OpenShift Dedicated 4

Planning your environment

An overview of planning for Dedicated 4

Last Updated: 2026-02-07

OpenShift Dedicated 4 Planning your environment

An overview of planning for Dedicated 4

Legal Notice

Copyright © Red Hat.

Except as otherwise noted below, the text of and illustrations in this documentation are licensed by Red Hat under the Creative Commons Attribution–Share Alike 3.0 Unported license . If you distribute this document or an adaptation of it, you must provide the URL for the original version.

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

Red Hat, the Red Hat logo, JBoss, Hibernate, and RHCE are trademarks or registered trademarks of Red Hat, LLC. or its subsidiaries in the United States and other countries.

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

XFS is a trademark or registered trademark of Hewlett Packard Enterprise Development LP or its subsidiaries in the United States and other countries.

The OpenStack ® Word Mark and OpenStack logo are trademarks or registered trademarks of the Linux Foundation, used under license.

All other trademarks are the property of their respective owners.

Abstract

This document provides planning considerations for OpenShift Dedicated cluster deployments.

Table of Contents

CHAPTER 1. LIMITS AND SCALABILITY 4
1.1. CLUSTER MAXIMUMS 4
1.2. OPENSHIFT CONTAINER PLATFORM TESTING ENVIRONMENT AND CONFIGURATION 5
1.3. CONTROL PLANE AND INFRASTRUCTURE NODE SIZING AND SCALING 5
1.3.1. Node sizing during installation 6
1.3.2. Node scaling after installation 6
1.3.3. Sizing considerations for larger clusters 7

CHAPTER 2. CUSTOMER CLOUD SUBSCRIPTIONS ON GOOGLE CLOUD 8
2.1. UNDERSTANDING CUSTOMER CLOUD SUBSCRIPTIONS ON GOOGLE CLOUD 8
2.2. CUSTOMER REQUIREMENTS 8
2.2.1. Account 8
2.2.2. Access requirements 9
2.2.3. Support requirements 9
2.2.4. Security requirements 9
2.3. REQUIRED CUSTOMER PROCEDURE 10
2.4. ROLES REQUIRED FOR GOOGLE CLOUD MARKETPLACE BILLING 13
2.5. WORKLOAD IDENTITY FEDERATION AUTHENTICATION TYPE PROCEDURE 14
2.6. SERVICE ACCOUNT AUTHENTICATION TYPE PROCEDURE 18
2.7. RED HAT MANAGED GOOGLE CLOUD RESOURCES 18
2.7.1. IAM service account and roles 19
2.7.2. IAM group and roles 19
2.8. PROVISIONED GOOGLE CLOUD INFRASTRUCTURE 21
2.8.1. Compute instances 21
2.8.2. Storage 21
2.8.3. VPC 21
2.8.4. Services 22
2.9. GOOGLE CLOUD ACCOUNT LIMITS 22
2.10. GOOGLE CLOUD FIREWALL PREREQUISITES 24
2.11. ADDITIONAL RESOURCES 27

CHAPTER 3. CUSTOMER CLOUD SUBSCRIPTIONS ON AWS 28
3.1. UNDERSTANDING CUSTOMER CLOUD SUBSCRIPTIONS ON AWS 28
3.2. CUSTOMER REQUIREMENTS 28
3.2.1. Account 28
3.2.2. Access requirements 29
3.2.3. Support requirements 29
3.2.4. Security requirements 29
3.3. REQUIRED CUSTOMER PROCEDURE 29
3.4. MINIMUM REQUIRED SERVICE CONTROL POLICY (SCP) 30
3.5. RED HAT MANAGED IAM REFERENCES FOR AWS 34
3.5.1. IAM policies 34
3.5.2. IAM users 35
3.5.3. IAM roles 36
3.6. PROVISIONED AWS INFRASTRUCTURE 36
3.6.1. AWS Elastic Computing (EC2) instances 36
3.6.2. AWS Elastic Block Store (EBS) storage 36
3.6.3. Elastic Load Balancing (ELB) load balancers 37
3.6.4. S3 storage 37
3.6.5. VPC 37
3.6.6. Sample VPC Architecture 38

3.6.7. Security groups 38
3.6.8. Additional custom security groups 38
3.7. NETWORKING PREREQUISITES 39
3.7.1. Networking prerequisites 39
3.7.1.1. Minimum bandwidth 39
3.7.2. Firewall AllowList requirements 39
3.7.2.1. Domains for installation packages and tools 39
3.7.2.2. Domains for telemetry 41
3.7.2.3. Domains for Amazon Web Services (AWS) APIs 41
3.7.2.4. Domains for OpenShift 42
3.7.2.5. Domains for your site reliability engineering (SRE) and management 42
3.8. AWS ACCOUNT LIMITS 43

CHAPTER 1. LIMITS AND SCALABILITY

This document details the tested cluster maximums for OpenShift Dedicated clusters, along with information about the test environment and configuration used to test the maximums. Information about control plane and infrastructure node sizing and scaling is also provided.

1.1. CLUSTER MAXIMUMS

Consider the following tested object maximums when you plan an OpenShift Dedicated cluster installation. The table specifies the maximum limits for each tested type in an OpenShift Dedicated cluster.

These guidelines are based on a cluster of 249 compute (also known as worker) nodes in a multiple availability zone configuration. For smaller clusters, the maximums are lower.

Table 1.1. Tested cluster maximums

Maximum type	4.x tested maximum
Number of pods [1]	25,000
Number of pods per node	250
Number of pods per core	There is no default value
Number of namespaces [2]	5,000
Number of pods per namespace [3]	25,000
Number of services [4]	10,000
Number of services per namespace	5,000
Number of back ends per service	5,000
Number of deployments per namespace [3]	2,000

1. The pod count displayed here is the number of test pods. The actual number of pods depends on the memory, CPU, and storage requirements of the application.

2. When there are a large number of active projects, etcd can suffer from poor performance if the keyspace grows excessively large and exceeds the space quota. Periodic maintenance of etcd, including defragmentation, is highly recommended to make etcd storage available.

3. There are several control loops in the system that must iterate over all objects in a given namespace as a reaction to some changes in state. Having a large number of objects of a type, in a single namespace, can make those loops expensive and slow down processing the state changes. The limit assumes that the system has enough CPU, memory, and disk to satisfy the application requirements.

4. Each service port and each service back end has a corresponding entry in **iptables**. The number of back ends of a given service impacts the size of the endpoints objects, which then impacts the size of data sent throughout the system.

1.2. OPENSHIFT CONTAINER PLATFORM TESTING ENVIRONMENT AND CONFIGURATION

The following table lists the OpenShift Container Platform environment and configuration on which the cluster maximums are tested for the AWS cloud platform.

Node	Type	vCPU	RAM(GiB)	Disk type	Disk size(GiB) /IOPS	Count	Region
Control plane/etcd [1]	m5.4xlarge	16	64	gp3	350 / 1,000	3	us-west-2
Infrastructure nodes [2]	r5.2xlarge	8	64	gp3	300 / 900	3	us-west-2
Workload [3]	m5.2xlarge	8	32	gp3	350 / 900	3	us-west-2
Compute nodes	m5.2xlarge	8	32	gp3	350 / 900	102	us-west-2

1. io1 disks are used for control plane/etcd nodes in all versions prior to 4.10.
2. Infrastructure nodes are used to host monitoring components because Prometheus can claim a large amount of memory, depending on usage patterns.
3. Workload nodes are dedicated to run performance and scalability workload generators.

Larger cluster sizes and higher object counts might be reachable. However, the sizing of the infrastructure nodes limits the amount of memory that is available to Prometheus. When creating, modifying, or deleting objects, Prometheus stores the metrics in its memory for roughly 3 hours prior to persisting the metrics on disk. If the rate of creation, modification, or deletion of objects is too high, Prometheus can become overwhelmed and fail due to the lack of memory resources.

1.3. CONTROL PLANE AND INFRASTRUCTURE NODE SIZING AND SCALING

When you install an OpenShift Dedicated cluster, the sizing of the control plane and infrastructure nodes are automatically determined by the compute node count.

If you change the number of compute nodes in your cluster after installation, the Red Hat Site Reliability Engineering (SRE) team scales the control plane and infrastructure nodes as required to maintain cluster stability.

1.3.1. Node sizing during installation

During the installation process, the sizing of the control plane and infrastructure nodes are dynamically calculated. The sizing calculation is based on the number of compute nodes in a cluster.

The following tables list the control plane and infrastructure node sizing that is applied during installation.

