# Red Hat OpenShift Container Platform 4.14 on Dell Infrastructure

## Enabled by Intel-powered Dell PowerEdge servers, Dell PowerSwitch networking, and Dell storage

June 2024

H19965.1

Design Guide

Abstract

This guide provides architecture and design information for the Dell Validated Design for a deployment of Red Hat® OpenShift® Container Platform 4.14 on Intel-powered Dell PowerEdge servers, Dell PowerSwitch networking, and Dell storage systems.

Dell Technologies Solutions

Dell Validated Design

Dell Technologies

Copyright

# Contents

# Chapter 1 Introduction

This chapter presents the following topics:

# Solution introduction

## Overview

Red Hat® OpenShift® Container Platform 4.14 can host the development and run-time execution of containerized applications. OpenShift Container Platform is based on Kubernetes, the acknowledged automation and life cycle management (LCM) platform leader for containerized workloads and services.

The Dell Validated Design (DVD) for OpenShift Container Platform 4.14 is a flexible infrastructure solution that has been designed, optimized, and validated for an on-premises bare-metal OpenShift Container Platform 4.14 deployment. The DVD includes Dell servers, switches, and storage to enable you to develop, validate, and deploy your containerized applications on OpenShift Container Platform. The focus is on running cloud-native applications on containers that are managed by OpenShift Container Platform.

## Key benefits

The DVD for OpenShift Container Platform 4.14 provides:

- A proven design to help organizations accelerate their container deployments and cloud-native adoption.
- A selection of validated OpenShift Container Platform hardware designs
- A rapid implementation and reduced time to value (TTV)

# Document purpose

This guide provides information about building an on-premises infrastructure solution to host OpenShift Container Platform 4.14. It describes the design decisions and configurations that enable solution architects to:

- Design and deploy a container platform solution.
- Extend or modify the design as necessary to meet customer requirements.

The guide addresses the following topics:

- Container ecosystem design overview
- Container and application storage design guidance
- Server requirements to support OpenShift Container Platform node roles
- Network infrastructure design guidance
- Configuration recommendations to optimize latency and networking performance
- General guidance for remote compute nodes at the edge
- Hardware platform configuration recommendations
- Rack-level design and power configuration considerations

The guide also provides information about:

- Red Hat OpenShift Container Platform 4.14 for application development and deployment

- Dell PowerEdge R660 and R760 servers based on Intel® 4th Generation Xeon® processors.
- Dell PowerSwitch S5200-series switches for infrastructure network enablement
- Dell PowerSwitch S3100-series switch for out-of-band (OOB) management of the cluster

**Note:** The contents of this document are valid for the software and hardware versions that it describes. For information about updated configurations for newer software and hardware versions, contact your Dell Technologies sales representative.

## Revision history

**Table 1. Revision History**

| Date | Part number | Change summary |
|---|---|---|
| June 2024 | H19965.1 | Corrected error in PowerSwitch fconfiguration topic |
| March 2024 | H19965 | Initial release |

## Audience

This design guide is intended for system administrators and system architects. Some experience with container technologies, Kubernetes, and Red Hat OpenShift Container Platform is recommended.

## We value your feedback

Dell Technologies and the authors of this document welcome your feedback on the solution and the solution documentation. Contact the Dell Technologies Solutions team by email.

# Chapter 2 Technology and Deployment Process

This chapter presents the following topics:

## Overview

This chapter describes the components and requirements of a viable architecture for an OpenShift Container Platform 4.14 cluster that can drive the demands of enterprise workloads.

## Red Hat OpenShift Container Platform

OpenShift Container Platform is an enterprise-grade declarative state machine that has been designed to automate application workload operations based on the upstream Kubernetes project. In this context, “declarative” means that developers can specify, in code, a configuration for an application or workload without knowing how that application is going to be deployed. OpenShift Container Platform provides production-oriented container automation and workload automation. Based on Kubernetes version 1.27, which includes native support for cluster snapshots, OpenShift Container Platform 4.14 gives administrators and developers the tools that they need to deploy and manage applications and services at scale.

---

**Note:** OpenShift Container Platform is a certified Kubernetes distribution. Certification for Kubernetes distributions is provided by the Cloud Native Computing Foundation.

---

## Server options

Dell Technologies has tested and fully validated the following PowerEdge server options for an OpenShift Container Platform 4.14 deployment.

### PowerEdge R660

The PowerEdge R660 server is powered by the 4th Generation Intel® Xeon® Scalable processors with up to 56 cores for faster and more accurate processing performance. The PowerEdge R660 is a two-socket/1U rack server that delivers outstanding performance for the most demanding workloads. The R660 model supports up to 32 DDR5 RDIMMS up to 4400 MT per second (2DPC) or 4800 MT per second for 1DPC (16 DDR5 RDIMMs at a maximum). The PowerEdge R660 model provides storage options up to eight 2.5” NVMe/SAS4/SATA, plus up to 10 x 2.5” NVMe/SAS4/SATA, and 14/16 x 5th Generation NVMe E3.S. It provides a New Smart Flow chassis that optimizes airflow to support the highest core count CPUs in an air-cooled environment within the current IT infrastructure. The R660 offers multiple 4th Generation and 5th Generation riser configurations (up to three PCIe slots) with interchangeable components that seamlessly integrate to address customer needs over time. These features make the R660 an ideal server for data center standardization on a wide range of workloads.

### PowerEdge R760

The PowerEdge R760 server is powered by the 4th Generation Intel® Xeon® Scalable processors with up to 56 cores for a faster and more accurate processing performance. The R760 is a two-socket/2U rack server that delivers peak performance for the most demanding workloads. The R760 model supports up to 32 DDR5 RDIMMS up to 4400 MT/sec (2DPC) or 4800 MT/sec for 1DPC (16 DDR5 RDIMMs at a maximum). The R760 model provides storage options up to 12 x 3.5” SAS3/SATA; or up to 24 x 2.5” SAS4/SATA, plus up to 24 x NVMe U.2 Gen4, 16 x NVMe E3.S Gen5. It offers multiple 4th Generation and 5th Generation riser configurations (up to eight PCIe slots) with interchangeable components that seamlessly integrate to address customer needs over time. The PowerEdge R760 is therefore an ideal server for data center standardization on a wide range of workloads requiring performance, extensive storage, and GPU support.

