

Self-managed Red Hat OpenShift subscription guide

Table of contents

Introduction	2
Definitions	2
Red Hat OpenShift subscription editions	3
What to count for self-managed OpenShift subscriptions	4
Subscription types	5
When to choose core-pair subscriptions	6
When to choose bare-metal node subscriptions	6
How to count AI Accelerator subscriptions	7
What not to count	8
Red Hat OpenShift control plane and infrastructure nodes	8
Red Hat Enterprise Linux entitlements	9
Bootstrap container registry for mirroring OpenShift container images	10
Hosted control planes	10
Exception 1: If you choose to run application instances on control plane or infrastructure nodes	11
Exception 2: Compact cluster deployments	11
Special use cases	11
Disaster recovery	11
Migration and swing upgrades	12
Entitlements for cores with hyperthreading	12
Alternative architectures (Arm, IBM Z, IBM LinuxONE , IBM Power)	12
Microsoft Windows Server containers support	13
Every situation is unique, so these are only guidelines not guarantees	13

Red Hat OpenShift Platform Plus component support	14
Red Hat OpenShift Virtualization Engine and related products	15
How to approach sizing your self-managed OpenShift environment	16
How to estimate your entitlement requirements	16
Step 1: Identify number and type of application instances needed	17
Step 2: Determine the total memory and total cores needed	17
Step 3: Choose a standard compute node (VM, cloud instance or bare-metal server)	18
Step 4: Calculate total subscriptions required	19
Step 5: Calculate the total AI Accelerator subscriptions (if applicable)	20
Example 1: Containerized application running in a hyperscaler cloud	20
Example 2: Containerized application running on premise on bare metal	21
Example 3: VM-only environment	22
Appendix 1: Self-managed OpenShift editions and what they include	25
Table 1: Overall feature differences by Red Hat OpenShift edition	25
Table 2: Detailed differences between Red Hat OpenShift editions, including operators providing those features	26
Common Red Hat software not included in any OpenShift edition	32

Introduction

This document will help you understand the subscription model for self-managed [Red Hat® OpenShift®](#) offerings and provide step-by-step instructions for how to approximate the number of entitlements needed for your Red Hat OpenShift environment. More accurate sizing information is available on request.

Definitions

First, we will define some terms that you will need to understand:

- ▶ **Core-pair:** This is 1 of the bases for self-managed OpenShift subscriptions. It is defined as 2 physical cores or as 4 virtual central processing units (vCPUs). On a bare-metal machine, it always refers to the physical core, regardless of any hyperthreading or symmetric multithreading technology. If hyperthreading is turned on, then 2 physical cores that present as 4 vCPU are still counted as a single core-pair. For hyperscaler deployments, e.g. AWS, Azure, and GCP, 4 vCPU are always considered a single core-pair. Core-pair subscriptions are counted at the cluster level, therefore can span multiple cloud instances, virtual machines (VMs), and physical servers.
- ▶ **Bare-metal node:** This is another base for self-managed OpenShift subscriptions. One bare-metal node subscription is required per physical server. A single bare-metal node subscription cannot span more than 1 server, nor can the cores in a subscription span more than 1 server.

- ▶ **AI Accelerator:** The number of Red Hat AI Accelerator subscriptions required is based on the number of physical hardware add-on devices that accelerate certain compute functions outside of the central processing units (CPUs). These can be, but are not limited to, discrete graphic processing units (GPUs), tensor processing units (TPUs), data processing units (DPUs), field-programmable gate arrays (FPGAs), application-specific integrated circuits (ASICs), and network processing units (NPUs) that are installed as add-ons. This does not include accelerators tightly integrated with the main CPU (like integrated GPU, NPU, and other types of accelerators). While these can often be virtualized and presented to more than 1 VM or instance, and may have 1 or many “cores” similar to CPU cores, the subscription is based on the number of physical devices.
- ▶ **SLA:** A choice of support service-level agreement (SLA) is required for Red Hat subscriptions. The 2 choices are Standard 8x5 or Premium 24x7 support SLA.
- ▶ **Compute node:** Instances (VMs or bare-metal hosts) of Red Hat Enterprise Linux® or Red Hat Enterprise Linux CoreOS where end-user application pods run. OpenShift environments can have many compute nodes. These nodes require OpenShift subscriptions. Control plane nodes and infrastructure nodes that support the compute nodes do not need subscriptions, but may incur infrastructure costs.
- ▶ **Control plane nodes:** Instances (VMs or bare-metal hosts) of Red Hat Enterprise Linux CoreOS that act as Kubernetes orchestration or the management layer for Red Hat OpenShift. Control plane node entitlements are included in self-managed OpenShift subscriptions and do not need to be counted to determine how many subscriptions to purchase. See the Red Hat OpenShift control plane and infrastructure nodes section for more details.
- ▶ **Infrastructure nodes:** Nodes that are running pods supporting an OpenShift cluster infrastructure and not application instances. (For example, nodes running the HAProxy-based load balancer for ingress traffic.) Infrastructure node entitlements are included in self-managed OpenShift subscriptions and do not need to be counted to determine how many subscriptions to purchase as long as no user application instances are running on them.
- ▶ **Cluster:** An OpenShift Kubernetes cluster consisting of a control plane, optional infrastructure nodes, and 1 or more compute nodes.
- ▶ **Application instance:** An application may be a single-pod instance or may be deployed across multiple-pod instances that make up an application service. For example, a highly available application service may consist of 2 or more pods. Application instances must always run on entitled compute nodes with Red Hat OpenShift subscriptions.

Red Hat OpenShift subscription editions

Red Hat OpenShift provides a consistent application development and management platform across an open hybrid cloud environment, and supports on-premise, virtual, and physical infrastructure, and private cloud, public cloud, and edge deployments. There are 2 ways to operate and use Red Hat OpenShift: self-managed OpenShift and fully managed OpenShift cloud services. This guide focuses on self-managed OpenShift.

Self-managed OpenShift allows you to install, operate, and manage Red Hat OpenShift environments with maximum control, flexibility, and customization, so you can operate your own environment starting with the infrastructure. Self-managed OpenShift is supported

on-premise—using physical servers, virtualization, and in a private cloud environment—and in supported public cloud environments. You control upgrades, manage the lower-level infrastructure, and maintain service-level agreements (SLA).

Managed OpenShift cloud services are fully managed and operated by Red Hat and its public cloud partners in major public clouds. A dedicated site reliability engineering (SRE) team manages and maintains Red Hat OpenShift core services and infrastructure, allowing your DevSecOps teams to concentrate on developing and deploying new applications and modernizing existing ones.

OpenShift cloud service offerings and where to find more information:

- ▶ **Red Hat OpenShift Dedicated:** Fully managed Red Hat OpenShift service on Amazon Web Services (AWS) and Google Cloud.
- ▶ **Microsoft Azure Red Hat OpenShift:** Fully managed Red Hat OpenShift service on Microsoft Azure, jointly managed by Red Hat and Microsoft.
- ▶ **Red Hat OpenShift Service on AWS:** Fully managed Red Hat OpenShift service on Amazon Web Services, jointly managed by Red Hat and AWS.
- ▶ **Red Hat OpenShift on IBM Cloud:** Fully managed Red Hat OpenShift service on IBM Cloud, jointly managed by Red Hat and IBM.

All editions of Red Hat OpenShift offer a consistent user experience for developers and operations teams across every environment, allowing you to transfer your skills and applications to the cloud environments where your applications run best.

What to count for self-managed OpenShift subscriptions

Self-managed OpenShift (Red Hat OpenShift Platform Plus, Red Hat OpenShift Container Platform, Red Hat OpenShift Kubernetes Engine, and Red Hat OpenShift Virtualization Engine) can be used anywhere 64-bit Red Hat Enterprise Linux is certified and supported. Refer to [the documentation](#) to learn more about [OpenShift deployment methods and supported infrastructure types](#).

Self-managed OpenShift software editions:

- ▶ **Red Hat OpenShift Kubernetes Engine:** A hybrid cloud, enterprise Kubernetes runtime engine that provides core Red Hat OpenShift functionality to deploy and run applications, which you install and manage in a datacenter, public cloud, or edge environment.
- ▶ **Red Hat OpenShift Container Platform:** A full-featured hybrid cloud, enterprise Kubernetes application platform used to build, deploy, and run applications, which you install and manage in a datacenter, public cloud, and edge environment.
- ▶ **Red Hat OpenShift Platform Plus:** A hybrid cloud platform that allows enterprises to build, deploy, run, and manage intelligent applications at scale across multiple clusters and cloud environments. Multiple layers of security, manageability, and automation provide consistency throughout the software supply chain. OpenShift Platform Plus subscriptions are available for x86-based clusters only.
- ▶ **Red Hat OpenShift Virtualization Engine:** A bare metal-only, virtualization infrastructure offering based on Red Hat OpenShift and the open source kernel-based virtual machines (KVM) hypervisor, purpose-built to provide enterprises with a reliable, enterprise-grade solution for

deploying, managing, and scaling VMs. A subset of OpenShift functionality, this edition of OpenShift is targeted at virtual machine-only workloads, with only infrastructure services supported in containers (i.e., no end-user application containers support).

Subscription types

There are 2 types of subscriptions (core-pair and bare-metal node) for self-managed OpenShift with 2 support levels available for each subscription.

Compute node subscriptions are required for the compute nodes in your environment. These can be entitled by core-pairs or by bare-metal nodes:

1. Core-pair (2 cores or 4vCPUs)

- ▶ This subscription option is available for OpenShift Kubernetes Engine, OpenShift Container Platform, and OpenShift Platform Plus. Core-pair subscriptions are **not** applicable to OpenShift Virtualization Engine.
- ▶ When entitling CPU cores, count the aggregate number of physical cores or vCPUs across all OpenShift compute nodes running across all OpenShift clusters you want to entitle using core-pair subscriptions.
- ▶ Available with Standard 8x5 or Premium 24x7 support SLA.

2. Bare-metal node (1 physical node)

- ▶ A physical node is 1 server regardless of the number of CPU sockets in the server, or cores in the CPUs.
- ▶ This subscription option is available for all self-managed OpenShift editions, and is the only option for OpenShift Virtualization Engine.
- ▶ This subscription is available only for x86 and Arm bare-metal physical nodes where OpenShift is installed directly to the hardware. No third-party hypervisor is allowed.
- ▶ This is explicitly **not** a “virtual datacenter” subscription (like Red Hat Enterprise Linux for Virtual Datacenters where a single subscription gives you unlimited VM guest operating system [OS] installations on any hypervisor host).
- ▶ Available with Standard 8x5 or Premium 24x7 support SLA.

Additionally, you will require Red Hat AI Accelerator subscriptions for the accelerators in your environment:

1. AI Accelerator (1 Accelerator)

- ▶ This subscription is required for accelerator cards (GPU, TPU, NPU, FPGA, DPU, etc.) that are providing compute acceleration for AI workloads that are discrete add-ons and not part of the CPU package.
- ▶ The same subscription is used for each physical AI Accelerator regardless of Red Hat OpenShift edition.
- ▶ A single AI accelerator subscription is sufficient for Red Hat OpenShift and OpenShift AI if both are installed on the cluster.

- ▶ This subscription is not required as long as the accelerator capability is not being used for compute acceleration (for example, DPUs used as SmartNICs for network acceleration only even if they have addressable Arm cores on them that are unused, or GPUs used for rendering graphics and not for AI Acceleration).
- ▶ Available with Standard 8x5 or Premium 24x7 support SLA. The SLA must match the SLA of the supporting core-pair or bare-metal node subscription.

When to choose core-pair subscriptions

You will most often choose core-pair subscriptions when you are deploying self-managed OpenShift to public cloud hyperscalers, within an Infrastructure-as-a-Service (IaaS) private cloud, or when you are deploying on a hypervisor such as VMware vSphere, Red Hat OpenStack® Platform, or Nutanix.

Core-pair subscriptions remove the need to attach subscriptions to physical servers and allow the freedom for pods to be deployed across your hybrid cloud whenever and wherever needed.

You can also use core-pair subscriptions on bare-metal servers or devices (i.e., with no hypervisor). Be aware there is usually a compute pod density where bare-metal node subscriptions may be more cost effective.

When using OpenShift Virtualization Engine as a dedicated virtualization platform, you can choose to entitle OpenShift containers on VMs using core-pair subscriptions on top of the bare-metal node subscriptions for the hypervisor itself. You would separately purchase OpenShift self-managed core-pair subscriptions and assign them to the VMs in this environment, just like any other application you may purchase and run as a VM. In this instance, there is a density of cores where switching to the bare-metal node model for self-managed OpenShift, which includes unlimited OpenShift containers on the bare-metal server and support for running those containers in OpenShift VMs, may be more cost effective.

Core-pair subscriptions can be distributed to cover all OpenShift compute nodes across all OpenShift clusters. For example, 100 core-pair Red Hat OpenShift Platform Plus subscriptions will provide 200 cores (400 vCPUs) that can be used across any number of compute nodes, across all running OpenShift clusters across your hybrid cloud environments.

When to choose bare-metal nodes subscriptions

Bare-metal node subscriptions are only an option for OpenShift compute nodes deployed to dedicated physical servers, either in your datacenter, in a hosted private cloud on a supported bare-metal offering, or with a hyperscaler on a supported bare-metal offering. Bare-metal node subscriptions are the only option for OpenShift Virtualization Engine, and are required for support of the OpenShift Virtualization feature in the other self-managed OpenShift editions.

Each bare-metal node subscription entitles a single physical node, regardless of the number of total CPU sockets or cores.

Each physical, bare-metal server using per node entitlements can only be used as a single OpenShift node due to the inherent architecture of Kubernetes. Since each node in Kubernetes can only belong to a single cluster, this means that all containers on a bare-metal server will be in the same cluster. This is suitable for resource intensive workloads like OpenShift Virtualization (where each workload is running a full VM), but may not be suitable for others. While OpenShift supports up to 2500

containers on a single node, there are situations, whether for performance or architectural reasons, where you might want to split containers between different nodes or different clusters. This is not possible without using virtualization to create separate compute nodes on the bare-metal server.

A common deployment model for containers is to architect a large number of clusters with smaller numbers of containers each. This model is common in hyperscaler environments, and can be achieved in the datacenter by using a hypervisor to create VMs which become the compute nodes on which containers are deployed. In the case of hypervisors such as VMware vSphere, Red Hat OpenStack Platform, and Nutanix, you must use core-pair subscriptions for OpenShift deployed in VMs.

OpenShift Kubernetes Engine, OpenShift Container Platform, and OpenShift Platform Plus clusters, deployed to bare-metal and entitled with node subscriptions, include OpenShift Virtualization and subscriptions for any virtual OpenShift clusters of the same product type deployed to them. This means that virtual OpenShift clusters deployed to, for example, a bare metal OpenShift Container Platform cluster inherit OpenShift Container Platform subscriptions from the hosting bare metal cluster.

An important note is that the OpenShift Virtualization Engine subscription does **not** include support for containerized application instances, the exception being infrastructure workloads as defined in the OpenShift Virtualization Engine section below. If you wish to run your own containerized application workloads with OpenShift Virtualization Engine you need to entitle the VMs using core-pair subscriptions for self-managed OpenShift. Alternatively, at higher densities you can choose instead to purchase a bare-metal node subscription to self-managed OpenShift Kubernetes Engine, OpenShift Container Platform, or OpenShift Platform Plus, which would result in the ability to run container-based applications natively on the bare metal cluster or virtual clusters inherit subscriptions as described in the previous paragraph.

It is not possible to mix OpenShift product types within the same cluster, all nodes must be subscribed using the same OpenShift Virtualization Engine, OpenShift Kubernetes Engine, OpenShift Container Platform, or OpenShift Platform Plus product type, however core-pair and node subscriptions can be used within a single cluster. For example, you cannot have a single bare-metal cluster with some number of OpenShift Virtualization Engine nodes for hosting VMs and other nodes subscribed using OpenShift Platform Plus for hosting containerized applications and virtual OpenShift instances.

How to count AI accelerators subscriptions

In recent years, specific hardware technologies have come onto the market that allow certain compute workloads to perform with increased speed. These types of hardware devices are collectively referred to as accelerators or AI accelerators in some Red Hat content. There are several types of hardware devices available for modern servers that can be classified as accelerators, including but not limited to GPUs, TPUs, ASICs, NPU, and FPGAs.

These accelerators are typically a card, board, or other physical device occupying a peripheral component interconnect (PCI) slot in a server. This is almost always the unit quantity you have purchased from the accelerator vendor. For example, in a server whose vendor says “it has 8 GPUs,” this almost always means 8 physical accelerator units.

Each AI accelerator subscription covers 1 physical accelerator device. For example, focusing just on the AI accelerator subscriptions:

- ▶ A physical compute node with 4 physical GPU devices would require 4 x AI accelerator subscriptions in addition to the CPU core-pair or node subscriptions covering the compute node.
- ▶ A single virtual compute node with 1 physical GPU device presented to the VMs as multiple vGPUs would require 1 AI Accelerator subscription, as counting is based on physical accelerators, not virtual accelerators.

Accelerators are only counted when they are used to execute a compute workload. A workload is considered a compute workload when the primary purpose of the workload is not actively drawing pixels on a user's screen in near real-time or moving data across a network.

This distinction is important because VFX (visual effects) and streaming applications may use GPUs and other accelerator hardware, but the primary purpose in these cases is drawing pixels on a screen. In some cases, the primary function is moving data across a network (such as data processing units dedicated to network functions), which also is not considered compute.

Examples of compute workloads include:

- ▶ Traditional software applications, such as Java, Python, and Perl.
- ▶ LLMs or other compute-intensive software.
- ▶ Data science model training and tuning.
- ▶ Scientific modeling and physical simulation like protein folding and fluid dynamics.

What not to count

Red Hat OpenShift control plane and infrastructure nodes

Each self-managed OpenShift subscription includes entitlements for Red Hat OpenShift and other OpenShift-related components. OpenShift Kubernetes Engine, OpenShift Container Platform, and OpenShift Platform Plus includes Red Hat Enterprise Linux entitlements for compute and infrastructure nodes. Red Hat OpenShift Virtualization Engine does not support standard Red Hat Enterprise Linux worker or infrastructure nodes, only the Red Hat Enterprise Linux CoreOS OS that is part of the OpenShift entitlement.

These entitlements are included to run the required OpenShift control plane and infrastructure, but do not need to be counted when determining the number of Red Hat subscriptions needed.

Red Hat OpenShift Platform Plus subscriptions also include the management of all nodes by Red Hat Advanced Cluster Security for Kubernetes and Red Hat Advanced Cluster Management for Kubernetes.

A default install deploys a highly available OpenShift control plane composed of 3 control plane nodes, in addition to a minimum of 2 OpenShift compute nodes to run end-user applications. By default, Kubernetes control plane components (e.g., application programming interface (API) server, etcd, scheduler) and supporting cluster services (e.g. monitoring, registry) will be deployed on the OpenShift control plane nodes. However, you may decide to move some of these supporting cluster services to dedicated "infrastructure nodes".

To qualify as an infrastructure node and use the included entitlement, only components that are supporting the cluster, and not part of an end-user application, may be running on those instances. Examples include:

- ▶ OpenShift registry.
- ▶ OpenShift Ingress Router (local and global and multicluster ingress).
- ▶ OpenShift Observability.
- ▶ HAProxy-based instances used for cluster ingress.
- ▶ Red Hat Quay.
- ▶ Red Hat OpenShift Data Foundation.
- ▶ Red Hat Advanced Cluster Management for Kubernetes.
- ▶ Red Hat Advanced Cluster Security for Kubernetes.
- ▶ Red Hat OpenShift GitOps.
- ▶ Red Hat OpenShift Pipelines.
- ▶ Hosted control planes for Red Hat OpenShift.
- ▶ Ansible Automation Platform, when used only for cluster management.

You may deploy and run custom and third-party agents and tools for monitoring, log data collection and forwarding, hardware drivers, infrastructure integration, such as virtualization agents, to infrastructure nodes without disqualifying the node for infrastructure licensing. However, this is limited only to agents and related components, including controller pods for operators and does not include the custom or third-party software suite. Examples of non-Red Hat software that qualify as infrastructure workload include:

- ▶ Custom and third-party monitoring agents.
- ▶ Container network interface (CNI) and container storage interface (CSI) drivers and controllers (plug-ins).
- ▶ Hardware or virtualization enablement accelerators.
- ▶ Controller pods used for Kubernetes custom resource definitions (CRD) or operators (custom or third-party software).

No other end-user application instances or types may be run on an infrastructure node using the included entitlement. To run other infrastructure workloads as application instances on Red Hat OpenShift, you must run those instances on regular application nodes with valid Red Hat OpenShift subscriptions. You can verify with Red Hat whether an application or service qualifies as an infrastructure workload.

Red Hat Enterprise Linux entitlements

Red Hat Enterprise Linux entitlements for OpenShift compute and infrastructure nodes are included with the OpenShift Kubernetes Engine, OpenShift Container Platform, and OpenShift Platform Plus. This also includes entitled Pods for applications and guest OS entitlements for VMs. However, Red Hat OpenShift subscriptions do not include other entitlements for Red Hat Enterprise Linux nodes with the following exception:

- ▶ A Red Hat Enterprise Linux node used specifically for bare-metal installer-provisioned infrastructure (IPI) provisioning.

OpenShift Virtualization Engine does **not** include Red Hat Enterprise Linux neither for compute and infrastructure nodes nor for guest OS entitlements of VMs. A Red Hat Enterprise Linux for Virtual Datacenters subscription or per-VM subscriptions are required for Red Hat Enterprise Linux guests on OpenShift Virtualization Engine.

Red Hat Enterprise Linux entitlements are not included for external nodes hosting services that OpenShift depends on, such as internet proxies, load balancers, or the mirror registry. Red Hat Satellite is not included as part of the entitlement.

Bootstrap container registry for mirroring OpenShift container images

The mirror registry for OpenShift is a Quay entitlement for the single purpose of easing the process of mirroring content required for bootstrapping disconnected OpenShift clusters and is included as part of the Red Hat OpenShift subscription. This is a limited support entitlement for a minimal Quay deployment created by a specific installer, which allows you to run a Quay registry on a preprovisioned and customer-managed Red Hat Enterprise Linux host.

Note: You are permitted to use Quay as a registry mirror limited to the purpose of mirroring the OpenShift release payload, OperatorHub content, sample operator images, Cincinnati graph image, and images related to the Red Hat OpenStack Platform offerings including Red Hat OpenStack Services on OpenShift, and Red Hat OpenShift Data Foundation.

The mirror registry for OpenShift is **not** intended to be a general-purpose registry working at arbitrary scale. Nonetheless, a limited set of custom images is permitted to be stored there, which contains required auxiliary software, for example agents and container images for qualified infrastructure workloads. Examples of these may include:

- ▶ Monitoring agents.
- ▶ CNI and CSI providers.
- ▶ Hardware or virtualization enablement agents.
- ▶ Operators supporting independent software vendor (ISV) services.
- ▶ Custom operators as deployment controllers.

Hosted control planes

The classic OpenShift infrastructure requires a minimum of 3 control plane nodes for each OpenShift cluster. An alternative is to use hosted control planes, where the control planes run on a central control plane cluster and provide the logical control plane for OpenShift clusters. As always, control plane nodes and infrastructure nodes do not count for subscriptions, but do need to be accounted for in the architecture.

Hosted control planes may be run on any bare metal OpenShift cluster, however the compute nodes for those clusters will need entitlements in line with the infrastructure they are running on. For example, virtual clusters hosted on OpenShift Virtualization Engine will need core-pair subscriptions for the compute nodes, whereas a cluster whose control plane is hosted by an OpenShift Virtualization Engine cluster and using bare-metal compute nodes would need bare-metal node subscriptions.

Exception 1: If you choose to run application instances on control plane or infrastructure nodes

OpenShift control plane nodes are not used as compute nodes by default and, therefore, will not run application instances. Whether a control plane node requires a full Red Hat OpenShift subscription depends on whether it runs supporting OpenShift cluster components only or end-user applications. If you choose to use a control plane node as a node for hosting any end-user applications, then it requires subscriptions for all cores. See the infrastructure nodes section to identify qualified workloads that do not require a subscription.

Exception 2: Compact cluster deployments

In compact cluster deployments like 3-node OpenShift, end-user application workloads are run on the control plane nodes. You would need to count the cores on the 3 nodes for Red Hat OpenShift subscriptions regardless of the role they play.

A single node OpenShift instance deploys all OpenShift services and end-user applications to a single physical or virtual node, with optimizations to reduce the footprint and maximize resources available to applications. As with compact 3-node clusters above, there are no special accommodations for this deployment model, all cores on the node need to be entitled.

Special use cases

Disaster recovery

Red Hat defines 3 types of disaster recovery (DR) environments: hot, warm, and cold. Paid Red Hat OpenShift subscriptions are needed for hot DR only.

- ▶ Hot DR systems are defined as fully functional and running concurrently with the production systems. They are ready to immediately receive traffic and take over in the event of a disaster within the primary environment. When data volumes are actively being replicated—either synchronously or asynchronously, by OpenShift-hosted software-defined storage between clusters—they are considered to be “hot” DR systems.
- ▶ Warm DR systems are defined as already prepared to deploy and host containerized workload representing a reasonable facsimile of that found in the primary site, but contain no customer workload from the source cluster(s). Warm DR clusters should **not** be participating in active data volume replication, either synchronously or asynchronously, between clusters when the secondary (DR) site is mounting said data into container workloads actively running on the cluster. External storage devices performing volume replication are still considered warm DR. For OpenShift-hosted software-defined storage, warm DR recovery requires the customer’s data be restored onto the existing cluster hardware from outside of the cluster.
- ▶ Cold DR systems are defined as having the infrastructure in place, but not the full technology (hardware, software, and data) needed to restore service.

Hibernating clusters that are not expressly configured and designed for warm or cold DR require subscriptions—such as clusters running on cloud services that are temporarily hibernating as a result of lower demand. Subscriptions are required when warm or cold DR clusters are brought out of hibernation for running workloads. Bringing a cluster out of hibernation temporarily for routine maintenance or tests does not require an additional subscription for any of the components in Red Hat OpenShift software offerings.

For both warm DR and cold DR, Red Hat OpenShift subscriptions can be transferred from the primary environment to the DR environment when the disaster occurs to restore service and maintain compliance with Red Hat's subscription terms.

M i g r a t i o n a n d s w i n g u p g r a d e s

Red Hat OpenShift 4 provides in-place upgrades between minor versions. Nevertheless, if you are upgrading from Red Hat OpenShift 3, or need to perform a swing upgrade between Red Hat OpenShift 4 minor versions as a result of maintenance windows or other considerations, your Red Hat OpenShift subscription will cover both the original and destination infrastructure of a 1-way migration until such migration is complete. During the migration, Red Hat's subscription management tools will show your environment as being out-of-compliance on the basis of the number of Red Hat OpenShift subscriptions purchased. Red Hat allows this for major version upgrades and will not require the purchase of additional subscriptions to get back into compliance during the migration. Finally, Red Hat OpenShift provides tooling to assist in these migrations, along with Red Hat consulting services, if desired. See documentation on the [migration toolkit for containers](#).

E n t i t l e m e n t s f o r c o r e s w i t h h y p e r t h r e a d i n g

Making a determination about whether or not a particular OpenShift node uses 1 or more physical cores is determined by whether or not that system has multiple threads per core enabled. [Learn how](#) to determine whether a particular system supports hyperthreading.

Virtualized OpenShift nodes using logical CPU threads, also known as simultaneous multithreading (SMT) for AMD EPYC CPUs or hyperthreading with Intel CPUs, calculate their core utilization for Red Hat OpenShift subscriptions based on the number of cores/CPU threads assigned to the node, however each subscription covers 4 vCPUs/cores when logical CPU threads are used. Red Hat's subscription management tools assume logical CPU threads are enabled by default on all systems.

Bare-metal subscriptions only count physical cores, logical CPU threads are not counted when calculating the number of subscriptions needed for bare metal.

A l t e r n a t i v e a r c h i t e c t u r e s (A r m | B M Z | B M L i n u x O N E | B M P o w e r)

Note: While this document refers only to IBM Z from here on, all information that references IBM Z also applies to IBM LinuxONE.

Red Hat OpenShift Container Platform can also run on Arm, IBM Z, and IBM Power for customers using these platforms as the standard for building and deploying cloud-native applications and microservices. Only the core-based subscription model is supported for IBM Z and IBM Power platforms. Red Hat OpenShift socket-based bare-metal subscriptions cannot be used with IBM Z and IBM Power platforms.

Arm clusters are entitled using the same rules as x86.

For IBM Z customers, Red Hat OpenShift does not require the entire physical node to be entitled, only the cores used by OpenShift. IBM Z customers know this as "subcapacity" entitlement. Customers using only a subset of the available cores (compute capacity) on their IBM Z environment for OpenShift Container Platform only require subscriptions for the subset that is used for the compute nodes. This applies regardless of how CPU partitioning is achieved, whether by CPU pooling, capping, separate logical partitions (LPARs), or other means.

For IBM Z, 1 Integrated Facility for Linux (IFL) requires 1 OpenShift core subscription (IBM Z is the only architecture where the base for subscriptions is a single core). There is no concept of vCPU in this architecture. When no partitioning is used, up to 3 IFLs per cluster may be identified per OpenShift cluster for control plane or infrastructure services running on the host. These must be actively used for control plane and/or infra services to qualify and do not require OpenShift entitlements. Three-node compact cluster deployments require all IFLs to be entitled.

For IBM Power, the core-pair subscription model is used. There is no concept of vCPU in this architecture.

The components of the Red Hat OpenShift Platform Plus subscription have different levels of support for alternative (non-x86) architectures. See the article, [“Red Hat OpenShift Container Platform 4.19 Multi Architecture Component Availability Matrix,”](#) on the Red Hat Customer Portal (requires login access). For details about interoperability between the OpenShift Platform Plus components and non-x86 architectures, see the compatibility matrices for [Red Hat OpenShift Container Platform](#), [Red Hat Advanced Cluster Management](#), [Red Hat Advanced Cluster Security](#), [Red Hat Quay](#) and [Red Hat OpenShift Data Foundation](#).

Red Hat OpenShift Kubernetes Engine and Red Hat OpenShift Virtualization Engine are not supported on IBM Z and IBM Power.

M i c r o s o f t W i n d o w s S e r v e r c o n t a i n e r s s u p p o r t

Self-managed OpenShift supports a subset of installation infrastructures and OpenShift features using Microsoft Windows Server containers. Windows Server containers are only supported on Microsoft Windows Server-based compute nodes. The control and infrastructure planes of the OpenShift environment must be running on x86 infrastructure using Red Hat Enterprise Linux or Red Hat Enterprise Linux CoreOS. For this reason, Windows Server container support is sold as a standalone subscription priced by core.

Red Hat OpenShift Platform Plus, Red Hat OpenShift Container Platform, and Red Hat OpenShift Kubernetes Engine infrastructure can be used to deploy and manage Windows Server compute nodes. Microsoft Windows Server container support for Red Hat OpenShift subscriptions must be purchased as a separate add-on.

Red Hat Advanced Cluster Management and Red Hat Advanced Cluster Security do not support managing Microsoft Windows nodes, but Red Hat Quay running on x86 architectures can manage container images for Microsoft Windows Server-based workloads.

E v e r y s i t u a t i o n i s n i q u e & t h e s e a r e o n l y g u i d e l i n e s n o t g u a r a n t e e s

Red Hat OpenShift supports many features and functions which affect scalability, Pod scheduling, idling, and resource quota/limiting features. The previous calculations are guidelines, and you may be able to tune your actual environment for better resource use or smaller total environment size. OpenShift Platform Plus customers should take into account the needs of the additional software applications (Red Hat Advanced Cluster Management, Red Hat Advanced Cluster Security, and Quay) including storage and compute resources, even though they may not require additional compute subscriptions.

If working with a third-party reseller, refer to their specific terms and agreements for Red Hat products and services.

Red Hat OpenShift Platform Plus component support

Red Hat OpenShift Platform Plus includes additional software beyond OpenShift Container Platform to help you manage and safeguard your Red Hat OpenShift environment at scale across multiple clusters and multiple clouds. OpenShift Platform Plus is available both in the core-pair and bare-metal node subscription models with the previously mentioned limitations.

The additional software included with Red Hat OpenShift Platform Plus is generally limited to managing the nodes entitled with OpenShift Platform Plus subscriptions. For example, the subscription for Red Hat Advanced Cluster Management included with OpenShift Platform Plus can only be used to manage OpenShift Platform Plus entitled nodes and clusters. Customers who also wish to manage non-OpenShift Platform Plus entitled nodes and clusters, for example Red Hat OpenShift Services on AWS clusters, would need to purchase additional Red Hat Advanced Cluster Management add-on subscriptions to cover those clusters.

The additional software subscriptions are also inseparable from the OpenShift Platform Plus subscription. For example, you cannot purchase 100 OpenShift Platform Plus subscriptions, install 200 cores of Red Hat OpenShift Container Platform subscriptions, and separately use Red Hat Advanced Cluster Management to manage 200 cores of Azure Red Hat OpenShift with the same subscription. The additional software can only be used to manage the same 200 cores on which the core OpenShift Platform Plus software is installed.

Specific rules for each layered product are:

- ▶ **Red Hat Advanced Cluster Management for Kubernetes:** An OpenShift Platform Plus subscription allows you to install as many Red Hat Advanced Cluster Management central instances as needed to manage your environment, and covers the management of all nodes and clusters entitled with OpenShift Platform Plus, including control plane and infrastructure nodes. If you wish to manage nodes and clusters without OpenShift Platform Plus entitlements (for example, if you also have self-managed OpenShift Container Platform or Red Hat OpenShift Kubernetes Engine entitled clusters, clusters running in a fully managed OpenShift cloud, or third-party Kubernetes environments supported by Red Hat Advanced Cluster Management), then you need to purchase Red Hat Advanced Cluster Management add-on subscriptions to cover those environments. You can choose to manage them centrally from the Red Hat Advanced Cluster Management console installed on OpenShift Platform Plus, or from a separate central application if that meets your requirement. [Learn more](#) about Red Hat Advanced Cluster Management subscriptions, Red Hat Advanced Cluster Management supported environments, and Red Hat Advanced Cluster Management best practices.
- ▶ **Red Hat Advanced Cluster Security for Kubernetes:** The OpenShift Platform Plus subscription allows you to install as many Red Hat Advanced Cluster Security central applications as needed to manage your environment, and covers the management of all nodes and clusters entitled with OpenShift Platform Plus, including control plane and infrastructure nodes. If you want to manage nodes and clusters without OpenShift Platform Plus entitlements (for example, if you also have self-managed OpenShift Container Platform or OpenShift Kubernetes Engine entitled clusters, clusters running in a fully managed Red Hat OpenShift cloud, or third-party Kubernetes environments supported by Red Hat Advanced Cluster Security) you need to purchase Red Hat Advanced Cluster Security add-on subscriptions to cover those environments. Red Hat's

suggested practice is to manage each environment with a separate Red Hat Advanced Cluster Security central application. [Learn more](#) about Red Hat Advanced Cluster Security supported environments.