AWS control plane and infrastructure node size:

Number of compute nodes	Control plane size	Infrastructure node size
1 to 25	m5.2xlarge	r5.xlarge
26 to 100	m5.4xlarge	r5.2xlarge
101 to 249	m5.8xlarge	r5.4xlarge

Google Cloud control plane and infrastructure node size:

Number of compute nodes	Control plane size	Infrastructure node size
1 to 25	custom-8-32768	custom-4-32768-ext
26 to 100	custom-16-65536	custom-8-65536-ext
101 to 249	custom-32-131072	custom-16-131072-ext

Google Cloud control plane and infrastructure node size for clusters created on or after 21 June 2024:

Number of compute nodes	Control plane size	Infrastructure node size
1 to 25	n2-standard-8	n2-highmem-4
26 to 100	n2-standard-16	n2-highmem-8
101 to 249	n2-standard-32	n2-highmem-16

NOTE

The maximum number of compute nodes on OpenShift Dedicated clusters version 4.14.14 and later is 249. For earlier versions, the limit is 180.

1.3.2. Node scaling after installation

If you change the number of compute nodes after installation, the control plane and infrastructure nodes are scaled by the Red Hat Site Reliability Engineering (SRE) team as required. The nodes are scaled to maintain platform stability.

Postinstallation scaling requirements for control plane and infrastructure nodes are assessed on a case-by-case basis. Node resource consumption and received alerts are taken into consideration.

Rules for control plane node resizing alerts

The resizing alert is triggered for the control plane nodes in a cluster when the following occurs:

- Control plane nodes sustain over 66% utilization on average in a cluster.

> **NOTE**
>
> The maximum number of compute nodes on OpenShift Dedicated is 180.

Rules for infrastructure node resizing alerts

Resizing alerts are triggered for the infrastructure nodes in a cluster when it has high-sustained CPU or memory utilization. This high-sustained utilization status is:

- Infrastructure nodes sustain over 50% utilization on average in a cluster with a single availability zone using 2 infrastructure nodes.
- Infrastructure nodes sustain over 66% utilization on average in a cluster with multiple availability zones using 3 infrastructure nodes.

> **NOTE**
>
> The maximum number of compute nodes on OpenShift Dedicated cluster versions 4.14.14 and later is 249. For earlier versions, the limit is 180.
>
> The resizing alerts only appear after sustained periods of high utilization. Short usage spikes, such as a node temporarily going down causing the other node to scale up, do not trigger these alerts.

The SRE team might scale the control plane and infrastructure nodes for additional reasons, for example to manage an increase in resource consumption on the nodes.

1.3.3. Sizing considerations for larger clusters

For larger clusters, infrastructure node sizing can become a significant impacting factor to scalability. There are many factors that influence the stated thresholds, including the etcd version or storage data format.

Exceeding these limits does not necessarily mean that the cluster will fail. In most cases, exceeding these numbers results in lower overall performance.

CHAPTER 2. CUSTOMER CLOUD SUBSCRIPTIONS ON GOOGLE CLOUD

OpenShift Dedicated provides a Customer Cloud Subscription (CCS) model that allows Red Hat to deploy and manage clusters in a customer's existing Google Cloud account.

2.1. UNDERSTANDING CUSTOMER CLOUD SUBSCRIPTIONS ON GOOGLE CLOUD

Red Hat OpenShift Dedicated provides a Customer Cloud Subscription (CCS) model that allows Red Hat to deploy and manage OpenShift Dedicated into a customer's existing Google Cloud account. Red Hat requires several prerequisites be met in order to provide this service.

Red Hat recommends the usage of a Google Cloud project, managed by the customer, to organize all of your Google Cloud resources. A project consists of a set of users and APIs, as well as billing, authentication, and monitoring settings for those APIs.

It is recommended for the OpenShift Dedicated cluster using a CCS model to be hosted in a Google Cloud project within a Google Cloud organization. The organization resource is the root node of the Google Cloud resource hierarchy and all resources that belong to an organization are grouped under the organization node. Customers have the choice of using service account keys or Workload Identity Federation when creating the roles and credentials necessary to access Google Cloud resources within a Google Cloud project.

For more information about creating and managing organization resources within Google Cloud, see Creating and managing organization resources.

2.2. CUSTOMER REQUIREMENTS

OpenShift Dedicated clusters using a Customer Cloud Subscription (CCS) model on Google Cloud must meet several prerequisites before they can be deployed.

2.2.1. Account

- The customer ensures that Google Cloud limits and allocation quotas that apply to Compute Engine are sufficient to support OpenShift Dedicated provisioned within the customer-provided Google Cloud account.
- The customer-provided Google Cloud account should be in the customer's Google Cloud Organization.
- The customer-provided Google Cloud account must not be transferable to Red Hat.
- The customer may not impose Google Cloud usage restrictions on Red Hat activities. Imposing restrictions severely hinders Red Hat's ability to respond to incidents.
- Red Hat deploys monitoring into Google Cloud to alert Red Hat when a highly privileged account, such as a root account, logs into the customer-provided Google Cloud account.
- The customer can deploy native Google Cloud services within the same customer-provided Google Cloud account.

NOTE

Customers are encouraged, but not mandated, to deploy resources in a Virtual Private Cloud (VPC) separate from the VPC hosting OpenShift Dedicated and other Red Hat supported services.

2.2.2. Access requirements

- To appropriately manage the OpenShift Dedicated service, Red Hat must have the **AdministratorAccess** policy applied to the administrator role at all times.

NOTE

This policy only provides Red Hat with permissions and capabilities to change resources in the customer-provided Google Cloud account.

- Red Hat must have Google Cloud console access to the customer-provided Google Cloud account. This access is protected and managed by Red Hat.
- The customer must not utilize the Google Cloud account to elevate their permissions within the OpenShift Dedicated cluster.
- Actions available in the OpenShift Cluster Manager must not be directly performed in the customer-provided Google Cloud account.

2.2.3. Support requirements

- Red Hat recommends that the customer have at least Enhanced Support from Google Cloud.
- Red Hat has authority from the customer to request Google Cloud support on their behalf.
- Red Hat has authority from the customer to request Google Cloud resource limit increases on the customer-provided account.
- Red Hat manages the restrictions, limitations, expectations, and defaults for all OpenShift Dedicated clusters in the same manner, unless otherwise specified in this requirements section.

2.2.4. Security requirements

- The customer-provided IAM credentials must be unique to the customer-provided Google Cloud account and must not be stored anywhere in the customer-provided Google Cloud account.
- Volume snapshots will remain within the customer-provided Google Cloud account and customer-specified region.
- To manage, monitor, and troubleshoot OpenShift Dedicated clusters, Red Hat must have direct access to the cluster's API server. You must not restrict or otherwise prevent Red Hat's access to the OpenShift Dedicated cluster's API server.

NOTE

SRE uses various methods to access clusters, depending on network configuration. Access to private clusters is restricted to Red Hat trusted IP addresses only. These access restrictions are managed automatically by Red Hat.

- OpenShift Dedicated requires egress access to certain endpoints over the internet. Only clusters deployed with Private Service Connect can use a firewall to control egress traffic. For additional information, see the *Google Cloud firewall prerequisites* section.

2.3. REQUIRED CUSTOMER PROCEDURE

The Customer Cloud Subscription (CCS) model allows Red Hat to deploy and manage OpenShift Dedicated into a customer's Google Cloud project. Red Hat requires several prerequisites to be completed before providing these services.

NOTE

The following requirements in this topic apply to OpenShift Dedicated on Google Cloud clusters created using both the Workload Identity Federation (WIF) and service account authentication types. Red Hat recommends using WIF as the authentication type for installing and interacting with an OpenShift Dedicated cluster deployed on Google Cloud because WIF provides enhanced security.

For information about creating a cluster using the WIF authentication type, see *Additional resources*.

For additional requirements that apply to the WIF authentication type only, see *Workload Identity Federation authentication type procedure*. For additional requirements that apply to the service account authentication type only, see *Service account authentication type procedure*.

Prerequisites

Before using OpenShift Dedicated in your Google Cloud project, confirm that the following organizational policy constraints are configured correctly where applicable:

- **constraints/iam.allowedPolicyMemberDomains**
  - This policy constraint is supported only if Red Hat's Directory Customer ID's **C02k0l5e8** and **C04j7mbwl** are included in the allowlist.
- **constraints/compute.restrictLoadBalancerCreationForTypes**
  - This organization policy constraint restricts the types of Google Cloud load balancer types that can be created within a project. Certain load balancer types are required depending on the type of OpenShift Dedicated cluster you are creating.
    - For private Workload Identity Federation (WIF)-enabled OpenShift Dedicated clusters with Google Cloud Private Service Connect (PSC), you must ensure the **INTERNAL_TCP_UDP** load balancer type is included in your organization's allowlist or excluded from the denylist.
    - For public WIF-enabled OpenShift Dedicated clusters, you must ensure the **INTERNAL_TCP_UDP**, **EXTERNAL_TCP_PROXY**, and

EXTERNAL_NETWORK_TCP_UDP load balancer types are permitted in your organization's allowlist or excluded from the denylist.