## Validated firmware

The Dell OpenShift team used R660 and R760 server components to validate an OpenShift Container Platform 4.14 deployment. The following table shows the versions that they used:

**Table 2. Validated R660 and R760 firmware for the DVD**

| Product | Version |
|---|---|
| BIOS | 2.0.0 |
| iDRAC with Lifecycle Controller | 7.10.05.00 |
| NVIDIA Connect X-6 Lx | 26.38.10.02 |
| Emulex LPe35002 | 03.07.23 |
| HBA 355i | 24.15.14.00 |
| BOSS-N1 Monolithic | 2.1.13.2025 |

## Deployment process

To deploy a Red Hat OpenShift cluster, use one of the following methods:

- **Full control:** User Provisioned Infrastructure (UPI)-based deployment offers maximum flexibility and enables you to provision, configure, and manage the required infrastructure and resources (bootstrap machine, networking, load balancer, DNS, DHCP, and so on) for the cluster. A Cluster System Admin Host (CSAH) facilitates the cluster deployment and hosts the required services. For more information, see Installing a user-provisioned cluster on bare metal.
- **Automated:** Installer Provisioned Infrastructure (IPI)-based deployment is an automated approach to OpenShift cluster deployment. The installer uses the baseboard management controller (BMC) on each cluster host for provisioning. A CSAH node or a provisioner node is required to initiate the cluster deployment. For more information, see Deploying installer-provisioned clusters on bare metal.
- **Interactive:** This method uses the Assisted Installer from the Red Hat Hybrid Cloud Console to enter configurable parameters and generate a discovery ISO. (Access to the console requires Red Hat login credentials.) The servers are booted using this ISO to install Red Hat Enterprise Linux CoreOS (RHCOS) and an agent. Preinstallation validation and installation are provided by the Assisted Installer. For more information, see Assisted Installer for OpenShift Container Platform.
- **Local agent-based:** This method uses the agent-based installer, which is available for download from the Red Hat Hybrid Cloud Console, for clusters that are confined in an air-gapped network. (Access to the console requires Red Hat login

credentials.) For more information, see [Installing an OpenShift Container Platform cluster with the Agent-based Installer].

The deployment process for OpenShift nodes varies depending on the cluster topology, as described in the following section.

## OpenShift cluster topologies

OpenShift Container Platform 4.14 offers different topologies for different workload requirements, with different levels of server hardware footprints and high availability (HA).

- **Single node cluster:** Single Node OpenShift (SNO) bundles control-plane and data-plane capabilities into a single server and provides users with a consistent experience across the sites where OpenShift is deployed. For more information, see Preparing to install on a single node.
- **Compact cluster**: A compact cluster consists of three nodes hosting both the control plane and the data plane, allowing small footprint deployments of OpenShift for testing, development, and production environments. An initial expansion requires the addition of at least two compute nodes simultaneously. This is because the ingress networking controller requires two router pods on compute nodes for full functionality. You can add compute nodes later as necessary. For more information, see Installing a user-provisioned cluster on bare metal.
- **Standard HA cluster:** A standard cluster or the remote compute node cluster consists of five or more nodes. The control plane is hosted on three dedicated nodes, separating it out from the compute nodes. At least two nodes are required for hosting the data plane. This topology offers the highest level of HA.

## Deployment prerequisites

Ensure that:

- The network switches and servers are provisioned.
- Network cabling is complete.
- Internet connectivity is available to the cluster.

## CSAH node

The deployment begins with initial switch provisioning. The provisioning enables the preparation and installation of the CSAH node by:

- Installing Red Hat Enterprise Linux 9
- Subscribing to the necessary repositories
- Creating an Ansible user account
- Cloning a GitHub Ansible playbook repository from the Dell ESG container repository
- Running an Ansible playbook to start the installation process.

Dell Technologies has generated Ansible playbooks that fully prepare both CSAH nodes. See User-provisioned infrastructure installation. For a SNO deployment, the Ansible playbook sets up a DHCP server and a DNS server. The CSAH node is also used as an admin host to perform operations and management tasks on SNO.

---

**Note:** For enterprise sites, consider deploying appropriately hardened DHCP and DNS servers and using resilient multiple-node HAProxy configuration. The Ansible playbook for this design can deploy multiple CSAH nodes for resilient HAProxy configuration. This guide provides CSAH Ansible playbooks for reference at the implementation stage.

## OpenShift cluster bootstrapping process

The Ansible playbook creates a YAML file called `install-config.yaml` to control deployment of the ***bootstrap*** node.

The following figure shows the installation workflow. An ignition configuration control file starts the bootstrap node.

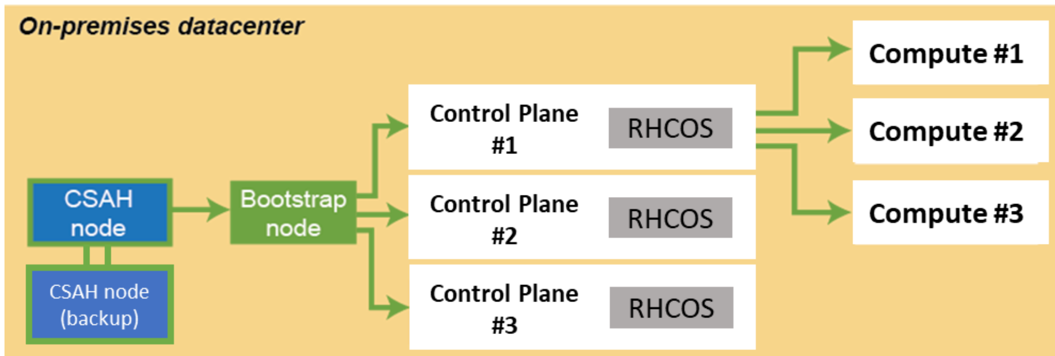


**Figure 1. Creating the bootstrap, control-plane nodes, and compute nodes**

**Note**: An installation that is driven by ignition configuration generates security certificates that expire after 24 hours. You must install the cluster before the certificates expire, and the cluster must operate in a viable (nondegraded) state so that the first certificate rotation can be completed.

The cluster bootstrapping process consists of the following phases:

1. After startup, the bootstrap virtual machine (VM) creates the resources that are required to start the control-plane nodes. **Interrupting this process can prevent the OpenShift control plane from coming up.**
2. The bootstrap machine starts a single-node `etcd` cluster and a temporary Kubernetes control plane.
3. The control-plane nodes pull resource information from the bootstrap VM to bring them up to a viable state.
4. The temporary control plane schedules the production control plane to the actual control-plane nodes.
5. The Cluster Version Operator (CVO) comes online and installs the `etcd` operator. The `etcd` operator scales up `etcd` on all the control-plane nodes.
6. The temporary control plane is shut down, handing control over to the now viable control-plane nodes.
7. OpenShift Container Platform components are pulled into the control of the control-plane nodes.
8. The bootstrap VM is shut down.
9. The control-plane nodes now drive creation and instantiation of the compute nodes.

10. The control plane adds operator-based services to complete the deployment of the OpenShift Container Platform ecosystem.

The cluster is now viable and can be placed into service in readiness for Day-2 operations. You can expand the cluster by adding more compute nodes for your requirements.

### User-provisioned infrastructure installation

Dell Technologies has generated Ansible playbooks that fully prepare both CSAH nodes. Before the installation of the OpenShift Container Platform 4.14 cluster begins, the Ansible playbook sets up the PXE server, DHCP server, DNS server, HAProxy, and HTTP server. If a second CSAH node is deployed, the playbook also sets up DNS, HAProxy, HTTP, and KeepAlived services on that node. The playbook creates ignition files to drive installation of the bootstrap, control-plane, and compute nodes, and also starts the bootstrap VM to initialize control-plane components. The playbook presents a list of node types that must be deployed in top-down order.

### Installer-provisioned infrastructure installation

The installer-provisioned infrastructure (IPI) installation on bare metal nodes provisions and configures the infrastructure on which an OpenShift cluster runs. OpenShift Container Platform manages all aspects of the cluster.

The CSAH node is used as the provisioner node and hosts infrastructure services such as DNS and an optional DHCP server. The bootstrap VM is hosted on the CSAH node for cluster setup. Dell-created Ansible playbooks set up CSAH nodes and automate the predeployment tasks, including configuring the CSAH node, downloading OpenShift installer, installing OpenShift client, and creating the manifest files that are required by the installer.

Cluster deployment using IPI installation is a two-phase process.

- In the first phase, the bootstrap VM is created on the CSAH node, and the bootstrap process is started. Two virtual IP addresses are used to access the cluster: API and Ingress.
- In the second phase, cluster nodes are provisioned using the virtual media-based Integrated Dell Remote Access Controller (iDRAC). After the nodes are booted, the temporary control plane is moved to the appropriate cluster nodes along with the virtual IP addresses. The bootstrap VM is destroyed, and the remaining cluster provisioning tasks are carried out by the controllers.

### Assisted Installer installation

Using the Assisted Installer deployment method, you create a cluster configuration using the web-based UI or the RESTful API. A CSAH node accesses the cluster and hosts the DNS and DHCP server.

The interface prompts for required values and provides default values for the remaining parameters unless you change them. After you have entered all the required details, a bootable discovery ISO is generated that is used to boot the cluster nodes. Along with RHCOS, the bootable ISO also contains an agent that handles cluster provisioning. The bootstrapping process completes on one of the cluster nodes. A bootstrap node or VM is not required. After the nodes are discovered on the Assisted Installer console, you can select node role (control plane or compute), installation disk, and networking options. The Assisted Installer performs prechecks before starting the installation.

You can monitor the cluster installation status or download installation logs and `kubeadmin` user credentials from the Assisted Installer console.

## Agent-based installation

This method is recommended for clusters with an air-gapped or disconnected network. You must download and install the agent-based installer on the [CSAH node](#).

The bootable ISO contains the assisted discovery agent and the assisted service. The assisted service runs only one of the control-plane nodes, and the node eventually becomes the bootstrap host. The assisted service ensures that all the cluster hosts meet the requirements, and then triggers an OpenShift cluster deployment.

The following figure shows the cluster node installation workflow:

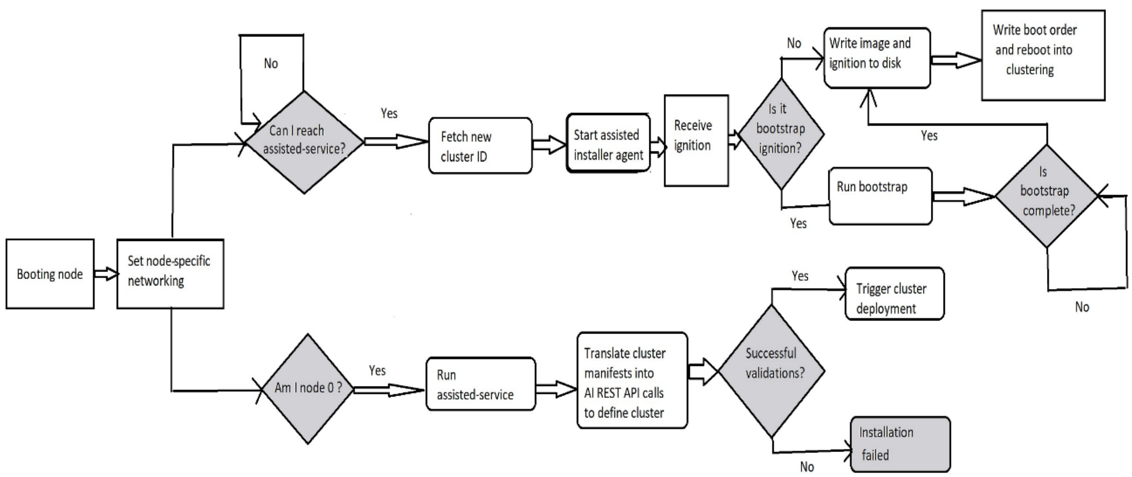


**Figure 1. Cluster node installation workflow**

### Disconnected installation mirroring

A mirror registry enables disconnected installations to host container images that are required for initial cluster deployment. The mirror registry must have access to the private network and to the internet to obtain the necessary container images. The options are:

- **Red Hat Quay**, an enterprise-quality container registry. Use Red Hat Quay to build and store container images, and then make them available to deploy across your enterprise. The Red Hat Quay operator provides a method to deploy and manage Red Hat Quay on an OpenShift cluster. For more information, see Quay for proof-of-concept purposes and Quay Operator.
- **Mirror registry for Red Hat OpenShift**, a small-scale container registry that is included with OpenShift Container Platform subscriptions. Use this registry if a large-scale registry is unavailable for initial cluster deployment. For more information, see Creating mirror registry for Red Hat OpenShift.

To use the mirror images for a disconnected installation, update the `install-config.yaml` file with the information in the `imageContentSources` section, as shown in the following example:

```
imageContentSources:
- mirrors:
  - ipicsah.dcws.lab:8443/ocp4/openshift414
  source: quay.io/openshift-release-dev/ocp-release
- mirrors:
  - ipicsah.dcws.lab:8443/ocp4/openshift414
  source: quay.io/openshift-release-dev/ocp-v4.0-art-dev
```

## Zero-Touch Provisioning

Zero-Touch Provisioning (ZTP) provisions new edge sites using declarative configurations. ZTP can deploy an OpenShift cluster quickly and reliably in any environment: edge, remote office/branch office (ROBO), disconnected, air-gapped. ZTP can deploy and deliver OpenShift 4.14 clusters in a hub-spoke architecture, where a unique hub cluster can manage multiple spoke clusters.

The following diagram shows how ZTP works in a far-edge environment:

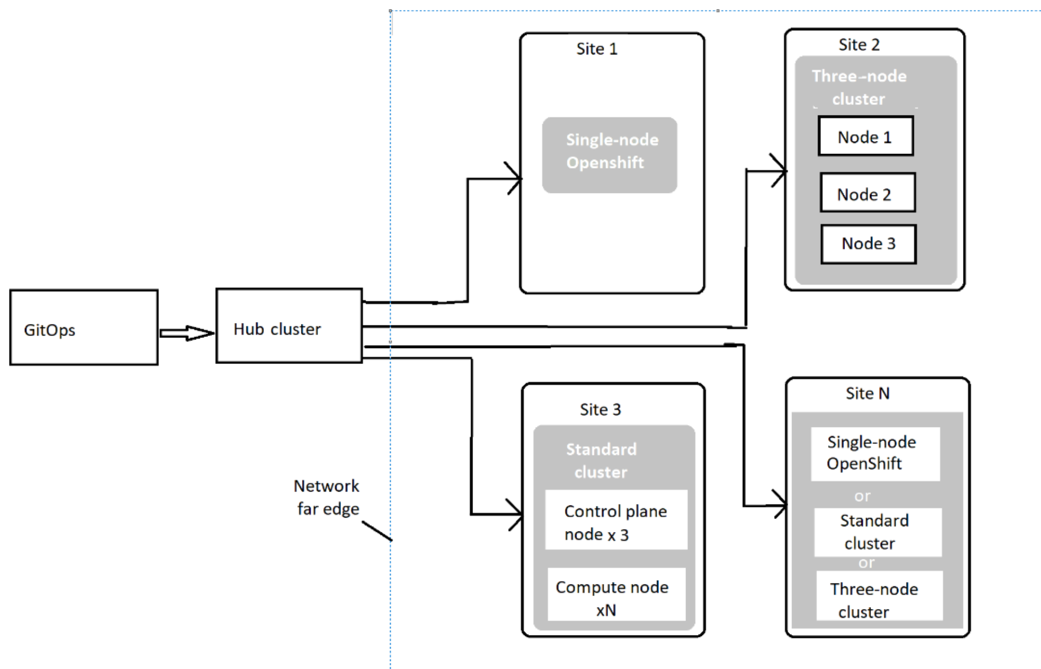


**Figure 2. OpenShift spoke cluster deployment using ZTP**

ZTP uses the GitOps deployment method for infrastructure deployment. Declarative specifications are stored in Git repositories in predefined patterns such as YAML. Red Hat Advanced Cluster Management (ACM) for Kubernetes uses the declarative output for multisite deployment. GitOps addresses reliability issues by providing traceability, role-based access controls (RBAC), and a single source of truth regarding the state of each site. SiteConfig uses ArgoCD as the engine for the GitOps method of site deployment. After completing a site plan that contains all the required parameters for deployment, a policy generator creates the manifests and applies them to the hub cluster.

For more information, see [Installing managed clusters with RHACM and SiteConfig resources](Installing managed clusters with RHACM and SiteConfig resources).

The following figure shows the ZTP flow that is used in a spoke-cluster deployment:

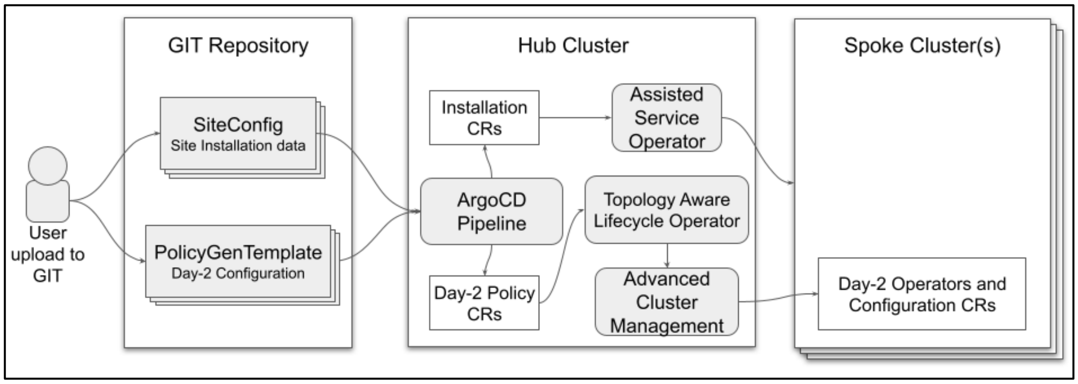


**Figure 3. ZTP deployment flow**

## OpenShift Virtualization

As an add-on to OpenShift Container Platform, OpenShift Virtualization enables you to run and manage VM workloads alongside container workloads. An enhanced web console provides a graphical portal to manage these virtualized resources alongside the OpenShift Container Platform cluster containers and infrastructure.

OpenShift Virtualization consists of the following components:

- Compute: virt-operator
- Storage: cdi-operator
- Network: cluster-network-addons-operator
- Scaling: ssp-operator
- Templating: tekton-tasks-operator

For more information, see OpenShift Virtualization architecture.

### Requirements

When planning the OpenShift Virtualization deployment, take into consideration that:

- Virtualization must be enabled in server processor settings.
- Control-plane nodes or the compute nodes must have RHCOS installed. Red Hat Enterprise Linux compute nodes are not supported.
- For HA to be enabled:
  - The OpenShift cluster must have at least two compute or data-plane nodes available at the time of cluster installation.
  - The live-migration feature requires the cluster-level HA flag to be set to true. This flag is set at the time of cluster installation and cannot be changed afterwards.
  - OpenShift Virtualization can be deployed on a single-node cluster, but HA is not available in this case.
- The live migration feature requires shared storage among the compute or data-plane nodes. For live migration to work, the underlying storage that is used for

OpenShift Virtualization must support and use ReadWriteMany (RWX) access mode.

- If Single Root I/O Virtualization (SR-IOV) will be used, the network interface controllers (NICs) must be supported.

For more information, see Preparing your cluster for OpenShift Virtualization.

As an add-on feature, OpenShift Virtualization imposes an overhead cost that must be accounted for during the planning phase. Oversubscribing the physical resources in a cluster can affect performance. See physical resource overhead requirements.

## Red Hat Hosted control planes

The introduction of hosted control planes for Red Hat OpenShift significantly simplifies the management of OpenShift clusters at scale, reducing both complexity and operational costs. This advancement enables you to create control planes as pods on a hosting cluster, eliminating the need for dedicated physical machines for each control plane.

### Requirements

To host control planes on OpenShift Container Platform version 4.14, you must:

- Set up a hub cluster on OpenShift Container Platform 4.14.
- Configure a default storage class.
- Install Red Hat Advanced Cluster Management and enable **Infrastructure management**.
- Configure DNS entries for the HyperShift cluster.
- Configure the host inventory and add hosts to the **Infrastructure** environment.
- After verifying that all hosts are available, go to **Infrastructure** and create the HyperShift cluster.

For more information, see Installation of Hosted control planes.

## Openshift AI

Red Hat OpenShift AI is a platform for data scientists and developers of artificial intelligence (AI) applications. OpenShift AI enables data scientists to analyze data and provides a cloud-based instance of Jupyter Notebook.

### Requirements

To use the OpenShift AI platform, you must:

- Configure a default storage class.
- Deploy the image registry.
- Install the OpenShift AI operator and its components.
- Add users to administrator groups to access the dashboard.
- Access the OpenShift AI dashboard.

For more information, see Overview of OpenShift AI.

# Infrastructure requirements

## Cluster infrastructure guidance

The following table provides basic cluster infrastructure guidance for validated hardware configurations. A container cluster can be deployed quickly and reliably when each node is within the validated design guidelines. For detailed configuration information, see Cluster Hardware Design.

**Table 3. Hardware configuration for OpenShift Container Platform 4.14 cluster deployment**

| Type | Description | Count | Notes |
|---|---|---|---|
| CSAH node | Dell PowerEdge R660 server | 2 | Creates a bootstrap VM.<br>CSAH runs a single instance of HAProxy for UPI-based deployment. For an enterprise HA deployment of OpenShift Container Platform 4.14, Dell Technologies recommends using a commercially supported L4 load-balancer or proxy service, or an additional PowerEdge R660 CSAH node running HAProxy and KeepAlived alongside the primary CSAH node. The options include commercial HAProxy, Nginx, and F5.<br>Used as a DNS server, DHCP server, and administration host. |
| Control-plane nodes | Dell PowerEdge R660 or R760 server | 3 | Deployed using the bootstrap VM. |
| Compute nodes | Dell PowerEdge R660 or R760 server | A minimum of 2* per rack. A maximum of 30 (R660) or 15 (R760) per rack | A standard deployment requires a minimum of two compute nodes. |
| Single-node OpenShift | Dell PowerEdge R660 or R760 server | 1 | Edge-optimized server options for deploying SNO. |
| Network switches | Either of the following switches:<br>• Dell PowerSwitch S5248F-ON<br>• Dell PowerSwitch S5232F-ON | 2 per rack | Configured at installation time.<br>**Note:**<br>• HA network configuration requires two data path switches per rack.<br>• Multirack clusters require network topology planning. Leaf-spine network switch configuration may be necessary. |
| OOB network | Dell PowerSwitch S3148 | 1 per rack | Used for iDRAC management. |
| Rack | Selected according to site standards | 1 to 7 racks | For multirack configurations, consult your Dell Technologies or Red Hat representative regarding custom engineering design. |

*A three-node cluster does not require any compute nodes. To expand a three-node cluster with additional compute machines, first expand the cluster to a five-node cluster using two additional compute nodes.

## Minimum viable solution requirements

Installing OpenShift Container Platform requires, at a minimum:

- One CSAH node, which is used to run the bootstrap VM. While the cluster is in production use, the CSAH node manages the cluster and performs load-balancing for the cluster. If resilient load-balancing is required, a second CSAH node can be used with the primary CSAH node to perform highly available load-balancing.

CSAH serves as an admin host and is required to access the cluster.

- Three nodes running both the control plane and the data plane.
- Single node running with the capabilities of both the control plane and the data plane but without HA, enabling customers to deploy OpenShift Container Platform 4.14 on the edge.

HA of the key services that make up the OpenShift Container Platform cluster is necessary to ensure run-time integrity. Redundancy of physical nodes for each cluster node type is an important aspect of HA for the bare-metal cluster. In this design guide, HA includes the provisioning of at least two NICs and two network switches that are configured to provide redundant pathing. The redundant pathing provides for network continuity if a NIC or a network switch fails.

# Chapter 3 Networking Infrastructure and Configuration

This chapter presents the following topics:

# Introduction

The components and operations that make up the container ecosystem each require network connectivity plus the ability to communicate with all other components and respond to incoming network requests. The DVD for OpenShift Container Platform 4.14 uses Dell PowerSwitch networking infrastructure.

# OpenShift network operations

## Operating components

Applications run on compute nodes. Each compute node is equipped with resources such as CPU cores, memory, storage, NICs, and add-in host adapters, including GPUs, SmartNICs, and FPGAs. Kubernetes provides a mechanism to enable orchestration of network resources through the Container Network Interface (CNI) API.

The CNI API uses the Multus CNI plug-in to enable attachment of multiple adapter interfaces on each pod. CRD objects are responsible for configuring Multus CNI plug-ins. For more information, see Multus CNI.

## Container communications

A pod is a basic unit of application deployment. Each pod consists of one or more containers that are deployed together on the same compute node. A pod shares the compute node network infrastructure with the other network resources that make up the cluster. As demand expands, additional identical pods are often deployed to the same or other compute nodes. For more information, see OpenShift Container Platform 4.14 Documentation.

Networking is critical to the operation of an OpenShift Container cluster. Four basic network communication flows occur within every cluster:

- Container-to-container connections (also known as “highly coupled communication”)
- Pod communication over the local host network (127.0.0.1)
- Pod-to-pod connections, as described in this guide
- Pod-to-service and ingress-to-service connections, which are handled by services

Containers that communicate within their pod use the local host network address. Containers that communicate with any external pod originate their traffic based on the IP address of the pod.

Application containers use shared storage volumes (configured as part of the pod resource) that are mounted as part of the shared storage for each pod. Network traffic that might be associated with nonlocal storage must be able to route across node network infrastructure. For more information, see Services, Load Balancing, and Networking.

## Services networking

Services are used to abstract access to Kubernetes pods. Every node in a Kubernetes cluster runs a kube-proxy and is responsible for implementing virtual IP (VIP) for services. Kubernetes supports two primary modes of finding (or resolving) a service:

- **Environment variables:** This method requires a reboot of the pods when the IP address of the service changes.
- **DNS:** OpenShift Container Platform 4.14 uses CoreDNS to resolve service IP addresses.

Some parts of the application (front ends, for example) might need to expose a service outside the application. If the service uses HTTP, HTTPS, or any other TLS-encrypted protocol, use an ingress controller; for other protocols, use a load balancer, external service IP address, or node port.

A node port exposes the service on a static port on the node IP address. A service with `NodePort-type` as a resource exposes the resource on a specific port on all nodes in the cluster. Ensure that external IP addresses are routed to the nodes. For more information, see Virtual IPs and Service Proxies.

## Ingress controller

OpenShift Container Platform uses an ingress controller to provide external access. The ingress controller defaults to running on two compute nodes, but it can be scaled up as needed. For more information, see Configuring ingress cluster traffic using an Ingress Controller.

An ingress controller accepts external HTTP, HTTPS, and TLS requests using SNI, and then proxies them based on the routes that are provisioned. Dell Technologies recommends creating a wildcard DNS entry and then setting up an ingress controller. This method enables you to work within the context of an ingress controller only.

You can expose a service by creating a route and using the cluster IP address. Cluster IP routes are created in the OpenShift Container Platform project, and a set of routes is admitted into ingress controllers.

By using sharding, the horizontal partitioning of data on route labels or name spaces, you can:

- Load-balance the incoming traffic.
- Hive off the required traffic to a single ingress controller.

For more information, see Ingress sharding in OpenShift Container Platform 4.14.

## Network operators

The following operators are available for network administration:

- **Cluster Network Operator (CNO):** Deploys the network provider plug-in that you selected during cluster installation. See Understanding the Cluster Network Operator.
- **DNS operator:** Deploys and manages CoreDNS and instructs pods to use the CoreDNS IP address for name resolution. See Understanding the DNS Operator.
- **Ingress operator:** Enables external access to OpenShift Cluster Platform cluster services and deploys and manages one or more HAProxy-based ingress controllers to handle routing. See Understanding the Ingress Operator.

## OpenShift SDN

OpenShift SDN creates an overlay network that is based on Open Virtual Switch (OVS). The overlay network enables communication between pods across the cluster. OVS operates in one of the following modes:

- **Network policy mode**: Allowing custom isolation policies (the default)
- **Multitenant mode**: Providing project-level isolation for pods and services
- **Subnet mode**: Providing a flat network

The Open Virtual Network (OVN)-Kubernetes as the CNI plug-in is a network provider for the default cluster network. OpenShift Container Platform 4.14 supports additional SDN orchestration and management for plug-ins that complies with the CNI specification. For more information, see About the OpenShift SDN network plugin.

## Multinetwork support

OpenShift Container Platform 4.14 also supports software-defined multiple networks. The platform comes with a default network. The cluster administrator uses the Multus CNI plug-in to define additional networks, and then chains the plug-ins. The additional networks are useful for increasing the networking capacity of the pods and meeting traffic separation requirements.

The following CNI plug-ins are available for creating additional networks:

- **Bridge:** The same host pods can communicate over a bridge-based additional network.
- **Host-device:** Pods can access the physical Ethernet network device of the host.
- **Macvlan:** Pods attached to a macvlan-based additional network have a unique MAC address and communicate using a physical network interface.
- **Ipvlan:** Pods communicate over an ipvlan-based additional network.
- **SR-IOV**: Pods attach to a virtual function (VF) interface on SR-IOV capable hardware on the host system.

For more information, see Attaching a pod to an additional network.

## Leaf-switch considerations

When pods are provisioned with additional network interfaces that are based on macvlan or ipvlan, corresponding leaf-switch ports must match the VLAN configuration of the host. Failure to properly configure the ports results in a loss of traffic.

For more information, see Updating node network configuration.

# Physical network design

## Design principles

OpenShift Container Platform 4.14 introduces advanced networking features to enable containers for high performance and monitoring. Dell Technologies recommends designs that apply the following principles:

- Meet network capacity and the separation requirements of the container pod.
- Configure dual homing of the OpenShift Container Platform node to two Virtual Link Trunked (VLT) switches.
- Create a scalable and resilient network fabric to increase cluster size.
- Allow monitoring and tracing of container communications.

### Container network capacity and separation

Container networking takes advantage of the high speed (25/100 GbE) network interfaces of the Dell server portfolio. To meet network capacity requirements, pods can attach to more networks by using available CNI plug-ins.

Additional networks are useful when network traffic isolation is required. Networking applications such as Container Network Functions (CNFs) have control traffic and data traffic. These different types of traffic have different processing, security, and performance requirements.

Pods can be attached to the SR-IOV virtual function (VF) interface on the host system for traffic isolation and to increase I/O performance.

### Dual homing

Dual homing means that each node that makes up the OpenShift cluster has at least two NICs, each connected to at least two switches. The switches require VLT connections so that they operate together as a single unit of connectivity to provide redundant data paths for network traffic. The NICs at each node and the ports that they connect to on each of the switches can be aggregated using bonding to assure HA operation.

### Network fabric

A nonblocking fabric is required to meet the needs of the microservices data traffic. Dell Technologies recommends deploying a leaf-spine network.

## Resilient networking

Dell networking products are designed to enable resilient network creation. Each server that has many NIC options in the rack is connected to:

- Two leaf switches with a network interface of choice: 10 GbE, 25 GbE, or 100 GbE
- A management switch (typically, 1 GbE) for iDRAC connectivity

This DVD employs a VLT connection between the two leaf switches. In a VLT environment, all paths are active and so it is possible to achieve high throughput while still protecting against hardware failures.

VLT technology allows a server to uplink multiple physical trunks into more than one PowerSwitch switch by treating the uplinks as one logical trunk. A VLT-connected pair of switches acts as a single switch to a connecting server. Both links from the bridge network can forward and receive traffic. VLT replaces Spanning Tree Protocol (STP) networks by providing both redundancy and full bandwidth utilization using multiple active paths.

The major benefits of VLT technology are:

- Dual control plane for highly available, resilient network services
- Full utilization of the active link aggregation (LAG) interfaces
- Active/active design for seamless operations during maintenance events

In this DVD, the VLTi configuration uses two 100 GbE ports between each ToR switch. The remaining 100 GbE ports can be used for high-speed connectivity to spine switches, or directly to the data center core network infrastructure.

## Scale out with a leaf-spine fabric

You can scale container solutions by adding multiple compute nodes and storage nodes. A container cluster can have multiple racks of servers. To create a nonblocking fabric that meets the needs of the microservices data traffic, the Dell OpenShift team used a leaf-spine network.

### Leaf-spine concepts

Layer 2 and Layer 3 leaf-spine topologies employ the following concepts:

- Each leaf switch connects to every spine switch in the topology.
- Servers, storage arrays, edge routers, and similar devices connect to leaf switches, but never to spines.

This DVD uses dual leaf switches at the top of each rack. The Dell OpenShift team employed VLT in the spine layer, allowing all connections to be active while also providing fault tolerance. As administrators add racks to the data center, leaf switches are added to each new rack.

The total number of leaf-spine connections is equal to the number of leaf switches multiplied by the number of spine switches. Administrators can increase the bandwidth of the fabric by adding connections between leaf switches and spine switches if the spine layer has capacity for additional connections.

### Layer 3 leaf-spine network

In a Layer 3 leaf-spine network, traffic is routed between leaf switches and spine switches. Spine switches are never connected to each other in a Layer 3 topology. The boundary between Layer 3 and Layer 2 is at the leaf switches. Equal cost multipath routing (ECMP) is used to load-balance traffic across the Layer 3 network. Connections within racks from hosts to leaf switches are Layer 2. Connections to external networks are made from a pair of edge or border leaf switches, as shown in the following figure:

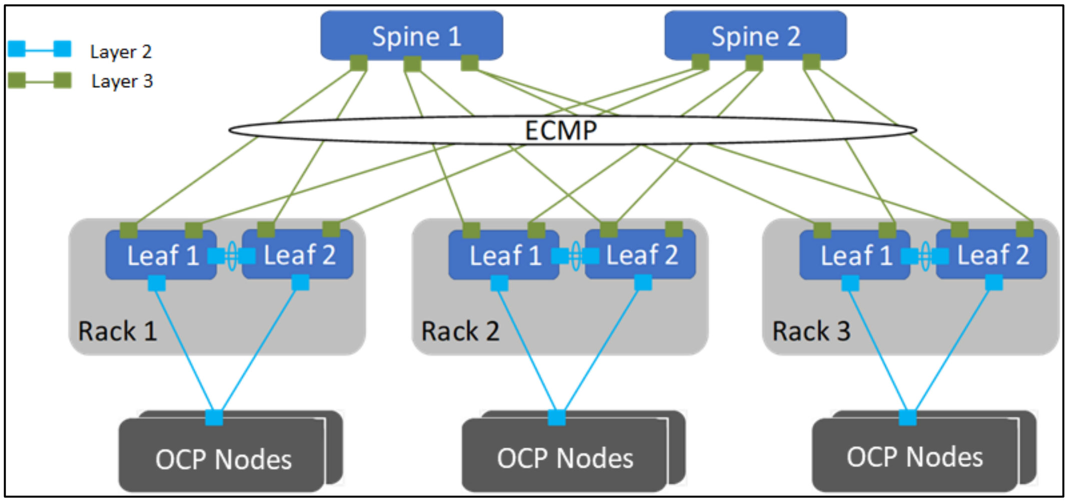


**Figure 4. Leaf-spine network configuration**

## PowerSwitch configuration

The high-capacity network switches from Dell Technologies are cost-effective and easy to deploy. The switches provide a clear path to a software-defined data center, offering:

- High density for 25, 40, 50, or 100 GbE deployments in ToR, middle-of-row, and end-of-row deployments
- A choice of: S5212F-ON, S5224F-ON, S5248F-ON, S5296F-ON, and S5232F-ON 25 GbE and 100 GbE switches

This DVD used Dell Network OS10. OS10 allows multilayered disaggregation of network functions that are layered on an open-source Linux-based operating system.

## High availability

This solution uses the following HA features:

- **Red Hat OpenShift Container Platform 4.14:** Multiple control-plane nodes and infrastructure nodes
- **Resilient load balancing:** Two CSAH nodes running infrastructure services
- **Dell cloud-native infrastructure:** PowerEdge servers with dual NICs
- **Dell PowerSwitch:** Spine-leaf fabric with VLT

# Chapter 4 Storage Overview

This chapter presents the following topics:

# OpenShift Container Platform storage

## Introduction

Stateful applications create a demand for persistent storage. All storage within OpenShift Container Platform 4.14 is managed separately from compute resources and from all networking and connectivity infrastructure facilities. The CSI API is designed to abstract storage use and enable storage portability.

This solution employs the following Kubernetes storage concepts:

- **Persistent volume (PV):** The physical LUN or file share on the storage array. PVs are internal objects against which persistent volume claims are created. PVs are unrelated to pods and pod storage life cycles.
- **Persistent volume claim (PVC):** An entitlement that the user creates for a specific PV.
- **Storage class:** A logical construct defining storage allocation for a given group of users.
- **CSI driver:** The software that orchestrates persistent volume provisioning and deprovisioning on the storage array.
- **Dell Container Storage Modules (CSMs):** These modules extend the enterprise storage capabilities by providing a consistent automation experience. CSMs have multiple components, including enterprise capabilities, CSI drivers (storage enablement), and other related applications (deployment, feature controllers, and more).

These resources are logical constructs that the Kubernetes container infrastructure uses to maintain storage for the container ecosystem components that depend on storage. Developers and operators can deploy applications and provision or deprovision persistent storage without having any specific knowledge of the underlying storage technology.

The OpenShift Container Platform administrator is responsible for provisioning storage classes and making them available to the cluster tenants.

Storage using PVCs is consumed or used in two ways: statically or dynamically. Static storage can be attached to one or more pods by static assignment of a PV to a PVC and then to a specific pod or pods.

- For more information about OpenShift storage concepts, see OpenShift Container Platform storage overview.
- For more information about Dell CSMs, see Container Storage Modules.

### Static persistent storage provisioning

An administrator preprovisions PVs for Kubernetes tenants. When a user makes a persistent storage request by creating a PVC, Kubernetes finds the closest available matching PV. Static provisioning is not the most efficient method for using storage, but it might be preferred when it is necessary to restrict users from PV provisioning.

The following figure shows the static storage provisioning workflow in this solution:

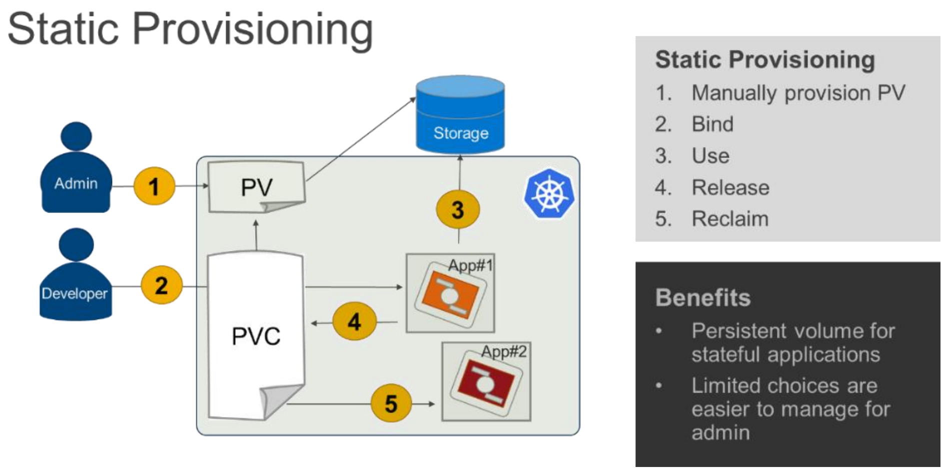


**Figure 5. Static storage provisioning workflow**

## Dynamic persistent storage provisioning

Dynamic persistent storage provisioning enables Kubernetes users to secure PV provisioning on demand. Dynamic provisioning has fully automated LUN export provisioning. For more information, see [OpenShift Container Platform storage overview](OpenShift Container Platform storage overview).

The following figure shows the dynamic storage provisioning workflow in this solution:

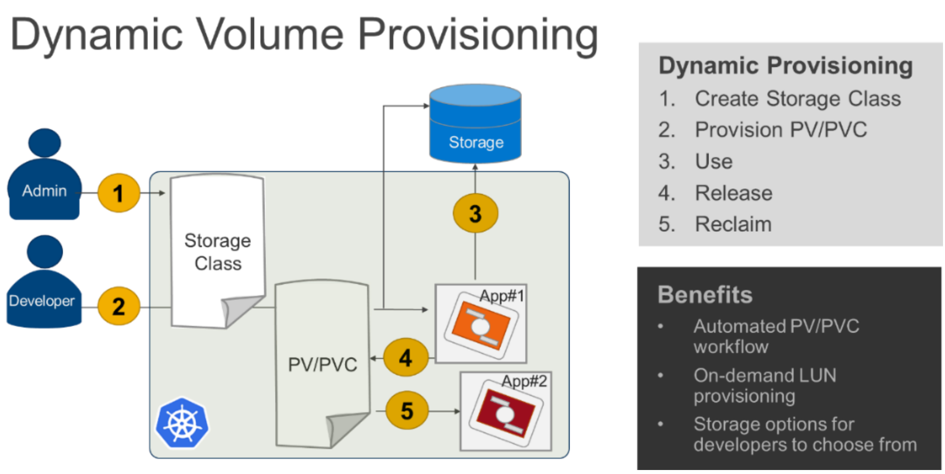


**Figure 6. Dynamic storage provisioning workflow and benefits**

After a PV is bound to a PVC, that PV cannot be bound to another PVC. This restriction binds the PV to the namespace of the binding project. A PV that has been created for dynamic use is a storage class object that functions as a cluster resource, and is automatically consumed as a cluster resource.

### *Persistent volume types*

OpenShift Container Platform supports the following PV types:

- AWS Elastic Block Store (EBS)
- Azure Disk
- Azure File

- Cinder
- Fibre Channel (FC)—can only be assigned and attached to a node
- GCE Persistent Disk
- HostPath (local disk)
- iSCSI (generic)
- Local volume
- NFS (generic)
- OpenStack Manila
- Red Hat OpenShift Data Foundation
- VMware vSphere

The CSI API extends the storage types that can be used within an OpenShift Container Platform solution.

#### *PV capacity*

Each PV has a predetermined storage capacity that is set in its `capacity` parameter. The storage capacity is set or requested by a pod that is launched within the container platform. Expect the choice of control parameters to expand as the CSI API is extended and matures.

#### *PV access modes*

A resource provider can determine how the PV is created and set the storage control parameters. Access mode support is specific to the type of storage volume that is provisioned as a PV. Provider capabilities determine the access modes of the PV, while the capabilities of each PV determine the modes which that volume supports. For example, NFS can support multiple read/write clients, but a specific NFS PV might be configured as read-only.

Pod claims are matched to volumes with compatible access modes based on two matching criteria: access modes and size. Access modes represent a request.

For more information, see Understanding persistent storage.

## Static persistent storage

The use of generic NFS or iSCSI is functional and stable. However, generic NFS and iSCSI do not contain a mechanism to provide service continuity if access to the storage subsystem fails. Also, generic NFS and iSCSI do not provide the advanced storage protection support that is available with CSI drivers. The Dell OpenShift team validated the functionality and capability of suitable storage drivers, as described in the following table:

**Table 4. Generic storage capabilities**

| Storage type | ReadWriteOnce | ReadOnlyMany | ReadWriteMany |
|---|---|---|---|
| HostPath (local disk) | Yes | N/A | N/A |
| iSCSI (generic) | Yes | Yes | N/A |
| NFS (generic) | Yes | Yes | Yes |

# Dell persistent storage options

For operations that are supported on Dell storage using Dell CSI drivers, see Container Storage Modules.

### PowerMax storage

The PowerMax storage array offers an extremely small footprint with fast end-to-end NVMe and SCM performance and rich data services. The array delivers the performance that your applications demand while consolidating block and file storage through iSCSI and Fibre Channel (FC) interfaces. For more information, see Dell PowerMax Data Sheet.

### PowerScale storage

PowerScale, a modern, scale-out NAS solution for file storage, is flexible and reliable at any scale. Whatever the kind of data, where it lives, or how big it gets, your data lake remains simple to manage, simple to grow, simple to protect, and simple enough to handle the most demanding workloads. For more information, see Dell PowerScale Family.

### PowerStore storage

The PowerStore standard deployment model provides organizations with all the benefits of a unified storage platform for block, file, and NFS storage, while also enabling flexible growth with the intelligent scale-up and scale-out capability of appliance clusters. Automated management of resources across the cluster results in superior storage utilization and simplified administration. The PowerStore data-centric design with NVMe and hardware-enabled advanced data reduction and 4:1 DRR delivers critical performance and storage efficiency for both traditional and modern applications. For more information, see Dell PowerStore.

For information about other Dell storage arrays, see the following documentation:

- Dell PowerFlex
- Dell Unity XT

# Data protection

### Introduction

Organizations must protect the business-critical applications that are developed and deployed on OpenShift Container Platform and also protect those applications' data. Protection for cloud-native workloads is a major challenge in container adoption. Protecting an OpenShift environment is not a simple matter of applying a traditional data backup and recovery solution.

### Dell PowerProtect Data Manager

As an innovative leader in protecting Kubernetes, Dell Technologies has evolved, innovated, and integrated with OpenShift to address the needs of the new container infrastructure, including data protection for OpenShift Container Platform and the unique complexities that come with that. Dell PowerProtect Data Manager offers a centralized platform to protect OpenShift workloads.

PowerProtect Data Manager protects OpenShift Kubernetes workloads by ensuring that data is easy to back up and restore and remains available, consistent, and durable in a disaster recovery (DR) situation. PowerProtect Data Manager provides a centralized management UI in which protection policies are defined to manage the clusters, namespaces, and other OpenShift components.

PowerProtect Data Manager provides software-defined data protection, automated discovery, deduplication, operational agility, self-service, and IT governance for physical, virtual, and cloud environments. You can:

- Orchestrate protection directly through an intuitive interface or empower data owners to perform self-service backup and restore operations from their native applications
- Ensure compliance with service level agreements (SLAs) and meet even the strictest SLA objectives
- Use your existing PowerProtect appliances

For general information, see Dell PowerProtect Data Manager.

For detailed information and deployment instructions, see PowerProtect Data Manager: Protecting OpenShift Workloads.

## Provisioning local storage

Use local volumes to provision OpenShift Container Platform with persistent storage. Local PVs enable you to access local storage devices, such as a disk or partition, using the standard PVC interface.

You can use local volumes without manually scheduling pods to nodes because the system is aware of the volume node constraints. Note that local volumes are still subject to the availability of the underlying node and are not suitable for all applications.

# Chapter 5 Cluster Hardware Design

This chapter presents the following topics:

# Overview

This chapter describes node design options that enable you to build a cluster for a wide range of workload-handling capabilities, expanding on information in the Technology and Deployment Process Overview. The goal is to help you get Day-2 operations under way as smoothly as possible. The platform design process ensures that the OpenShift Container Platform 4.14 cluster can meet initial workload requirements. The cluster must also be scalable as the demand for workload handling grows. With a clear understanding of your workloads, it is easier to approach CPU sizing, memory configuration, network bandwidth capacity specification, and storage needs. Many operational factors can affect how the complexity of a container ecosystem affects operational latencies, so it is a best practice to add a safety margin to all physical resource estimates.

# Cluster scaling

The design and architecture of OpenShift Container Platform place resource hosting limits on an OpenShift cluster. Red Hat offers support for OpenShift Container Platform 4.14 up to these limits, as described in planning your environment according to object maximums.

# Cluster hardware planning

## Design limits

The DVD for OpenShift Container Platform 4.14 requires a minimum of four servers for a three-node cluster, with each node running as both a control-plane node and a compute node. A three-node cluster can be expanded to a standard cluster as needed. You can also expand the standard cluster with more compute nodes at any time.

## Server, switch, and rack configuration

This guide uses a server-node base configuration for the PowerEdge R660 and R760 nodes that can be used in each node role for a multinode OpenShift cluster.

For the PowerEdge server baseline configurations that the Dell OpenShift team used in validating the cluster design, see Appendix A.

The following table provides additional cluster configuration information:

**Table 5. Cluster configuration reference information**

| Quantity | Description | Dell Technologies reference |
| --- | --- | --- |
| 1 | Rack enclosure:<br>APC AR3300 NetShelter SZ 42U | APC AR3300 NetShelter SX 42U Enclosure |
| 1 | Management switch:<br>Dell Networking S3148-ON | Dell PowerSwitch S series 1 GbE switches |
| 2 | Data switch:<br>• Dell Networking S5248F-ON<br>or<br>• Dell Networking S5232-ON | Dell PowerSwitch S series 25/40/50/100 GbE switches |

| Quantity | Description | Dell Technologies reference |
|---|---|---|
| 7-210 | • CSAH, Control-plane: Dell PowerEdge R660<br>• