer administrator, as well as namespaces created by the user.

2.4.1.6. Customer access

Customer access is limited to namespaces created by the customer and permissions that are granted using RBAC by the customer administrator role. Access to the underlying infrastructure or product namespaces is generally not permitted without **cluster-admin** access. More information on customer access and authentication can be found in the Understanding Authentication section of the documentation.

2.4.1.7. Access approval and review

New SRE user access requires management approval. Separated or transferred SRE accounts are removed as authorized users through an automated process. Additionally, SRE performs periodic access review including management sign-off of authorized user lists.

2.4.2. SRE cluster access

SRE access to OpenShift Dedicated clusters is controlled through several layers of required authentication, all of which are managed by strict company policy. All authentication attempts to access a cluster and changes made within a cluster are recorded within audit logs, along with the specific account identity of the SRE responsible for those actions. These audit logs help ensure that all changes made by SREs to a customer's cluster adhere to the strict policies and procedures that make up Red Hat's managed services guidelines.

The information presented below is an overview of the process an SRE must perform to access a customer's cluster.

- SRE requests a refreshed ID token from the Red Hat SSO (Cloud Services). This request is authenticated. The token is valid for fifteen minutes. After the token expires, you can refresh the token again and receive a new token. The ability to refresh to a new token is indefinite; however, the ability to refresh to a new token is revoked after 30 days of inactivity.

- SRE connects to the Red Hat VPN. The authentication to the VPN is completed by the Red Hat Corporate Identity and Access Management system (RH IAM). With RH IAM, SREs are multifactor and can be managed internally per organization by groups and existing onboarding and offboarding processes. After an SRE is authenticated and connected, the SRE can access the cloud services fleet management plane. Changes to the cloud services fleet management plane require many layers of approval and are maintained by strict company policy.

- After authorization is complete, the SRE logs into the fleet management plane and receives a service account token that the fleet management plane created. The token is valid for 15 minutes. After the token is no longer valid, it is deleted.

- With access granted to the fleet management plane, SRE uses various methods to access clusters, depending on network configuration.

    - Accessing a private or public cluster: Request is sent through a specific Network Load Balancer (NLB) by using an encrypted HTTP connection on port 6443.

    - Accessing a PrivateLink cluster: Request is sent to the Red Hat Transit Gateway, which then connects to a Red Hat VPC per region. The VPC that receives the request will be dependent on the target private cluster's region. Within the VPC, there is a private subnet that contains the PrivateLink endpoint to the customer's PrivateLink cluster.

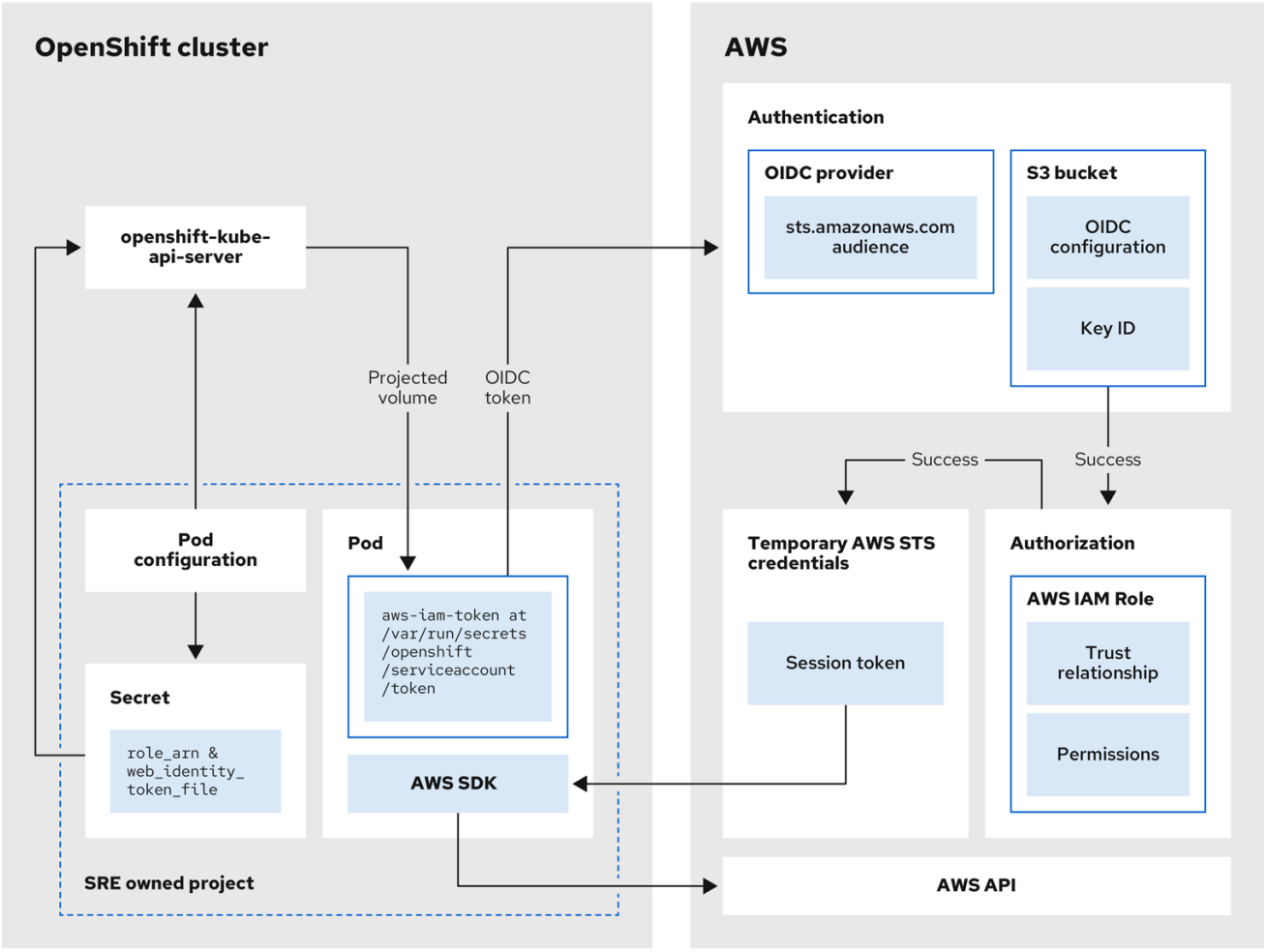
2.4.3. How service accounts assume AWS IAM roles in SRE owned projects

When you install a OpenShift Dedicated cluster that uses the AWS Security Token Service (STS), cluster-specific Operator AWS Identity and Access Management (IAM) roles are created. These IAM roles permit the OpenShift Dedicated cluster Operators to run core OpenShift functionality.

Cluster Operators use service accounts to assume IAM roles. When a service account assumes an IAM role, temporary STS credentials are provided for the service account to use in the cluster Operator's pod. If the assumed role has the necessary AWS privileges, the service account can run AWS SDK operations in the pod.

Workflow for assuming AWS IAM roles in SRE owned projects
The following diagram illustrates the workflow for assuming AWS IAM roles in SRE owned projects:

Figure 2.1. Workflow for assuming AWS IAM roles in SRE owned projects



The workflow has the following stages:

1. Within each project that a cluster Operator runs, the Operator's deployment spec has a volume mount for the projected service account token, and a secret containing AWS credential configuration for the pod. The token is audience-bound and time-bound. Every hour, OpenShift Dedicated generates a new token, and the AWS SDK reads the mounted secret containing the AWS credential configuration. This configuration has a path to the mounted token and the AWS IAM Role ARN. The secret's credential configuration includes the following:

    - An **$AWS_ARN_ROLE** variable that has the ARN for the IAM role that has the permissions required to run AWS SDK operations.

    - An **$AWS_WEB_IDENTITY_TOKEN_FILE** variable that has the full path in the pod to the OpenID Connect (OIDC) token for the service account. The full path is **/var/run/secrets/openshift/serviceaccount/token**.

2. When a cluster Operator needs to assume an AWS IAM role to access an AWS service (such as EC2), the AWS SDK client code running on the Operator invokes the **AssumeRoleWithWebIdentity** API call.

3. The OIDC token is passed from the pod to the OIDC provider. The provider authenticates the service account identity if the following requirements are met:

    - The identity signature is valid and signed by the private key.

- The **sts.amazonaws.com** audience is listed in the OIDC token and matches the audience configured in the OIDC provider.

NOTE

In OpenShift Dedicated with STS clusters, the OIDC provider is created during install and set as the service account issuer by default. The **sts.amazonaws.com** audience is set by default in the OIDC provider.

- The OIDC token has not expired.
- The issuer value in the token has the URL for the OIDC provider.

4. If the project and service account are in the scope of the trust policy for the IAM role that is being assumed, then authorization succeeds.
5. After successful authentication and authorization, temporary AWS STS credentials in the form of an AWS access token, secret key, and session token are passed to the pod for use by the service account. By using the credentials, the service account is temporarily granted the AWS permissions enabled in the IAM role.
6. When the cluster Operator runs, the Operator that is using the AWS SDK in the pod consumes the secret that has the path to the projected service account and AWS IAM Role ARN to authenticate against the OIDC provider. The OIDC provider returns temporary STS credentials for authentication against the AWS API.

2.5. UNDERSTANDING AVAILABILITY FOR OPENSHIFT DEDICATED

Availability and disaster avoidance are extremely important aspects of any application platform. OpenShift Dedicated provides many protections against failures at several levels, but customer-deployed applications must be appropriately configured for high availability. In addition, to account for cloud provider outages that might occur, other options are available, such as deploying a cluster across multiple availability zones or maintaining multiple clusters with failover mechanisms.

2.5.1. Potential points of failure

OpenShift Container Platform provides many features and options for protecting your workloads against downtime, but applications must be architected appropriately to take advantage of these features.

OpenShift Dedicated can help further protect you against many common Kubernetes issues by adding Red Hat Site Reliability Engineer (SRE) support and the option to deploy a multi-zone cluster, but there are a number of ways in which a container or infrastructure can still fail. By understanding potential points of failure, you can understand risks and appropriately architect both your applications and your clusters to be as resilient as necessary at each specific level.

NOTE

An outage can occur at several different levels of infrastructure and cluster components.

2.5.1.1. Container or pod failure

By design, pods are meant to exist for a short time. Appropriately scaling services so that multiple instances of your application pods are running protects against issues with any individual pod or

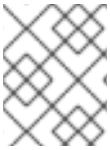
container. The node scheduler can also ensure that these workloads are distributed across different worker nodes to further improve resiliency.

When accounting for possible pod failures, it is also important to understand how storage is attached to your applications. Single persistent volumes attached to single pods cannot leverage the full benefits of pod scaling, whereas replicated databases, database services, or shared storage can.

To avoid disruption to your applications during planned maintenance, such as upgrades, it is important to define a pod disruption budget. These are part of the Kubernetes API and can be managed with the OpenShift CLI (**oc**) like other object types. They allow the specification of safety constraints on pods during operations, such as draining a node for maintenance.

2.5.1.2. Worker node failure

Worker nodes are the virtual machines that contain your application pods. By default, an OpenShift Dedicated cluster has a minimum of four worker nodes for a single availability-zone cluster. In the event of a worker node failure, pods are relocated to functioning worker nodes, as long as there is enough capacity, until any issue with an existing node is resolved or the node is replaced. More worker nodes means more protection against single node outages, and ensures proper cluster capacity for rescheduled pods in the event of a node failure.



NOTE

When accounting for possible node failures, it is also important to understand how storage is affected.

2.5.1.3. Cluster failure

OpenShift Dedicated clusters have at least three control plane nodes and three infrastructure nodes that are preconfigured for high availability, either in a single zone or across multiple zones depending on the type of cluster you have selected. This means that control plane and infrastructure nodes have the same resiliency of worker nodes, with the added benefit of being managed completely by Red Hat.

In the event of a complete control plane node outage, the OpenShift APIs will not function, and existing worker node pods will be unaffected. However, if there is also a pod or node outage at the same time, the control plane nodes will have to recover before new pods or nodes can be added or scheduled.

All services running on infrastructure nodes are configured by Red Hat to be highly available and distributed across infrastructure nodes. In the event of a complete infrastructure outage, these services will be unavailable until these nodes have been recovered.

2.5.1.4. Zone failure

A zone failure from a public cloud provider affects all virtual components, such as worker nodes, block or shared storage, and load balancers that are specific to a single availability zone. To protect against a zone failure, OpenShift Dedicated provides the option for clusters that are distributed across three availability zones, called multi-availability zone clusters. Existing stateless workloads are redistributed to unaffected zones in the event of an outage, as long as there is enough capacity.

2.5.1.5. Storage failure

If you have deployed a stateful application, then storage is a critical component and must be accounted for when thinking about high availability. A single block storage PV is unable to withstand outages even at the pod level. The best ways to maintain availability of storage are to use replicated storage solutions, shared storage that is unaffected by outages, or a database service that is independent of the cluster.

2.6. OPENSHIFT DEDICATED UPDATE LIFE CYCLE

2.6.1. Overview

Red Hat provides a published product life cycle for OpenShift Dedicated in order for customers and partners to effectively plan, deploy, and support their applications running on the platform. Red Hat publishes this life cycle to provide as much transparency as possible and might make exceptions from these policies as conflicts arise.

OpenShift Dedicated is a managed instance of Red Hat OpenShift and maintains an independent release schedule. More details about the managed offering can be found in the OpenShift Dedicated service definition. The availability of Security Advisories and Bug Fix Advisories for a specific version are dependent upon the Red Hat OpenShift Container Platform life cycle policy and subject to the OpenShift Dedicated maintenance schedule.

Additional resources

- OpenShift Dedicated service definition

2.6.2. Definitions

Table 2.3. Version reference

Version format	Major	Minor	Patch	Major.minor.patch
	x	y	z	x.y.z
Example	4	5	21	4.5.21

Major releases or X-releases

Referred to only as *major releases* or *X-releases* (X.y.z).

Examples

- "Major release 5" → 5.y.z
- "Major release 4" → 4.y.z
- "Major release 3" → 3.y.z

Minor releases or Y-releases

Referred to only as *minor releases* or *Y-releases* (x.Y.z).

Examples

- "Minor release 4" → 4.4.z
- "Minor release 5" → 4.5.z
- "Minor release 6" → 4.6.z

Patch releases or Z-releases

Referred to only as *patch releases* or *Z-releases* (x.y.Z).

Examples

- "Patch release 14 of minor release 5" → 4.5.14
- "Patch release 25 of minor release 5" → 4.5.25
- "Patch release 26 of minor release 6" → 4.6.26

2.6.3. Major versions (X.y.z)

Major versions of OpenShift Dedicated, for example version 4, are supported for one year following the release of a subsequent major version or the retirement of the product.

Example

- If version 5 were made available on OpenShift Dedicated on January 1, version 4 would be allowed to continue running on managed clusters for 12 months, until December 31. After this time, clusters would need to be upgraded or migrated to version 5.

2.6.4. Minor versions (x.Y.z)

Starting with the 4.8 OpenShift Container Platform minor version, Red Hat supports all minor versions for at least a 16 month period following general availability of the given minor version. Patch versions are not affected by the support period.

Customers are notified 60, 30, and 15 days before the end of the support period. Clusters must be upgraded to the latest patch version of the oldest supported minor version before the end of the support period, or the cluster will enter a "Limited Support" status.

Example

1. A customer's cluster is currently running on 4.13.8. The 4.13 minor version became generally available on May 17, 2023.
2. On July 19, August 16, and September 2, 2024, the customer is notified that their cluster will enter "Limited Support" status on September 17, 2024 if the cluster has not already been upgraded to a supported minor version.
3. The cluster must be upgraded to 4.14 or later by September 17, 2024.
4. If the upgrade has not been performed, the cluster will be flagged as being in a "Limited Support" status.

2.6.5. Patch versions (x.y.Z)

During the period in which a minor version is supported, Red Hat supports all OpenShift Container Platform patch versions unless otherwise specified.

For reasons of platform security and stability, a patch release may be deprecated, which would prevent installations of that release and trigger mandatory upgrades off that release.

Example

1. 4.7.6 is found to contain a critical CVE.
2. Any releases impacted by the CVE will be removed from the supported patch release list. In addition, any clusters running 4.7.6 will be scheduled for automatic upgrades within 48 hours.

2.6.6. Limited support status

When a cluster transitions to a *Limited Support* status, Red Hat no longer proactively monitors the cluster, the SLA is no longer applicable, and credits requested against the SLA are denied. It does not mean that you no longer have product support. In some cases, the cluster can return to a fully-supported status if you remediate the violating factors. However, in other cases, you might have to delete and recreate the cluster.

A cluster might transition to a Limited Support status for many reasons, including the following scenarios:

If you do not upgrade a cluster to a supported version before the end-of-life date

Red Hat does not make any runtime or SLA guarantees for versions after their end-of-life date. To receive continued support, upgrade the cluster to a supported version before the end-of-life date. If you do not upgrade the cluster before the end-of-life date, the cluster transitions to a Limited Support status until it is upgraded to a supported version.

Red Hat provides commercially reasonable support to upgrade from an unsupported version to a supported version. However, if a supported upgrade path is no longer available, you might have to create a new cluster and migrate your workloads.

If you remove or replace any native OpenShift Dedicated components or any other component that is installed and managed by Red Hat

If cluster administrator permissions were used, Red Hat is not responsible for any of your or your authorized users' actions, including those that affect infrastructure services, service availability, or data loss. If Red Hat detects any such actions, the cluster might transition to a Limited Support status. Red Hat notifies you of the status change and you should either revert the action or create a support case to explore remediation steps that might require you to delete and recreate the cluster.

If you have questions about a specific action that might cause a cluster to transition to a Limited Support status or need further assistance, open a support ticket.

2.6.7. Supported versions exception policy

Red Hat reserves the right to add or remove new or existing versions, or delay upcoming minor release versions, that have been identified to have one or more critical production impacting bugs or security issues without advance notice.

2.6.8. Installation policy

While Red Hat recommends installation of the latest support release, OpenShift Dedicated supports installation of any supported release as covered by the preceding policy.

2.6.9. Mandatory upgrades

If a critical or important CVE, or other bug identified by Red Hat, significantly impacts the security or stability of the cluster, the customer must upgrade to the next supported patch release within two business days.

In extreme circumstances and based on Red Hat's assessment of the CVE criticality to the environment,

Red Hat will notify customers that they have two business days to schedule or manually update their cluster to the latest, secure patch release. In the case that an update is not performed after two business days, Red Hat will automatically update the cluster to the latest, secure patch release to mitigate potential security breach(es) or instability. Red Hat might, at its own discretion, temporarily delay an automated update if requested by a customer through a support case.

2.6.10. Life cycle dates

Version	General availability	End of life
4.16	Jul 2, 2024	Nov 2, 2025
4.15	Feb 27, 2024	Jun 30, 2025
4.14	Oct 31, 2023	Feb 28, 2025
4.13	May 17, 2023	Sep 17, 2024
4.12	Jan 17, 2023	Jul 17, 2024
4.11	Aug 10, 2022	Dec 10, 2023
4.10	Mar 10, 2022	Sep 10, 2023
4.9	Oct 18, 2021	Dec 18, 2022
4.8	Jul 27, 2021	Sep 27, 2022