- ▶ **Red Hat Quay:** The OpenShift Platform Plus subscription allows you to install Red Hat Quay on any cluster that has an OpenShift Platform Plus entitlement. There is no limit on the number of Quay deployments you can install on your OpenShift Platform Plus clusters. Quay can then serve any supported Kubernetes environment you wish, including the OpenShift Platform Plus environment, other self-managed OpenShift clusters, managed OpenShift services, and supported third-party Kubernetes. If you wish to install Quay in a non-OpenShift Platform Plus environment, you will need to purchase a separate Red Hat Quay subscription. Quay is also available as a fully managed [Software-as-a-Service \(SaaS\) offering](#).
- ▶ **Red Hat OpenShift Data Foundation:** The OpenShift Platform Plus subscription allows you to install Red Hat OpenShift Data Foundation Essentials on any cluster that has an OpenShift Platform Plus entitlement. The Red Hat Data Foundation entitlement is limited to the features available in Essentials, and to 256 TB of data storage per OpenShift cluster. You can choose to extend functionality and capacity through additional subscriptions. See the [OpenShift Data Foundation Subscription guide](#) (Customer Portal login required) or consult with a Red Hat sales representative for further guidance.
- ▶ **Argo CD Agent:** The OpenShift Platform Plus subscription allows you to install the Argo CD Agent component of Red Hat OpenShift GitOps on any cluster that has an OpenShift Platform Plus subscription. Argo CD Agent can run on Red Hat OpenShift or supported Kubernetes environments. (Red Hat OpenShift GitOps is included with OpenShift Container Platform.)

Red Hat OpenShift Virtualization Engine and related products

OpenShift Virtualization has long been an included feature with all self-managed OpenShift editions, so that customers can include VM workloads in cloud-native applications and modernize VMs into microservices and containers.

Recent changes in the virtualization market have increased the demand for alternative virtualization platforms, especially one that offers a path to modernization. These customers are focused exclusively on virtualization with an option to adopt modern application development and containerization capabilities at their own pace.

Red Hat OpenShift Virtualization Engine is a self-managed OpenShift edition specifically aimed at customers who want an alternative virtualization platform for running VMs. OpenShift Virtualization Engine is entitled by bare-metal node subscriptions on supported physical servers and on supported hyperscaler bare-metal instances. When users are ready to upgrade to the broader portfolio, they can change their subscription entitlements and do not need to redeploy existing clusters.

OpenShift Virtualization Engine includes only the features required to deploy, manage and run virtual machines, such as:

- ▶ Unlimited VMs on subscribed hosts.
- ▶ Cannot be used to run application instances (like commercial software or customer applications) in containers, only VMs.
- ▶ Does not include subscriptions for running any Red Hat OpenShift editions inside VMs (separate purchase of core-pair subscriptions required).

- ▶ Red Hat Enterprise Linux guest entitlements for VMs are *not included* (separate purchase of Red Hat Enterprise Linux for Virtual Datacenters or per-VM subscriptions required).

Red Hat Advanced Cluster Management and Red Hat Ansible® Automation Platform both support OpenShift Virtualization Engine. Add-on subscriptions for both products are available specifically for virtualization-only use cases at a reduced cost compared to the full version.

For existing Red Hat Advanced Cluster Management and Ansible Automation Platform customers, you can use your existing installations of these central applications to support the rest of your environment and to manage the OpenShift Virtualization Engine hosts with the purchase of the mentioned add-on subscriptions.

For customers who do not have existing Red Hat Advanced Cluster Management or Ansible Automation Platform central applications installed, your subscription for the mentioned add-ons also includes support for installing the central applications as infrastructure containers on your OpenShift Virtualization Engine hosts. See the documentation for Red Hat Advanced Cluster Management or Ansible Automation Platform on installing these central applications and best practices for architecture.

For customers who need Day 2 automation for their VMs, Ansible Automation Platform is recommended. In addition to the hypervisor node subscriptions listed above, customers will need a node subscription for each VM instance. [See Ansible Automation Platform documentation](#) for more details.

While the entitlement for OpenShift Virtualization Engine restricts customer application instances to VMs only, many of the infrastructure requirements, such as storage drivers, backup applications, forwarding agents, Red Hat Advanced Cluster Management, and Ansible Automation Platform, run as infrastructure containers in Red Hat OpenShift. Therefore your entitlement does allow you to run those infrastructure features in containers. In addition, software defined storage solutions that provide storage for VMs, but themselves run in containers are also included in the entitlement. The guidance on what qualifies for these infrastructure containers is the same as what workloads are allowed on Infrastructure nodes for other editions of Red Hat OpenShift. See the information about infrastructure nodes in the “What not to count” section. Verify with a Red Hatter whether an application or service qualifies as an infrastructure workload for OpenShift Virtualization Engine.

How to approach sizing your self-managed OpenShift environment

To determine how many self-managed OpenShift (Red Hat OpenShift Platform Plus, Red Hat OpenShift Container Platform, or Red Hat OpenShift Kubernetes Engine) or add-on subscriptions you need, use the following questions and examples.

In summary:

- ▶ Applications are packaged in container images or VMs.
- ▶ Containers and VMs are deployed as pods.
- ▶ Pods run on OpenShift compute nodes, which are managed by the OpenShift control plane nodes.

How to estimate your entitlement requirements

Red Hat OpenShift subscriptions do not limit application instances. You can run as many application instances in the Red Hat OpenShift environment as the underlying hardware and infrastructure will support. Larger-capacity hardware can run many application instances on a small number of hosts,

while smaller-capacity hardware may require many hosts to run many application instances. The primary factor in determining the size of a Red Hat OpenShift environment is how many pods, or application instances, will be running at any given time.

Step 1: Determine the number of application instances needed

Start with the applications. Determine how many application instances, or pods, you plan to deploy. When sizing the environment, any application component deployed on Red Hat OpenShift—such as a database, front-end static server, or VM instance—is considered an application instance.

This figure can be an approximation to help calculate a gross estimate of your Red Hat OpenShift environment size. CPU, memory oversubscription, quotas and limits, and other features can be used to further refine this estimate.

Table 1: Application and instance sizing questions

Relevant questions	Example answers
<ul style="list-style-type: none"> How many application instances do you anticipate deploying in each Red Hat OpenShift environment? 	<ul style="list-style-type: none"> We have around 1,250 application instances in our development environment and around 250 application instances in production.
<ul style="list-style-type: none"> What type of applications are they (e.g., language, framework, database)? 	<ul style="list-style-type: none"> We mainly deploy Java, but have some Microsoft.NET Core and Ruby applications. We also use MySQL.
<ul style="list-style-type: none"> For VM workloads, what are the standard configurations of the VMs? Are they all custom or are they standardized by application, and what are their requirements? 	<ul style="list-style-type: none"> Our average VM requires 4 vCPU and 8 GB of RAM Our average container requires 2 vCPU and 2 GB of RAM

Step 2: Determine the total memory and total core needed

Once you have determined the requirements for 1 application instance and the number of application instances, it is trivial to determine the total amount of resources needed, both compute and memory.

At this point, you should determine a maximum utilization that you want your compute nodes to meet. This should make room for the management overhead of OpenShift (for more details, see documentation, but typically 1 core or 1 vCPU and <1 GB of read access memory [RAM] per compute node) and to allow for normal variation in the application before autoscaling is required. If you assume an aggressive utilization (greater than 80%) you may need to explicitly take the OpenShift overhead requirements into account in your calculation of memory and core resources.

For virtualization use cases, you need to consider CPU and memory overcommitment, redundancy and high-availability concerns, and the overall architecture of the environment. We will go through an illustration in sizing [Example 3](#).

Table 2: Preferred maximum OpenShift node utilization questions

Relevant questions	Example answers
<ul style="list-style-type: none"> How much space do I want to reserve for increased demand? 	<ul style="list-style-type: none"> We want to run nodes at a maximum average of 80% of total capacity (20% in reserve).

Maximum utilization = % chosen by the architect

Total cores (or vCPU) required = "Cores for 1 application" X "number of application instances" X 1 / "percent utilization"

Total memory required = "Memory for 1 application" X "number of application instances" X 1 / "percent utilization"

Step 3: Choose a standard compute node (VM, cloud instance or bare-metal server)

Now you have a total for cores and memory you want to accommodate. You must now choose a standard compute node to deploy the worker nodes for your applications.

In a virtualization infrastructure, your environment may allow you to choose the configuration of the VMs that will serve as your compute node, or you may need to pick 1 of several standard "t-shirt size" VM types.

In a hyperscaler cloud, you will generally need to choose a cloud instance type with different compute, memory and access to optional equipment like AI accelerators.

If you are deploying on bare metal, either on premise or in a hyperscaler bare metal instance, you may have a standard server configuration, and sometimes you can specify the configuration.

Where possible, you should choose compute nodes that best match the ratio of your core requirements and your memory requirements. For example, if my total requirement for my containers is 400 vCPU and 1600 GB of RAM, then I will get the best average consolidation ratios by choosing compute nodes that have around 1:4 ratio of vCPU to memory.

Table 3: VM and hardware sizing questions

Relevant questions	Example answers
<ul style="list-style-type: none"> What is the memory capacity of the VMs you will use for nodes? 	<ul style="list-style-type: none"> Our VMs have 64 GB of memory and 4 vCPUs and hyperthreading is used.
<ul style="list-style-type: none"> What is the number of vCPUs for the VMs you will use for nodes? 	<ul style="list-style-type: none"> In our datacenter, we can customize the VMs for our container workloads.
<ul style="list-style-type: none"> Is hyperthreading in use? 	<ul style="list-style-type: none"> We use Amazon EC2 and have access to M6i and M7i instance types.
<ul style="list-style-type: none"> What number of AI accelerators are in use? 	<ul style="list-style-type: none"> Our bare-metal systems have 128 GB of RAM, 2 CPU sockets (64 cores), hyperthreading is used, and a single GPU is present.

Step 4: Calculate total subscriptions required

Finally, determine the number of Red Hat OpenShift subscriptions needed based on the data gathered in steps 1–3. In the case of subscriptions on VMs or hyperscaler cloud instances, you must use the core-pair subscription model. In the case of bare-metal servers, you should calculate the subscription using both models, compare the cost and flexibility differences of each model, and make the choice that is best suited for your needs.

For the core-pair subscription model

- ▶ Number of self-managed OpenShift core-pair subscriptions¹
 - ▶ = Total cores / 2, rounded up **or**
 - ▶ = Total vCPUs / 4, rounded up

For the bare-metal node subscription model

- ▶ First, calculate the number of core-pair subscriptions needed for your application, since you can use the core-pair model in a bare metal installation
- ▶ Second, determine the number of bare-metal subscriptions: Each physical server will require a bare-metal node subscription (you need 1 subscription per node).
- ▶ Third, compare the cost and flexibility differences between the 2 models.
 - ▶ Financially:
 - ▶ One subscription model will probably be less expensive than the other for your sized environment.
 - ▶ In planning for the future, there may be a break-even capacity point beyond which makes the other model a better choice financially.
 - ▶ Architecturally:
 - ▶ Core-pair subscriptions can be used anywhere in your environment (VMs, cloud instances, bare-metal servers), where bare-metal node subscriptions can only be used on bare-metal servers.
 - ▶ Core-pair subscriptions on bare-metal servers must run containers on bare metal and must be in the same single cluster.
 - ▶ You can choose to install OpenShift Virtualization Engine on the bare-metal servers and then core-pair OpenShift subscriptions on top for containers (OpenShift-on-OpenShift). This is best suited for a mixed VM environment with a relatively small amount of OpenShift per server.
 - ▶ In the case where you need a high density of Red Hat OpenShift workloads on a bare-metal server, choosing the bare-metal node subscription gives you unlimited OpenShift containers either directly on bare metal, or, using the OpenShift Virtualization feature, inside the VMs running on the server.

¹ If hyperthreading is in use, 2 virtual cores count only as 1 core of a subscription. See the section on Cores versus vCPUs and hyperthreading for details on whether to use effective or actual cores in this calculation.

Step 5: Calculate the total AI Accelerator subscriptions (if applicable)

Add all of the AI Accelerator subscriptions above. You need one AI Accelerator subscription per physical accelerator on premise or on a bare-metal server. On a hyperscaler cloud environment, the instance types for accelerated compute will list the number of physical GPUs or accelerators included in that instance type. Remember to match the SLA of the AI Accelerator subscription to the corresponding Red Hat OpenShift subscription.

Note: Red Hat OpenShift supports many features and functions which affect scalability, pod scheduling, idling, and resource quota/limiting features. The previous calculations are guidelines, and you may be able to tune your actual environment for better resource use or smaller total environment size. OpenShift Platform Plus customers should consider the needs of the additional software applications (Red Hat Advanced Cluster Management, Red Hat Advanced Cluster Security, and Quay), including storage and compute resources, even though they may not require additional compute subscriptions.

If working with a third-party reseller, refer to their specific terms and agreements for Red Hat products and services.

Example 1: Containerized application running in a hyperscaler cloud

We have an application consisting of 200 container instances, each of which consumes an average of 1 vCPU and 4 GB of RAM. Our preferred maximum utilization is 80%. We are running the application on AWS and have access to EC2 M6i instance types. Our application does not require any specific hardware or AI accelerators. We have chosen Red Hat OpenShift Platform Plus as our self-managed edition, and at the moment we just need an estimate of the number of subscriptions we require.

Step 1: Identify number and type of application instances needed

From the information in our example:

- ▶ Number of application instances = **200**
- ▶ Preferred maximum node utilization = **80%**
- ▶ Average application vCPU requirement = **1 vCPU**
- ▶ Average application memory footprint = **4 GB**

Step 2: Determine the total memory and total cores needed

Using the numbers from step 1 in our calculations:

- ▶ Maximum utilization = **80%**
- ▶ Total vCPU required = $1 \text{ vCPU} \times 200 \times 1/80\% = \mathbf{250 \text{ vCPU}}$
- ▶ Total memory required = $4 \text{ GB} \times 200 \times 1/80\% = \mathbf{1000 \text{ GB}}$

Step 3: Choose a standard compute node

From the information provided, we do not need an instance type with GPU or specialized hardware. Our ratio of vCPU:Memory is 250vCPU:1000 GB, or 1:4. Fortunately, there are plenty of instance types with that ratio. After analysis of the other factors our application requires, we determined that the **m6i.4xlarge** instance type is the best suited for our needs. Each instance has **16 vCPU and 64 GB** of memory.

Step 4: Calculate total subscriptions required

In this case, we know we need core-pair subscriptions since we are running in a hyperscaler cloud, and our hyperscaler cloud is using vCPU, so we use the core-pair subscription formula in step 2.

Total core-pair subscriptions = Total vCPU required / 4, rounded up

250 vCPU / 4 = 62.5, or 63 rounded up

Step 5: Calculate the total AI Accelerator subscriptions (if applicable)

For this sizing example, we are not using AI Accelerators so this does not apply.

Result

For this example 1, **63 OpenShift Platform Plus core-pair subscriptions** would be needed.

Example 2: Containerized application running on premise on bare metal

We now want to run this same application on premise and on bare metal, provisioned into VM compute nodes using the included OpenShift Virtualization feature of Red Hat OpenShift. The application consists of 200 container instances, each of which consumes 1 vCPU and 4 GB of RAM. Our preferred maximum utilization is 80%. Our application does not require any specific hardware or AI accelerators. We have chosen Red Hat OpenShift Platform Plus as our self-managed edition, and at this time we just need an estimate of the number of subscriptions we require. We have some flexibility in the bare-metal configuration we choose, but we are using standard off-the-shelf server configurations.

Step 1: Identify number and type of application instances needed

From the information in our example:

- ▶ Number of application instances = **200**
- ▶ Preferred maximum node utilization = **80%**
- ▶ Average application vCPU requirement = **1 vCPU**
- ▶ Average application memory footprint = **4 GB**

Step 2: Determine the total memory and total cores needed

Using the numbers from step 1 in our calculations:

- ▶ Maximum utilization = **80%**
- ▶ Total vCPU required = 1 vCPU X 200 X 1/80% = **250 vCPU**
- ▶ Total memory required = 4 GB X 200 X 1/80% = **1000 GB**

Step 3: Choose a standard compute node

Our current bare-metal server standard is a 2 socket server with 32 cores/64 vCPU per socket. We have our choice of RAM. Since vCPU to RAM ratio is 1:4, we will stock our servers with 256 GB of RAM. Therefore our bare-metal server choice is a 2-socket server with 32 cores/64 vCPU per socket (64 cores/128 vCPU per server) and 256 GB RAM.

Step 4: Calculate total subscriptions required

On a bare-metal server, we have our choice of core-pair subscriptions or bare-metal node subscriptions. So let's calculate both:

Total core-pair subscriptions = Total vCPU required / 4, rounded up

250 vCPU / 4 = 62.5, or 63 rounded up

Total node subscriptions = Total number of physical servers

▶ **Total number of servers needed** = 250 vCPU / 128 vCPU per server, rounded up = 2 servers

▶ 2 servers x 1 subscription/server = **2 bare-metal node subscriptions**

In this case, since we were able to select a server to give us the right vCPU to memory ratio, we do not need to calculate the number of servers required to meet memory requirements and take the larger of the 2. If the server we chose had a different amount of CPU cores or RAM, we would need to take that difference into account.

Step 5: Calculate the total AI Accelerator subscriptions (if applicable)

For this sizing example, we are not using AI Accelerators so this does not apply.

Result

For this example, **63 OpenShift Platform Plus core-pair subscriptions** or **2 bare-metal node subscriptions** would be needed. The decision at this point would come down to the best financial and architectural decision.

Example 3: VM-only environment

We are moving our virtual machines from another hypervisor to Red Hat. This is a mixed environment, but we have identified the following number of different sizes of virtual machines:

- ▶ 1000 small = 1000 vCPU, 4000GiB. Plus, 228 GiB memory overhead
- ▶ 300 medium = 600 vCPU, 2400GiB. Plus, 73 GiB memory overhead
- ▶ 200 large = 800 vCPU, 4800GiB. Plus, 58 GiB memory overhead
- ▶ 200 x-large = 1600 vCPU, 9600GiB. Plus, 64 GiB memory overhead

Step 1: Identify number and type of application instances needed

From the information in our example:

- ▶ Number of application instances = **1700**
- ▶ Total vCPU = **4000 vCPUs**
- ▶ Total memory = 20800 GB + 423 GB overhead = **21223 GB**

- ▶ Average application vCPU requirement = **2.4 vCPU**
- ▶ Average application memory footprint = **12.5 GB**

Step 2: Determine the total memory and total cores needed

Using the numbers from step 1 in our calculations:

- ▶ Total memory required = 20800 GB + 423 GB overhead = **21223 GB**
- ▶ Total vCPU required = **4000 vCPUs**

Now the maximum utilization metric for VMs is a slightly different exercise than for containers, but should be familiar to virtualization administrators.

In general, we do not recommend memory overcommitment for VMs, so total memory requirements generally become the leading factor for the number of compute nodes.

For compute resources, we expect CPU overcommitment since on average most VMs are not using all their compute resources. The maximum CPU overcommitment ratio for OpenShift Virtualization is set at 10:1, so choosing a ratio like 4:1 is conservative. Here is where we can also choose if we want to use cores or threads (with hyperthreading support each core can be 2 threads) to equal a vCPU. We can be conservative and choose 1 vCPU = 1 core and no hyperthreading. Now, our requirements are:

- ▶ Total memory required = 20,800 GB + 423 GB overhead = **21,223 GB**
- ▶ Total cores required = 4000 vCPUs x 1/4 x 1 core/vCPU = **1000 cores**

Step 3: Choose a standard compute node

Choosing a bare-metal compute node for virtualization depends on many things, including redundancy and failure domains, size of clusters, etc. On premise we have some choices, including increasing RAM per server. Since memory generally guides the server requirements, we can start there.

We have 1700 VMs and we are choosing a 2 socket server with 32 cores per socket. If we use our core count to fit the number of servers, we will need:

- ▶ 1000 cores / 64 cores/server, rounded up = 16 servers

With 16 servers, we will need 21,223 GB / 16 servers = 1,326 GB ram per server. With our server, we can choose 1,536 GB of RAM. Now, our bare-metal configuration is:

- ▶ 2 socket server with 32cores/socket (64 cores total), 1,536 GB RAM

Finally, 16 servers with this configuration yields a total of:

- ▶ 16 x 64 cores = 1024 cores
- ▶ 16 x 1536 GB = 24576 GB memory

This is sufficient to run the VM loads, but we will require additional servers for redundancy. Right now we cannot afford the loss of any one server with an outage or severe performance impacts. Our virtualization administrators advise us to have 25% spare capacity for failover and redundancy. Therefore we will need:

- ▶ 16 servers + (16 servers * 25%) = 20 servers total

We will spread out VMs across all 20 servers so that even if we lose 1–4 servers, we will still be able to meet our VM's requirements. (Your resiliency requirements may differ.)

Step 4: Calculate total subscriptions required

For a virtualization-only use case, we are using OpenShift Virtualization Engine, which is only available in bare-metal node subscriptions.

Total node subscriptions =

- ▶ **Total number of servers needed** = 20
- ▶ 20 servers x 1 subscription/server = **20 bare-metal node subscriptions**

Step 5: Calculate the total AI Accelerator subscriptions (if applicable)

For this sizing example, we are not using AI Accelerators so this does not apply.

Result

For this example, **20 OpenShift Virtualization Engine bare-metal node subscriptions** would be needed.

Appendix 1: Self-managed OpenShift editions and what they include

Table 1: Overall feature differences by Red Hat OpenShift edition

Feature	Red Hat OpenShift Virtualization Engine	Red Hat OpenShift Kubernetes Engine	Red Hat OpenShift Container Platform	Red Hat OpenShift Platform Plus
Auth integrations, role-based access control (RBAC), security context constraints (SCC), Multi-tenancy admission controller	Yes	Yes	Yes	Yes
Autoscaling (cluster and pod)	Yes	Yes	Yes	Yes
cluster monitoring	Yes	Yes	Yes	Yes
cost management SaaS service	Yes	Yes	Yes	Yes
Kubernetes Container Runtime Interface (CRI) for Open Container Initiative (OCI)-compatible runtimes, or CRI-O runtime	Yes	Yes	Yes	Yes
Enterprise Secured Kubernetes	Yes	Yes	Yes	Yes
fully automated installers	Yes	Yes	Yes	Yes
kubectl and oc automated command line	Yes	Yes	Yes	Yes
OpenShift Virtualization	Yes	Yes	Yes	Yes
Operator Lifecycle Manager (OLM)	Yes	Yes	Yes	Yes
Over the air smart upgrades	Yes	Yes	Yes	Yes
secrets management	Yes	Yes	Yes	Yes
storage drivers	Yes	Yes	Yes	Yes
user-provided virtual machine workloads	Yes	Yes	Yes	Yes
Red Hat OpenShift Lightspeed	Yes	Yes	Yes	Yes
web console	Yes	Yes	Yes	Yes
Red Hat OpenShift GitOps	For VM use cases only	For VM use cases only	Yes	Yes
Platform Logging	For VM use cases only	For VM use cases only	Yes	Yes
User Workload Monitoring	For VM use cases only	For VM use cases only	Yes	Yes
Red Hat Enterprise Linux support and entitlement for worker and infrastructure nodes		Yes	Yes	Yes
Entitlement for Red Hat Enterprise Linux VM guests operating system and container builds		Yes	Yes	Yes
user-provided container workloads		Yes	Yes	Yes

Feature	Red Hat OpenShift Virtualization Engine	Red Hat OpenShift Kubernetes Engine	Red Hat OpenShift Container Platform	Red Hat OpenShift Platform Plus
Builds for Red Hat OpenShift			Yes	Yes
developer application catalog			Yes	Yes
Embedded Component of IBM Cloud Pak and Red Hat Application Services bundles			Yes	Yes
integrated development environments (IDEs)			Yes	Yes
distributed tracing			Yes	Yes
odo			Yes	Yes
Red Hat OpenShift Pipelines			Yes	Yes
Red Hat OpenShift sandboxed containers			Yes	Yes
Red Hat OpenShift Serverless			Yes	Yes
Red Hat OpenShift Service Mesh			Yes	Yes
Argo CD Agent				Yes
Red Hat Advanced Cluster Management for Kubernetes				Yes
Red Hat Advanced Cluster Security for Kubernetes				Yes
Red Hat OpenShift Data Foundation Essentials				Yes
Red Hat Quay				Yes

Table 2: Detailed differences between Red Hat OpenShift editions, including operators providing those features

Feature	Red Hat OpenShift Virtualization Engine	Red Hat OpenShift Kubernetes Engine	Red Hat OpenShift Container Platform	Red Hat OpenShift Platform Plus	Operator name
AWS EFS CSI Driver Operator	Yes	Yes	Yes	Yes	aws-efs-csi-driver-operator
AWS Load Balancer Operator	Yes	Yes	Yes	Yes	aws-load-balancer-operator
Builds for Red Hat OpenShift			Yes	Yes	openshift-builds-operator
cert-manager operator for Red Hat OpenShift	Yes	Yes	Yes	Yes	openshift-cert-manager-operator
cluster monitoring	Yes	Yes	Yes	Yes	Cluster Monitoring
cluster observability operator	Yes	Yes	Yes	Yes	cluster-observability-operator

Feature	Red Hat OpenShift Virtualization Engine	Red Hat OpenShift Kubernetes Engine	Red Hat OpenShift Container Platform	Red Hat OpenShift Platform Plus	Operator name
ClusterResourceOverride Operator	Yes	Yes	Yes	Yes	clusterresourceoverride
CNI plugin ISV Compatibility	Yes	Yes	Yes	Yes	N/A
compliance operator	Yes	Yes	Yes	Yes	Compliance Operator
cost management	Yes	Yes	Yes	Yes	costmanagement-metrics-operator
CSI plugin ISV Compatibility	Yes	Yes	Yes	Yes	N/A
custom metrics autoscaler	Yes	Yes	Yes	Yes	openshift-custom-metrics-autoscaler-operator
developer application catalog			Yes	Yes	N/A
device manager (for example, GPU)	Yes	Yes	Yes	Yes	N/A
distributed tracing			Yes	Yes	tempo-operator
DPU network operator	Yes	Yes	Yes	Yes	dpu-network-operator
Egress Pod and Namespace Granular Control	Yes	Yes	Yes	Yes	N/A
Embedded Marketplace	Yes	Yes	Yes	Yes	N/A
Embedded OperatorHub	Yes	Yes	Yes	Yes	N/A
Embedded Registry	Yes	Yes	Yes	Yes	N/A
external DNS operator	Yes	Yes	Yes	Yes	external-dns-operator
fence agents remediation	Yes	Yes	Yes	Yes	fence-agents-remediation
file integrity operator	Yes	Yes	Yes	Yes	File Integrity Operator
Gateway API	Yes	Yes	Yes	Yes	N/A
GCP FileStore CSI Driver Operator	Yes	Yes	Yes	Yes	gcp-filestore-csi-driver-operator
HAProxy Ingress Controller	Yes	Yes	Yes	Yes	N/A
Helm	Yes	Yes	Yes	Yes	N/A
Ingress Cluster-wide Firewall	Yes	Yes	Yes	Yes	N/A
Ingress Non-Standard Ports	Yes	Yes	Yes	Yes	N/A

Feature	Red Hat OpenShift Virtualization Engine	Red Hat OpenShift Kubernetes Engine	Red Hat OpenShift Container Platform	Red Hat OpenShift Platform Plus	Operator name
IPv6 Single and Dual Stack	Yes	Yes	Yes	Yes	N/A
Kiali			Yes	Yes	Kiali Operator
Kourier Ingress Controller			Yes	Yes	OpenShift Serverless
Kube Descheduler Operator provided by Red Hat	Yes	Yes	Yes	Yes	Kube Descheduler Operator
Kubernetes NMState Operator	Yes	Yes	Yes	Yes	kubernetes-nmstate-operator
Leader Worker Set	No	No	Yes	Yes	Leader Worker Set
local storage operator	Yes	Yes	Yes	Yes	Local Storage Operator
Log Forwarding	Yes	Yes	Yes	Yes	Red Hat OpenShift Logging Operator
logging for Red Hat OpenShift			Yes	Yes	cluster-logging
logical volume manager storage	Yes	Yes	Yes	Yes	lvms-operator
Loki Operator	For VM use cases only	For VM use cases only	Yes	Yes	loki-operator
machine deletion remediation	Yes	Yes	Yes	Yes	machine-deletion-remediation
MetalLB Operator	Yes	Yes	Yes	Yes	metallb-operator
migration toolkit for applications			Yes	Yes	mta-operator
migration toolkit for containers	For VM use cases only	For VM use cases only	Yes	Yes	mtc-operator
migration toolkit for runtimes			Yes	Yes	mtr-operator
migration toolkit for virtualization	Yes	Yes	Yes	Yes	mtv-operator
multiarch tuning	Yes	Yes	Yes	Yes	multiarch-tuning-operator
Multus and Available Multus plugins	Yes	Yes	Yes	Yes	N/A
network bound disk encryption (nbde) tang server	Yes	Yes	Yes	Yes	nbde-tang-server

Feature	Red Hat OpenShift Virtualization Engine	Red Hat OpenShift Kubernetes Engine	Red Hat OpenShift Container Platform	Red Hat OpenShift Platform Plus	Operator name
Network observability operator	Yes	Yes	Yes	Yes	netobserv-operator
network policies	Yes	Yes	Yes	Yes	N/A
Node Feature Discovery provided by Red Hat	Yes	Yes	Yes	Yes	nfd
node health check	Yes	Yes	Yes	Yes	node-healthcheck-operator
node maintenance	Yes	Yes	Yes	Yes	node-maintenance-operator
NUMA Resources Operator	Yes	Yes	Yes	Yes	numaresources-operator
ODF Multicluster Orchestrator				Yes	odf-multicluster-orchestrator
OpenShift APIs for Data Protection (OADP)	Yes	Yes	Yes	Yes	OADP Operator
OpenShift Cloud Manager SaaS service	Yes	Yes	Yes	Yes	N/A
OpenShift DR Cluster Operator				Yes	odr-cluster-operator
OpenShift DR Hub Operator				Yes	odr-hub-operator
OpenShift Elasticsearch Operator (Note 1)			Yes	Yes	elasticsearch-operator
OpenShift Update Service	Yes	Yes	Yes	Yes	cincinnati-operator
OpenShift Virtualization	Yes	Yes	Yes	Yes	OpenShift Virtualization Operator
OVS and OVN SDN	Yes	Yes	Yes	Yes	N/A
Platform Logging	For VM use cases only	For VM use cases only	Yes	Yes	Red Hat OpenShift Logging Operator
power monitoring for Red Hat OpenShift	Yes	Yes	Yes	Yes	power-monitoring-operator
PTP Operator provided by Red Hat	Yes	Yes	Yes	Yes	PTP Operator
Quay Bridge Operator provided by Red Hat			Yes	Yes	Quay Bridge Operator

Feature	Red Hat OpenShift Virtualization Engine	Red Hat OpenShift Kubernetes Engine	Red Hat OpenShift Container Platform	Red Hat OpenShift Platform Plus	Operator name
Quay Container Security provided by Red Hat			Yes	Yes	Container Security Operator
Red Hat build of Keycloak			Yes	Yes	keycloak-operator
Red Hat build of Kueue		Yes	Yes	Yes	Red Hat Build of Kueue
Red Hat build of OpenTelemetry		Yes	Yes	Yes	opentelemetry-operatorclusterlet-product
Red Hat build of Quarkus			Yes	Yes	N/A
Red Hat build of Trustee			Yes	Yes	trustee-operator
Red Hat Enterprise Linux Software Collections and RHT SSO Common Service	Yes	Yes	Yes	Yes	N/A
Red Hat JBoss Web Server			Yes	Yes	N/A
Red Hat OpenShift Dev Spaces			Yes	Yes	devspaces
Argo CD Agent				Yes	argocd-agent
Red Hat OpenShift GitOps (excluding Argo CD Agent)	For VM use cases only	For VM use cases only	Yes	Yes	openshift-gitops-operator
Red Hat OpenShift Local			Yes	Yes	N/A
Red Hat OpenShift Pipelines			Yes	Yes	openshift-pipelines-operator-rh
Red Hat OpenShift sandboxed containers			Yes	Yes	sandboxed-containers-operator
Red Hat OpenShift Serverless			Yes	Yes	serverless-operator
Red Hat OpenShift Service Mesh			Yes	Yes	servicemeshoperator
Red Hat Quay compatibility	Yes	Yes	Yes	Yes	N/A
Run Once Duration Override Operator	Yes	Yes	Yes	Yes	run-once-duration-override-operator
secondary scheduler operator for Red Hat OpenShift	Yes	Yes	Yes	Yes	openshift-secondary-scheduler-operator
secret store CSI	Yes	Yes	Yes	Yes	Secrets Store CSI Operator

Feature	Red Hat OpenShift Virtualization Engine	Red Hat OpenShift Kubernetes Engine	Red Hat OpenShift Container Platform	Red Hat OpenShift Platform Plus	Operator name
security profile	Yes	Yes	Yes	Yes	Security Profiles Operator
Service Binding operator	Yes	Yes	Yes	Yes	rh-service-binding-operator
single sign-on			Yes	Yes	rhsso-operator
source to image and builder automation			Yes	Yes	OpenShift Pipelines
SR-IOV network operator	Yes	Yes	Yes	Yes	SR-IOV Network Operator
Telemeter and Insights Connected Experience	Yes	Yes	Yes	Yes	N/A
Tempo Operator	Yes	Yes	Yes	Yes	tempo-operator
topology aware lifecycle manager			Yes	Yes	topology-aware-lifecycle-manager
User Workload Monitoring	For VM use cases only	For VM use cases only	Yes	Yes	cluster-monitoring-operator
Vertical Pod Autoscaler	Yes	Yes	Yes	Yes	Vertical Pod Autoscaler
VolSync			Yes	Yes	volsync-product
Web Terminal	Yes	Yes	Yes	Yes	web-terminal
Web Terminal provided by Red Hat			Yes	Yes	Web Terminal
Windows Machine Config Operator		Yes	Yes	Yes	Windows Machine Config Operator

Note 1: The OpenShift Elasticsearch Operator included in Red Hat OpenShift is only licensed for supporting the internal infrastructure search needs of the OpenShift cluster and cannot be used standalone for customer applications.

Common Red Hat software not included in any Red Hat OpenShift editions

Unless otherwise specified, the following Red Hat software offerings (commonly used with Red Hat OpenShift) need to be entitled separately. Red Hat OpenShift Platform Plus includes Red Hat Advanced Cluster Management, Red Hat Advanced Cluster Security, Red Hat Quay, and Red Hat OpenShift Data Foundation Essentials. Other self-managed OpenShift subscriptions do not include those additional products, but they can be purchased separately. Other Red Hat software used with Red Hat OpenShift but not included in this subscription guide are:

- ▶ Red Hat Advanced Developer Suite
- ▶ Red Hat Ansible Automation Platform
- ▶ Red Hat Application Foundations
- ▶ Red Hat Enterprise Linux AI
- ▶ Red Hat Integration portfolio, including 3scale, AMQ, Camel K, Fuse, etc.
- ▶ Red Hat JBoss EAP
- ▶ Red Hat Middleware Bundles
- ▶ Red Hat OpenShift Data Foundation Advanced
- ▶ Red Hat OpenShift Developer Hub (Red Hat's build of the Backstage project)
- ▶ Red Hat OpenShift AI
- ▶ Red Hat Satellite (for management of Red Hat Enterprise Linux)
- ▶ IBM CloudPaks

Ask your Red Hat seller or partner for more information on the above offerings.



About Red Hat

Red Hat is the world's leading provider of enterprise open source software solutions, using a community-powered approach to deliver reliable and high-performing Linux, hybrid cloud, container, and Kubernetes technologies. Red Hat helps customers develop cloud-native applications, integrate existing and new IT applications, and automate and manage complex environments. [A trusted adviser to the Fortune 500](#), Red Hat provides [award-winning](#) support, training, and consulting services that bring the benefits of open innovation to any industry. Red Hat is a connective hub in a global network of enterprises, partners, and communities, helping organizations grow, transform, and prepare for the digital future.

f facebook.com/redhatinc
x @RedHat
in linkedin.com/company/red-hat

North America
1 888 REDHAT1

**Europe, Middle East,
and Africa**
00800 7334 2835
europe@redhat.com

Asia Pacific
+65 6490 4200
apac@redhat.com

Latin America
+54 11 4329 7300
info-latam@redhat.com