IMPORTANT

Although the **EXTERNAL_NETWORK_TCP_UDP** load balancer type is not required when creating a private cluster with PSC, disallowing it with this constraint will prevent the cluster from being able to create externally accessible load balancers.

- **constraints/compute.requireShieldedVm**
  - This policy constraint is supported only if the cluster is created with **Enable Secure Boot support for Shielded VMs** selected during the initial cluster creation.
- **constraints/compute.vmExternalIpAccess**
  - This policy constraint is supported only when creating a private cluster with Google Cloud Private Service Connect (PSC). For all other cluster types, this policy constraint is supported only after cluster creation.
- **constraints/compute.trustedImageProjects**
  - This policy constraint is supported only when the projects **redhat-marketplace-public**, **rhel-cloud**, and **rhcos-cloud** are included in the allowlist. If this policy constraint is enabled and these projects are not included in the allowlist, cluster creation will fail.

For more information about configuring Google Cloud organization policy constraints, see Organization policy constraints.

Procedure

1. Create a Google Cloud project to host the OpenShift Dedicated cluster.
2. Enable the following required APIs in the project that hosts your OpenShift Dedicated cluster:

Table 2.1. Required API services

API service	Console service name	Purpose
Cloud Deployment Manager V2 API	**deploymentmanager.googleapis.com**	Used for automated deployment and management of infrastructure resources.
Compute Engine API	**compute.googleapis.com**	Used for creating and managing virtual machines, firewalls, networks, persistent disk volumes, and load balancers.
Google Cloud APIs	**cloudapis.googleapis.com**	Used for managing Google Cloud services and resources.

API service	Console service name	Purpose
Cloud Resource Manager API	**cloudresourcemanager.googleapis.com**	Used for getting projects, getting or setting an IAM policy for projects, validating required permissions, and tagging.
Cloud DNS API	**dns.googleapis.com**	Used for creating DNS zones and managing DNS records for the cluster domains.
Network Security API	**networksecurity.googleapis.com**	Used for creating, managing, and enforcing network security policies for your applications and resources within Google Cloud.
IAM Service Account Credentials API	**iamcredentials.googleapis.com**	Used for creating short-lived credentials for impersonating IAM service accounts.
Identity and Access Management (IAM) API	**iam.googleapis.com**	Used for managing the IAM configuration for the cluster.
Service Management API	**servicemanagement.googleapis.com**	Used indirectly to fetch quota information for Google Cloud resources.
Service Usage API	**serviceusage.googleapis.com**	Used for determining what services are available in the customer's Google Cloud account.
Cloud Storage JSON API	**storage-api.googleapis.com**	Used for accessing Cloud Storage for the image registry, ignition, and cluster backups (if applicable).
Cloud Storage	**storage-component.googleapis.com**	Used for managing Cloud Storage for the image registry, ignition, and cluster backups (if applicable).
Organization Policy API	**orgpolicy.googleapis.com**	Used to identify governance rules applied to customer's Google Cloud that might impact cluster creation or management.

API service	Console service name	Purpose
Cloud Commerce Consumer Procurement API	**cloudcommerceconsumerprocurement.googleapis.com**	Enables users to procure products from the Google Cloud Marketplace. Specifically, this permission is required to validate through this API that customers have accepted the Marketplace terms and conditions for OpenShift Dedicated. This API is required when transacting through the Google Cloud Marketplace.
Cloud Identity-Aware Proxy API	**iap.googleapis.com**	Used in emergency situations to troubleshoot cluster nodes that are otherwise inaccessible. This API is required for clusters deployed with Private Service Connect.

2.4. ROLES REQUIRED FOR GOOGLE CLOUD MARKETPLACE BILLING

To deploy an OpenShift Dedicated cluster using Google Cloud Marketplace-based billing, your Google Cloud account must first be prepared. This involves accepting the Google Cloud Marketplace terms and agreements for the OpenShift Dedicated product listing. Contact your Google Cloud administrator who has the **Consumer Procurement Entitlement Manager** role to enable OpenShift Dedicated cluster deployments in your Google Cloud project.

To automate the checking and acceptance of these terms and agreements during OpenShift Dedicated cluster creation, you must grant the **Consumer Procurement Entitlement Viewer** role to the Google Cloud identity (user or service account) that is creating the cluster. The **Consumer Procurement Entitlement Viewer** role includes the necessary permissions to check for existing consent to the Google Cloud terms and agreements and grants consent if that has not yet been given.

The following table lists the permissions that are included in the **Consumer Procurement Entitlement Viewer** role.

Table 2.2. Required permissions in the Consumer Procurement Entitlement Viewer role

Role and description	Console role name	Permissions

Role and description	Console role name	Permissions
Consumer Procurement Entitlement Viewer Allows for the inspecting of entitlements and service states for a consumer project.	**consumerprocurement.entitlementViewer**	commerceoffercatalog.offers.get consumerprocurement.consents.check consumerprocurement.consents.list consumerprocurement.entitlements.get consumerprocurement.entitlements.list consumerprocurement.freeTrials.get consumerprocurement.freeTrials.list orgpolicy.policy.get resourcemanager.projects.get resourcemanager.projects.list serviceusage.consumerpolicy.analyze serviceusage.consumerpolicy.get serviceusage.effectivepolicy.get serviceusage.groups.list serviceusage.groups.listExpandedMembers serviceusage.groups.listMembers serviceusage.services.get serviceusage.services.list serviceusage.values.test

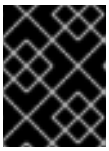
For more information about Google Cloud Marketplace roles and permissions, see Access control with IAM in the Google Cloud documentation.

2.5. WORKLOAD IDENTITY FEDERATION AUTHENTICATION TYPE PROCEDURE

Besides the required customer procedures listed in *Required customer procedure*, there are other specific actions that you must take when creating an OpenShift Dedicated cluster on Google Cloud using Workload Identity Federation (WIF) as the authentication type.

Procedure

1. Assign the following roles to the service account of the user implementing the WIF authentication type:



IMPORTANT

The following roles are only required when creating, updating, or deleting WIF configurations.

Table 2.3. Required roles

Role and description	Console role name	Permissions
Role Admin Required by the Google Cloud client in the OCM CLI for creating custom role.	**roles/iam.roleAdmin**	iam.roles.create iam.roles.delete iam.roles.get iam.roles.list iam.roles.undelete iam.roles.update resourcemanager.projects.get resourcemanager.projects.getIamPolicy
Service Account Admin Required for the pre-creation of the service accounts used by the deployer, support, and Operators.	**roles/iam.serviceAccountAdmin**	iam.serviceAccountApiKeyBindings.create iam.serviceAccountApiKeyBindings.delete iam.serviceAccountApiKeyBindings.undelete iam.serviceAccounts.create iam.serviceAccounts.create iam.serviceAccounts.create iam.serviceAccounts.createTagBinding iam.serviceAccounts.delete iam.serviceAccounts.deleteTagBinding iam.serviceAccounts.disable iam.serviceAccounts.enable iam.serviceAccounts.get iam.serviceAccounts.getIamPolicy iam.serviceAccounts.list iam.serviceAccounts.listEffectiveTags iam.serviceAccounts.listTagBindings iam.serviceAccounts.setIamPolicy iam.serviceAccounts.undelete iam.serviceAccounts.update resourcemanager.projects.get resourcemanager.projects.list

Role and description	Console role name	Permissions
Workload Identity Pool Admin Required to create and configure the workload identity pool.	**roles/iam.workloadIdentityPoolAdmin**	iam.googleapis.com/workloadIdentityPoolProviderKeys.create iam.googleapis.com/workloadIdentityPoolProviderKeys.delete iam.googleapis.com/workloadIdentityPoolProviderKeys.get iam.googleapis.com/workloadIdentityPoolProviderKeys.list iam.googleapis.com/workloadIdentityPoolProviderKeys.undelete iam.googleapis.com/workloadIdentityPoolProviders.create iam.googleapis.com/workloadIdentityPoolProviders.delete iam.googleapis.com/workloadIdentityPoolProviders.get iam.googleapis.com/workloadIdentityPoolProviders.list iam.googleapis.com/workloadIdentityPoolProviders.undelete iam.googleapis.com/workloadIdentityPoolProviders.up iam.googleapis.com/workloadIdentityPools.delete iam.googleapis.com/workloadIdentityPools.get iam.googleapis.com/workloadIdentityPools.list iam.googleapis.com/workloadIdentityPools.undelete iam.googleapis.com/workloadIdentityPools.update iam.workloadIdentityPools.createPolicyBinding iam.workloadIdentityPools.deletePolicyBinding iam.workloadIdentityPools.searchPolicyBindings iam.workloadIdentityPools.updatePolicyBinding resourcemanager.projects.get resourcemanager.projects.list

Role and description	Console role name	Permissions
Project IAM Admin Required for assigning roles to the service account and giving permissions to those roles that are necessary to perform operations on cloud resources.	**roles/resourcemanager.projectIamAdmin**	iam.policybindings.get iam.policybindings.list resourcemanager.projects.createPolicyBinding resourcemanager.projects.deletePolicyBinding resourcemanager.projects.get resourcemanager.projects.getIamPolicy resourcemanager.projects.searchPolicyBindings resourcemanager.projects.setIamPolicy resourcemanager.projects.updatePolicyBinding

2. Install the OpenShift Cluster Manager API command-line interface (**ocm**).

> **IMPORTANT**
>
> The OpenShift Cluster Manager API command-line interface (**ocm**) is a Developer Preview feature only. For more information about the support scope of Red Hat Developer Preview features, see Developer Preview Support Scope.

3. To authenticate against your Red Hat OpenShift Cluster Manager account, run one of the following commands.

   a. If your system supports a web-based browser, run the Red Hat single sign-on (SSO) authorization code command for secure authentication:

   ```
   $ ocm login --use-auth-code
   ```

   Running this command will redirect you to the Red Hat SSO login. Log in with your Red Hat login or email.

   b. If you are working with containers, remote hosts, and other environments without a web browser, run the Red Hat single sign-on (SSO) device code command for secure authentication:

   Syntax

   ```
   $ ocm login --use-device-code
   ```

   Running this command will redirect you to the Red Hat SSO login and provide a log in code.

   To switch accounts, logout from https://sso.redhat.com and run the **ocm logout** command in your terminal before attempting to login again.

4. Install the gcloud CLI.

5. Authenticate the gcloud CLI with the Application Default Credentials (ADC).

2.6. SERVICE ACCOUNT AUTHENTICATION TYPE PROCEDURE

Besides the required customer procedures listed in *Required customer procedure*, there are other specific actions that you must take when creating an OpenShift Dedicated cluster on Google Cloud using a service account as the authentication type.

Procedure

1. To ensure that Red Hat can perform necessary actions, you must create an **osd-ccs-admin** IAM service account user within the Google Cloud project.
   The following roles must be granted to the service account :

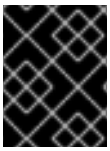
   Table 2.4. Required roles

Role	Console role name
Compute Admin	**roles/compute.admin**
DNS Administrator	**roles/dns.admin**
Organization Policy Viewer	**roles/orgpolicy.policyViewer**
Service Management Administrator	**roles/servicemanagement.admin**
Service Usage Admin	**roles/serviceusage.serviceUsageAdmin**
Storage Admin	**roles/storage.admin**
Compute Load Balancer Admin	**roles/compute.loadBalancerAdmin**
Role Viewer	**roles/viewer**
Role Administrator	**roles/iam.roleAdmin**
Security Admin	**roles/iam.securityAdmin**
Service Account Key Admin	**roles/iam.serviceAccountKeyAdmin**
Service Account Admin	**roles/iam.serviceAccountAdmin**
Service Account User	**roles/iam.serviceAccountUser**

2. Create the service account key for the **osd-ccs-admin** IAM service account. Export the key to a file named **osServiceAccount.json**; this JSON file will be uploaded in Red Hat OpenShift Cluster Manager when you create your cluster.

2.7. RED HAT MANAGED GOOGLE CLOUD RESOURCES

Red Hat is responsible for creating and managing the following IAM Google Cloud resources.



IMPORTANT

The *IAM service account and roles* and *IAM group and roles* topics are only applicable to clusters created using the service account authentication type.

2.7.1. IAM service account and roles

The **osd-managed-admin** IAM service account is created immediately after taking control of the customer-provided Google Cloud account. This is the user that will perform the OpenShift Dedicated cluster installation.

The following roles are attached to the service account:

Table 2.5. IAM roles for osd-managed-admin

Role	Console role name	Description
Compute Admin	**roles/compute.admin**	Provides full control of all Compute Engine resources.
DNS Administrator	**roles/dns.admin**	Provides read-write access to all Cloud DNS resources.
Security Admin	**roles/iam.securityAdmin**	Security admin role, with permissions to get and set any IAM policy.
Storage Admin	**roles/storage.admin**	Grants full control of objects and buckets. When applied to an individual **bucket**, control applies only to the specified bucket and objects within the bucket.
Service Account Admin	**roles/iam.serviceAccountAdmin**	Create and manage service accounts.
Service Account Key Admin	**roles/iam.serviceAccountKeyAdmin**	Create and manage (and rotate) service account keys.
Service Account User	**roles/iam.serviceAccountUser**	Run operations as the service account.
Role Administrator	**roles/iam.roleAdmin**	Provides access to all custom roles in the project.

2.7.2. IAM group and roles

The **sd-sre-platform-gcp-access** Google group is granted access to the Google Cloud project to allow Red Hat Site Reliability Engineering (SRE) access to the console for emergency troubleshooting purposes.

NOTE

- For information regarding the roles within the **sd-sre-platform-gcp-access** group that are specific to clusters created when using the Workload Identity Federation (WIF) authentication type, see managed-cluster-config.
- For information about creating a cluster using the Workload Identity Federation authentication type, see *Additional resources*.

The following roles are attached to the group:

Table 2.6. IAM roles for sd-sre-platform-gcp-access

Role	Console role name	Description
Compute Admin	**roles/compute.admin**	Provides full control of all Compute Engine resources.
Editor	**roles/editor**	Provides all viewer permissions, plus permissions for actions that modify state.
Organization Policy Viewer	**roles/orgpolicy.policyViewer**	Provides access to view Organization Policies on resources.
Project IAM Admin	**roles/resourcemanager.projectIamAdmin**	Provides permissions to administer IAM policies on projects.
Quota Administrator	**roles/servicemanagement.quotaAdmin**	Provides access to administer service quotas.
Role Administrator	**roles/iam.roleAdmin**	Provides access to all custom roles in the project.
Service Account Admin	**roles/iam.serviceAccountAdmin**	Create and manage service accounts.
Service Usage Admin	**roles/serviceusage.serviceUsageAdmin**	Ability to enable, disable, and inspect service states, inspect operations, and consume quota and billing for a consumer project.

Role	Console role name	Description
Tech Support Editor	**roles/cloudsupport.techSupportEditor**	Provides full read-write access to technical support cases.

2.8. PROVISIONED GOOGLE CLOUD INFRASTRUCTURE

This is an overview of the provisioned Google Cloud components on a deployed OpenShift Dedicated cluster. For a more detailed listing of all provisioned Google Cloud components, see the OpenShift Container Platform documentation.

2.8.1. Compute instances

Google Cloud compute instances are required to deploy the control plane and data plane functions of OpenShift Dedicated in Google Cloud. Instance types might vary for control plane and infrastructure nodes depending on worker node count.

- Single availability zone
  - 2 infra nodes (n2-highmem-4 machine type: 4 vCPU and 32 GB RAM)
  - 3 control plane nodes (n2-standard-8 machine type: 8 vCPU and 32 GB RAM)
  - 2 worker nodes (default n2-standard-4 machine type: 4 vCPU and 16 GB RAM)
- Multiple availability zones
  - 3 infra nodes (n2-highmem-4 machine type: 4 vCPU and 32 GB RAM)
  - 3 control plane nodes (n2-standard-8 machine type: 8 vCPU and 32 GB RAM)
  - 3 worker nodes (default n2-standard-4 machine type: 4 vCPU and 16 GB RAM)

2.8.2. Storage

- Infrastructure volumes:
  - 300 GB SSD persistent disk (deleted on instance deletion)
  - 110 GB Standard persistent disk (kept on instance deletion)
- Worker volumes:
  - 300 GB SSD persistent disk (deleted on instance deletion)
- Control plane volumes:
  - 350 GB SSD persistent disk (deleted on instance deletion)

2.8.3. VPC

NOTE

Installing a new OpenShift Dedicated cluster into a VPC that was automatically created by the installer for a different cluster is not supported.

- **Subnets:** One master subnet for the control plane workloads and one worker subnet for all others. An additional subnet is required for Google Private Service Connect (PSC) when a private cluster is deployed using PSC.
- **Router tables:** One global route table per VPC.
- **Internet gateways:** One internet gateway per cluster.
- **NAT gateways:** One master NAT gateway and one worker NAT gateway per cluster.

2.8.4. Services

For a list of services that must be enabled on a Google Cloud CCS cluster, see the *Required API services* table.

2.9. GOOGLE CLOUD ACCOUNT LIMITS

The OpenShift Dedicated cluster uses a number of Google Cloud components, but the default quotas do not affect your ability to install an OpenShift Dedicated cluster.

A standard OpenShift Dedicated cluster uses the following resources. Note that some resources are required only during the bootstrap process and are removed after the cluster deploys.

NOTE

3 subnets are required to deploy a private cluster with Private Service Connect (PSC). These subnets are a control plane subnet, a worker subnet, and a subnet used for the PSC service attachment with the purpose set to Private Service Connect.

48 vCPUs for a default multi-AZ OpenShift Dedicated cluster consists of 3 compute nodes (4 vCPUs each, one per availability zone), 3 infra nodes (4 vCPU each), and 3 control plane nodes (8 vCPU each).

40 vCPUs for a default single-AZ OpenShift Dedicated cluster consists of 2 compute nodes (4 vCPUs each), 2 infra nodes (4 vCPU each) and 3 control plane nodes (8 vCPU each).

Table 2.7. Google Cloud resources used in a default cluster

Service	Component	Location	Total resources required	Resources removed after bootstrap
Service account	IAM	Global	10	0
Firewall Rules	Compute	Global	11	1
Forwarding Rules	Compute	Global	2	0

Service	Component	Location	Total resources required	Resources removed after bootstrap
In-use global IP addresses	Compute	Global	4	1
Health checks	Compute	Global	3	0
Images	Compute	Global	1	0
Networks	Compute	Global	2	0
Static IP addresses	Compute	Region	4	1
Routers	Compute	Global	1	0
Routes	Compute	Global	2	0
Subnetworks	Compute	Global	3	0
Target Pools	Compute	Global	3	0
CPUs	Compute	Region	48	4
Persistent Disk SSD (GB)	Compute	Region	1060	128

NOTE

If any of the quotas are insufficient during installation, the installation program displays an error that states both which quota was exceeded and the region.

Be sure to consider your actual cluster size, planned cluster growth, and any usage from other clusters that are associated with your account. The CPU, Static IP addresses, and Persistent Disk SSD (Storage) quotas are the ones that are most likely to be insufficient.

If you plan to deploy your cluster in one of the following regions, you will exceed the maximum storage quota and are likely to exceed the CPU quota limit:

- asia-east2
- asia-northeast2
- asia-south1
- australia-southeast1
- europe-north1
- europe-west2

- europe-west3
- europe-west6
- northamerica-northeast1
- southamerica-east1
- us-west2

You can increase resource quotas from the Google Cloud console, but you might need to file a support ticket. Be sure to plan your cluster size early so that you can allow time to resolve the support ticket before you install your OpenShift Dedicated cluster.

2.10. GOOGLE CLOUD FIREWALL PREREQUISITES

If you are using a firewall to control egress traffic from OpenShift Dedicated on Google Cloud, you must configure your firewall to grant access to certain domains and port combinations listed in the tables below. OpenShift Dedicated requires this access to provide a fully managed OpenShift service.

Procedure

1. Add the following URLs that are used to install and download packages and tools to an allowlist:

Domain	Port	Function
registry.redhat.io	443	Provides core container images.
quay.io	443	Provides core container images.
cdn01.quay.io **cdn02.quay.io** **cdn03.quay.io** **cdn04.quay.io** **cdn05.quay.io** **cdn06.quay.io**	443	Provides core container images.
sso.redhat.com	443	Required. The https://console.redhat.com/openshift site uses authentication from sso.redhat.com to download the pull secret and use Red Hat SaaS solutions to facilitate monitoring of your subscriptions, cluster inventory, chargeback reporting, and so on.
quayio-production-s3.s3.amazonaws.com	443	Provides core container images.

Domain	Port	Function
pull.q1w2.quay.rhcloud.com	443	Provides core container images.
registry.access.redhat.com	443	Hosts all the container images that are stored on the Red Hat Ecosytem Catalog. Additionally, the registry provides access to the **odo** CLI tool that helps developers build on OpenShift and Kubernetes.
registry.connect.redhat.com	443	Required for all third-party images and certified Operators.
console.redhat.com	443	Required. Allows interactions between the cluster and Red Hat OpenShift Cluster Manager to enable functionality, such as scheduling upgrades.
catalog.redhat.com	443	The **registry.access.redhat.com** and **https://registry.redhat.io** sites redirect through **catalog.redhat.com**.

2. Add the following telemetry URLs to an allowlist:

Domain	Port	Function
cert-api.access.redhat.com	443	Required for telemetry.
api.access.redhat.com	443	Required for telemetry.
infogw.api.openshift.com	443	Required for telemetry.
console.redhat.com	443	Required for telemetry and Red Hat Lightspeed.
observatorium-mst.api.openshift.com	443	Required for managed OpenShift-specific telemetry.
observatorium.api.openshift.com	443	Required for managed OpenShift-specific telemetry.

NOTE

Managed clusters require the enabling of telemetry to allow Red Hat to react more quickly to problems, better support the customers, and better understand how product upgrades impact clusters. For more information about how remote health monitoring data is used by Red Hat, see *About remote health monitoring* in the *Additional resources* section.

3. Add the following OpenShift Dedicated URLs to an allowlist:

Domain	Port	Function
mirror.openshift.com	443	Used to access mirrored installation content and images. This site is also a source of release image signatures.
api.openshift.com	443	Used to check if updates are available for the cluster.

4. Add the following site reliability engineering (SRE) and management URLs to an allowlist:

Domain	Port	Function
api.pagerduty.com	443	This alerting service is used by the in-cluster alertmanager to send alerts notifying Red Hat SRE of an event to take action on.
events.pagerduty.com	443	This alerting service is used by the in-cluster alertmanager to send alerts notifying Red Hat SRE of an event to take action on.
api.deadmanssnitch.com	443	Alerting service used by OpenShift Dedicated to send periodic pings that indicate whether the cluster is available and running.
nosnch.in	443	Alerting service used by OpenShift Dedicated to send periodic pings that indicate whether the cluster is available and running.
http-inputs-osdsecuritylogs.splunkcloud.com	443	Used by the **splunk-forwarder-operator** as a logging forwarding endpoint to be used by Red Hat SRE for log-based alerting.
sftp.access.redhat.com (Recommended)	22	The SFTP server used by **must-gather-operator** to upload diagnostic logs to help troubleshoot issues with the cluster.

5. Add the following URLs for the Google Cloud API endpoints to an allowlist:

Domain	Port	Function
accounts.google.com	443	Used to access your Google Cloud account.

Domain	Port	Function
***.googleapis.com** OR **storage.googleapis.com** **iam.googleapis.com** **serviceusage.googleapis.com** **cloudresourcemanager.googleapis.com** **compute.googleapis.com** **oauth2.googleapis.com** **dns.googleapis.com** **iamcredentials.googleapis.com**	443	Used to access Google Cloud services and resources. Review Cloud Endpoints in the Google Cloud documentation to determine the endpoints to allow for your APIs.

NOTE

Required Google APIs can be exposed using the Private Google Access restricted virtual IP (VIP), with the exception of the Service Usage API (serviceusage.googleapis.com). To circumvent this, you must expose the Service Usage API using the Private Google Access private VIP .

2.11. ADDITIONAL RESOURCES

- About remote health monitoring
- Creating a cluster on Google Cloud with Workload Identity Federation authentication
- For more information about the specific roles and permissions that are specific to clusters created when using the Workload Identity Federation (WIF) authentication type, see managed-cluster-config.

CHAPTER 3. CUSTOMER CLOUD SUBSCRIPTIONS ON AWS

OpenShift Dedicated provides a Customer Cloud Subscription (CCS) model that allows Red Hat to deploy and manage clusters into a customer's existing Amazon Web Service (AWS) account.

3.1. UNDERSTANDING CUSTOMER CLOUD SUBSCRIPTIONS ON AWS

To deploy OpenShift Dedicated into your existing Amazon Web Services (AWS) account using the Customer Cloud Subscription (CCS) model, Red Hat requires several prerequisites be met.

Red Hat recommends the usage of an AWS Organization to manage multiple AWS accounts. The AWS Organization, managed by the customer, hosts multiple AWS accounts. There is a root account in the organization that all accounts will refer to in the account hierarchy.

It is recommended for the OpenShift Dedicated cluster using a CCS model to be hosted in an AWS account within an AWS Organizational Unit. A service control policy (SCP) is created and applied to the AWS Organizational Unit that manages what services the AWS sub-accounts are permitted to access. The SCP applies only to available permissions within a single AWS account for all AWS sub-accounts within the Organizational Unit. It is also possible to apply a SCP to a single AWS account. All other accounts in the customer's AWS Organization are managed in whatever manner the customer requires. Red Hat Site Reliability Engineers (SRE) will not have any control over SCPs within the AWS Organization.

3.2. CUSTOMER REQUIREMENTS

OpenShift Dedicated clusters using a Customer Cloud Subscription (CCS) model on Amazon Web Services (AWS) must meet several prerequisites before they can be deployed.

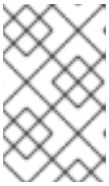
3.2.1. Account

- The customer ensures that AWS limits are sufficient to support OpenShift Dedicated provisioned within the customer-provided AWS account.
- The customer-provided AWS account should be in the customer's AWS Organization with the applicable service control policy (SCP) applied.

NOTE

It is not a requirement that the customer-provided account be within an AWS Organization or for the SCP to be applied, however Red Hat must be able to perform all the actions listed in the SCP without restriction.

- The customer-provided AWS account must not be transferable to Red Hat.
- The customer may not impose AWS usage restrictions on Red Hat activities. Imposing restrictions severely hinders Red Hat's ability to respond to incidents.
- Red Hat deploys monitoring into AWS to alert Red Hat when a highly privileged account, such as a root account, logs into the customer-provided AWS account.
- The customer can deploy native AWS services within the same customer-provided AWS account.



NOTE

Customers are encouraged, but not mandated, to deploy resources in a Virtual Private Cloud (VPC) separate from the VPC hosting OpenShift Dedicated and other Red Hat supported services.

3.2.2. Access requirements

- To appropriately manage the OpenShift Dedicated service, Red Hat must have the **AdministratorAccess** policy applied to the administrator role at all times.

NOTE

This policy only provides Red Hat with permissions and capabilities to change resources in the customer-provided AWS account.

- Red Hat must have AWS console access to the customer-provided AWS account. This access is protected and managed by Red Hat.
- The customer must not utilize the AWS account to elevate their permissions within the OpenShift Dedicated cluster.
- Actions available in OpenShift Cluster Manager must not be directly performed in the customer-provided AWS account.

3.2.3. Support requirements

- Red Hat recommends that the customer have at least Business Support from AWS.
- Red Hat has authority from the customer to request AWS support on their behalf.
- Red Hat has authority from the customer to request AWS resource limit increases on the customer-provided account.
- Red Hat manages the restrictions, limitations, expectations, and defaults for all OpenShift Dedicated clusters in the same manner, unless otherwise specified in this requirements section.

3.2.4. Security requirements

- The customer-provided IAM credentials must be unique to the customer-provided AWS account and must not be stored anywhere in the customer-provided AWS account.
- Volume snapshots will remain within the customer-provided AWS account and customer-specified region.
- Red Hat must have ingress access to EC2 hosts and the API server through white-listed Red Hat machines.
- Red Hat must have egress allowed to forward system and audit logs to a Red Hat managed central logging stack.

3.3. REQUIRED CUSTOMER PROCEDURE

The Customer Cloud Subscription (CCS) model allows Red Hat to deploy and manage OpenShift Dedicated into a customer's Amazon Web Services (AWS) account. Red Hat requires several prerequisites in order to provide these services.

Procedure

1. If the customer is using AWS Organizations, you must either use an AWS account within your organization or create a new one .
2. To ensure that Red Hat can perform necessary actions, you must either create a service control policy (SCP) or ensure that none is applied to the AWS account.
3. Attach the SCP to the AWS account.
4. Within the AWS account, you must create an **osdCcsAdmin** IAM user with the following requirements:
   - This user needs at least **Programmatic access** enabled.
   - This user must have the **AdministratorAccess** policy attached to it.
5. Provide the IAM user credentials to Red Hat.
   - You must provide the **access key ID** and **secret access key** in OpenShift Cluster Manager.

3.4. MINIMUM REQUIRED SERVICE CONTROL POLICY (SCP)

Service control policy (SCP) management is the responsibility of the customer. These policies are maintained in the AWS Organization and control what services are available within the attached AWS accounts.

Required/optional	Service	Actions	Effect
Required	Amazon EC2	All	Allow
	Amazon EC2 Auto Scaling	All	Allow
	Amazon S3	All	Allow
	Identity And Access Management	All	Allow
	Elastic Load Balancing	All	Allow
	Elastic Load Balancing V2	All	Allow
	Amazon CloudWatch	All	Allow
	Amazon CloudWatch Events	All	Allow

Required/optional	Service	Actions	Effect
	Amazon CloudWatch Logs	All	Allow
	AWS Support	All	Allow
	AWS Key Management Service	All	Allow
	AWS Security Token Service	All	Allow
	AWS Resource Tagging	All	Allow
	AWS Route53 DNS	All	Allow
	AWS Service Quotas	ListServices GetRequestedServiceQuotaChange GetServiceQuota RequestServiceQuotaIncrease ListServiceQuotas	Allow
Optional	AWS Billing	ViewAccount Viewbilling ViewUsage	Allow
	AWS Cost and Usage Report	All	Allow
	AWS Cost Explorer Services	All	Allow

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ec2:*"
            ],
            "Resource": [
                "*"
```

```
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "autoscaling:*"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:*"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:*"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "elasticloadbalancing:*"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "cloudwatch:*"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "events:*"
            ],
            "Resource": [
                "*"
            ]
        },
```

```
        {
            "Effect": "Allow",
            "Action": [
                "logs:*"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "support:*"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "kms:*"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "sts:*"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "tag:*"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "route53:*"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Effect": "Allow",
```

```
            "Action": [
                "servicequotas:ListServices",
                "servicequotas:GetRequestedServiceQuotaChange",
                "servicequotas:GetServiceQuota",
                "servicequotas:RequestServiceQuotaIncrease",
                "servicequotas:ListServiceQuotas"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

3.5. RED HAT MANAGED IAM REFERENCES FOR AWS

Red Hat is responsible for creating and managing the following Amazon Web Services (AWS) resources: IAM policies, IAM users, and IAM roles.

3.5.1. IAM policies

NOTE

IAM policies are subject to modification as the capabilities of OpenShift Dedicated change.

- The **AdministratorAccess** policy is used by the administration role. This policy provides Red Hat the access necessary to administer the OpenShift Dedicated cluster in the customer-provided AWS account.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": "*",
            "Resource": "*",
            "Effect": "Allow"
        }
    ]
}
```

- The **CustomerAdministratorAccess** role provides the customer access to administer a subset of services within the AWS account. At this time, the following are allowed:
  - VPC Peering
  - VPN Setup
  - Direct Connect (only available if granted through the service control policy)

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
```

```
            "Effect": "Allow",
            "Action": [
                "ec2:AttachVpnGateway",
                "ec2:DescribeVpnConnections",
                "ec2:AcceptVpcPeeringConnection",
                "ec2:DeleteVpcPeeringConnection",
                "ec2:DescribeVpcPeeringConnections",
                "ec2:CreateVpnConnectionRoute",
                "ec2:RejectVpcPeeringConnection",
                "ec2:DetachVpnGateway",
                "ec2:DeleteVpnConnectionRoute",
                "ec2:DeleteVpnGateway",
                "ec2:DescribeVpcs",
                "ec2:CreateVpnGateway",
                "ec2:ModifyVpcPeeringConnectionOptions",
                "ec2:DeleteVpnConnection",
                "ec2:CreateVpcPeeringConnection",
                "ec2:DescribeVpnGateways",
                "ec2:CreateVpnConnection",
                "ec2:DescribeRouteTables",
                "ec2:CreateTags",
                "ec2:CreateRoute",
            "directconnect:*"
            ],
            "Resource": "*"
        }
    ]
}
```

- If enabled, the **BillingReadOnlyAccess** role provides read-only access to view billing and usage information for the account.
  Billing and usage access is only granted if the root account in the AWS Organization has it enabled. This is an optional step the customer must perform to enable read-only billing and usage access and does not impact the creation of this profile and the role that uses it. If this role is not enabled, users will not see billing and usage information. See this tutorial on how to enable access to billing data.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "aws-portal:ViewAccount",
                "aws-portal:ViewBilling"
            ],
            "Resource": "*"
        }
    ]
}
```

3.5.2. IAM users

The **osdManagedAdmin** user is created immediately after taking control of the customer-provided AWS account. This is the user that will perform the OpenShift Dedicated cluster installation.

3.5.3. IAM roles

- The **network-mgmt** role provides customer-federated administrative access to the AWS account through a separate AWS account. It also has the same access as a read-only role. The **network-mgmt** role only applies to non-Customer Cloud Subscription (CCS) clusters. The following policies are attached to the role:
  - AmazonEC2ReadOnlyAccess
  - CustomerAdministratorAccess
- The **read-only** role provides customer-federated read-only access to the AWS account through a separate AWS account. The following policies are attached to the role:
  - AWSAccountUsageReportAccess
  - AmazonEC2ReadOnlyAccess
  - AmazonS3ReadOnlyAccess
  - IAMReadOnlyAccess
  - BillingReadOnlyAccess

3.6. PROVISIONED AWS INFRASTRUCTURE

This is an overview of the provisioned Amazon Web Services (AWS) components on a deployed OpenShift Dedicated cluster. For a more detailed listing of all provisioned AWS components, see the OpenShift Container Platform documentation.

3.6.1. AWS Elastic Computing (EC2) instances

AWS EC2 instances are required to deploy the control plane and data plane functions of OpenShift Dedicated in the AWS public cloud. Instance types might vary for control plane and infrastructure nodes depending on worker node count.

- Single availability zone
  - 3 m5.2xlarge minimum (control plane nodes)
  - 2 r5.xlarge minimum (infrastructure nodes)
  - 2 m5.xlarge minimum but highly variable (worker nodes)
- Multiple availability zones
  - 3 m5.2xlarge minimum (control plane nodes)
  - 3 r5.xlarge minimum (infrastructure nodes)
  - 3 m5.xlarge minimum but highly variable (worker nodes)

3.6.2. AWS Elastic Block Store (EBS) storage

Amazon EBS block storage is used for both local node storage and persistent volume storage.

Volume requirements for each EC2 instance:

- Control plane volumes
  - Size: 350 GB
  - Type: io1
  - Input/output operations per second: 1000
- Infrastructure volumes
  - Size: 300 GB
  - Type: gp2
  - Input/output operations per second: 900
- Worker volumes
  - Size: 300 GB
  - Type: gp2
  - Input/output operations per second: 900

3.6.3. Elastic Load Balancing (ELB) load balancers

Up to two Network Load Balancers for API and up to two Classic Load Balancers for application router. For more information, see the ELB documentation for AWS.

3.6.4. S3 storage

The image registry and Elastic Block Store (EBS) volume snapshots are backed by AWS S3 storage. Pruning of resources is performed regularly to optimize S3 usage and cluster performance.

NOTE

Two buckets are required with a typical size of 2 TB each.

3.6.5. VPC

Customers should expect to see one VPC per cluster. Additionally, the VPC needs the following configurations:

NOTE

Installing a new OpenShift Dedicated cluster into a VPC that was automatically created by the installer for a different cluster is not supported.

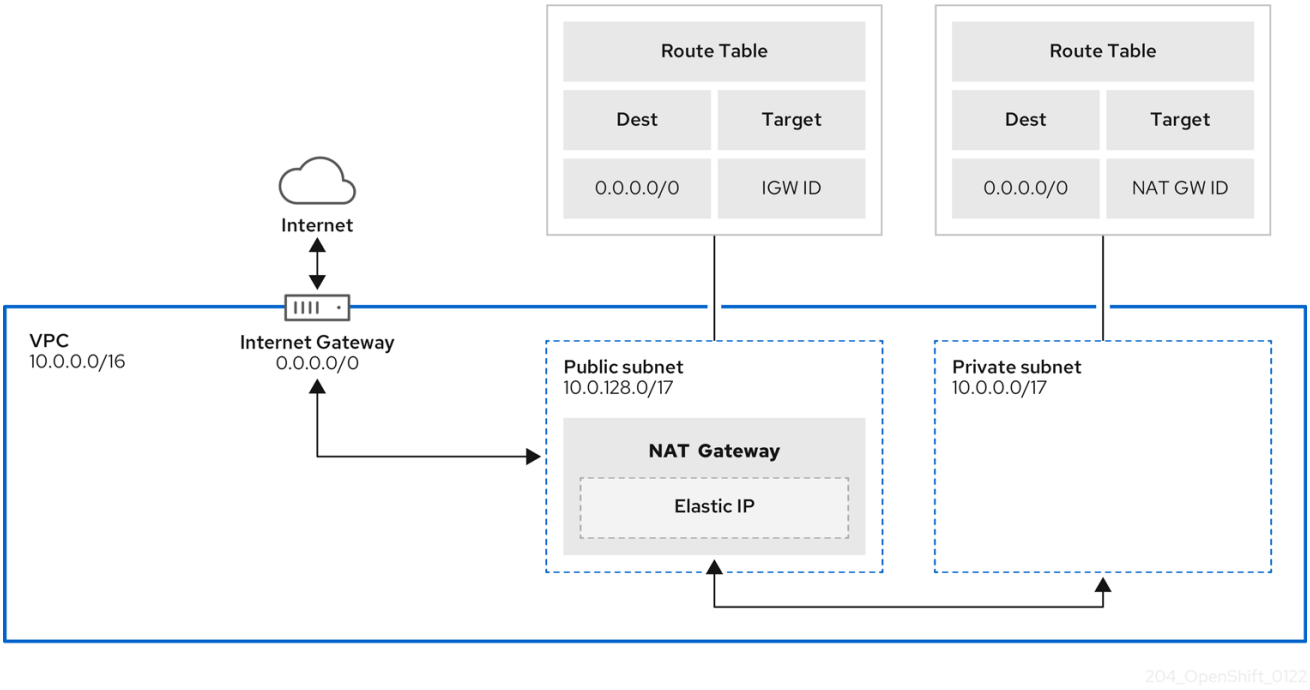
- **Subnets**: Two subnets for a cluster with a single availability zone, or six subnets for a cluster with multiple availability zones.

NOTE

A **public subnet** connects directly to the internet through an internet gateway. A **private subnet** connects to the internet through a network address translation (NAT) gateway.

- **Route tables**: One route table per private subnet, and one additional table per cluster.
- **Internet gateways**: One Internet Gateway per cluster.
- **NAT gateways**: One NAT Gateway per public subnet.

3.6.6. Sample VPC Architecture



3.6.7. Security groups

AWS security groups provide security at the protocol and port-access level; they are associated with EC2 instances and Elastic Load Balancing. Each security group contains a set of rules that filter traffic coming in and out of an EC2 instance. You must ensure the ports required for the OpenShift Container Platform installation are open on your network and configured to allow access between hosts.

3.6.8. Additional custom security groups

When you create a cluster by using a non-managed VPC, you can add custom security groups during cluster creation. Custom security groups are subject to the following limitations:

- You must create the custom security groups in AWS before you create the cluster. For more information, see Amazon EC2 security groups for Linux instances .
- You must associate the custom security groups with the VPC that the cluster will be installed into. Your custom security groups cannot be associated with another VPC.

- You might need to request additional quota for your VPC if you are adding additional custom security groups. For information on requesting an AWS quota increase, see Requesting a quota increase.

3.7. NETWORKING PREREQUISITES

3.7.1. Networking prerequisites

The following sections detail the requirements to create your cluster.

3.7.1.1. Minimum bandwidth

During cluster deployment, OpenShift Dedicated requires a minimum bandwidth of 120 Mbps between cluster infrastructure and the public internet or private network locations that provide deployment artifacts and resources. When network connectivity is slower than 120 Mbps (for example, when connecting through a proxy) the cluster installation process times out and deployment fails.

After cluster deployment, network requirements are determined by your workload. However, a minimum bandwidth of 120 Mbps helps to ensure timely cluster and operator upgrades.

3.7.2. Firewall AllowList requirements

If you are using a firewall to control egress traffic from OpenShift Dedicated, you must configure your firewall to grant access to the certain domain and port combinations below. OpenShift Dedicated requires this access to provide a fully managed OpenShift service.

3.7.2.1. Domains for installation packages and tools

Domain	Port	Function
registry.redhat.io	443	Provides core container images.
quay.io	443	Provides core container images.
cdn01.quay.io	443	Provides core container images.
cdn02.quay.io	443	Provides core container images.
cdn03.quay.io	443	Provides core container images.
cdn04.quay.io	443	Provides core container images.
cdn05.quay.io	443	Provides core container images.
cdn06.quay.io	443	Provides core container images.

Domain	Port	Function
sso.redhat.com	443	Required. The **https://console.redhat.com/openshift** site uses authentication from **sso.redhat.com** to download the pull secret and use Red Hat SaaS solutions to facilitate monitoring of your subscriptions, cluster inventory, chargeback reporting, and so on.
quay-registry.s3.amazonaws.com	443	Provides core container images.
quayio-production-s3.s3.amazonaws.com	443	Provides core container images.
registry.access.redhat.com	443	Hosts all the container images that are stored on the Red Hat Ecosytem Catalog. Additionally, the registry provides access to the **odo** CLI tool that helps developers build on OpenShift and Kubernetes.
access.redhat.com	443	Required. Hosts a signature store that a container client requires for verifying images when pulling them from **registry.access.redhat.com**.
registry.connect.redhat.com	443	Required for all third-party images and certified Operators.
console.redhat.com	443	Required. Allows interactions between the cluster and OpenShift Console Manager to enable functionality, such as scheduling upgrades.
sso.redhat.com	443	The **https://console.redhat.com/openshift** site uses authentication from **sso.redhat.com**.
pull.q1w2.quay.rhcloud.com	443	Provides core container images as a fallback when quay.io is not available.
catalog.redhat.com	443	The **registry.access.redhat.com** and **https://registry.redhat.io** sites redirect through **catalog.redhat.com**.
oidc.op1.openshiftapps.com	443	Used by OpenShift Dedicated for STS implementation with managed OIDC configuration.

3.7.2.2. Domains for telemetry

Domain	Port	Function
cert-api.access.redhat.com	443	Required for telemetry.
api.access.redhat.com	443	Required for telemetry.
infogw.api.openshift.com	443	Required for telemetry.
console.redhat.com	443	Required for telemetry and Red Hat Lightspeed.
observatorium-mst.api.openshift.com	443	Required for managed OpenShift-specific telemetry.
observatorium.api.openshift.com	443	Required for managed OpenShift-specific telemetry.

Managed clusters require enabling telemetry to allow Red Hat to react more quickly to problems, better support the customers, and better understand how product upgrades impact clusters. For more information about how remote health monitoring data is used by Red Hat, see *About remote health monitoring* in the *Additional resources* section.

3.7.2.3. Domains for Amazon Web Services (AWS) APIs

Domain	Port	Function
.amazonaws.com	443	Required to access AWS services and resources.

Alternatively, if you choose to not use a wildcard for Amazon Web Services (AWS) APIs, you must allowlist the following URLs:

Domain	Port	Function
ec2.amazonaws.com	443	Used to install and manage clusters in an AWS environment.
events.<aws_region>.amazonaws.com	443	Used to install and manage clusters in an AWS environment.
iam.amazonaws.com	443	Used to install and manage clusters in an AWS environment.
route53.amazonaws.com	443	Used to install and manage clusters in an AWS environment.

Domain	Port	Function
sts.amazonaws.com	443	Used to install and manage clusters in an AWS environment, for clusters configured to use the global endpoint for AWS STS.
sts.<aws_region>.amazonaws.com	443	Used to install and manage clusters in an AWS environment, for clusters configured to use regionalized endpoints for AWS STS. See AWS STS regionalized endpoints for more information.
tagging.us-east-1.amazonaws.com	443	Used to install and manage clusters in an AWS environment. This endpoint is always us-east-1, regardless of the region the cluster is deployed in.
ec2.<aws_region>.amazonaws.com	443	Used to install and manage clusters in an AWS environment.
elasticloadbalancing.<aws_region>.amazonaws.com	443	Used to install and manage clusters in an AWS environment.
tagging.<aws_region>.amazonaws.com	443	Allows the assignment of metadata about AWS resources in the form of tags.

3.7.2.4. Domains for OpenShift

Domain	Port	Function
mirror.openshift.com	443	Used to access mirrored installation content and images. This site is also a source of release image signatures.
api.openshift.com	443	Used to check if updates are available for the cluster.

3.7.2.5. Domains for your site reliability engineering (SRE) and management

Domain	Port	Function
api.pagerduty.com	443	This alerting service is used by the in-cluster alertmanager to send alerts notifying Red Hat SRE of an event to take action on.

Domain	Port	Function
events.pagerduty.com	443	This alerting service is used by the in-cluster alertmanager to send alerts notifying Red Hat SRE of an event to take action on.
api.deadmanssnitch.com	443	Alerting service used by OpenShift Dedicated to send periodic pings that indicate whether the cluster is available and running.
nosnch.in	443	Alerting service used by OpenShift Dedicated to send periodic pings that indicate whether the cluster is available and running.
http-inputs-osdsecuritylogs.splunkcloud.com	443	Required. Used by the **splunk-forwarder-operator** as a logging forwarding endpoint to be used by Red Hat SRE for log-based alerting.
sftp.access.redhat.com (Recommended)	22	The SFTP server used by **must-gather-operator** to upload diagnostic logs to help troubleshoot issues with the cluster.

Additional resources

- About remote health monitoring

3.8. AWS ACCOUNT LIMITS

The OpenShift Dedicated cluster uses a number of Amazon Web Services (AWS) components, and the default service limits affect your ability to install OpenShift Dedicated clusters. If you use certain cluster configurations, deploy your cluster in certain AWS regions, or run multiple clusters from your account, you might need to request additional resources for your AWS account.

The following table summarizes the AWS components whose limits can impact your ability to install and run OpenShift Dedicated clusters.

Component	Number of clusters available by default	Default AWS limit	Description

Component	Number of clusters available by default	Default AWS limit	Description
Instance Limits	Varies	Varies	At a minimum, each cluster creates the following instances: ● One bootstrap machine, which is removed after installation ● Three control plane nodes ● Two infrastructure nodes for a single availability zone; three infrascture nodes for multi-availability zones ● Two worker nodes for a single availability zone; three worker nodes for multi-availability zones These instance type counts are within a new account's default limit. To deploy more worker nodes, deploy large workloads, or use a different instance type, review your account limits to ensure that your cluster can deploy the machines that you need. In most regions, the bootstrap and worker machines uses an **m4.large** machines and the control plane machines use **m4.xlarge** instances. In some regions, including all regions that do not support these instance types, **m5.large** and **m5.xlarge** instances are used instead.

Component	Number of clusters available by default	Default AWS limit	Description
Elastic IPs (EIPs)	0 to 1	5 EIPs per account	To provision the cluster in a highly available configuration, the installation program creates a public and private subnet for each availability zone within a region. Each private subnet requires a NAT Gateway, and each NAT gateway requires a separate elastic IP. Review the AWS region map to determine how many availability zones are in each region. To take advantage of the default high availability, install the cluster in a region with at least three availability zones. To install a cluster in a region with more than five availability zones, you must increase the EIP limit. **IMPORTANT** To use the **us-east-1** region, you must increase the EIP limit for your account.
Virtual Private Clouds (VPCs)	5	5 VPCs per region	Each cluster creates its own VPC.
Elastic Load Balancing (ELB)	3	20 per region	By default, each cluster creates internal and external Network Load Balancers for the primary API server and a single Classic Load Balancer for the router. Deploying more Kubernetes LoadBalancer Service objects will create additional load balancers.
NAT Gateways	5	5 per availability zone	The cluster deploys one NAT gateway in each availability zone.
Elastic Network Interfaces (ENIs)	At least 12	350 per region	The default installation creates 21 ENIs and an ENI for each availability zone in your region. For example, the **us-east-1** region contains six availability zones, so a cluster that is deployed in that zone uses 27 ENIs. Review the AWS region map to determine how many availability zones are in each region. Additional ENIs are created for additional machines and load balancers that are created by cluster usage and deployed workloads.

Component	Number of clusters available by default	Default AWS limit	Description
VPC Gateway	20	20 per account	Each cluster creates a single VPC Gateway for S3 access.
S3 buckets	99	100 buckets per account	Because the installation process creates a temporary bucket and the registry component in each cluster creates a bucket, you can create only 99 OpenShift Dedicated clusters per AWS account.
Security Groups	250	2,500 per account	Each cluster creates 10 distinct security groups.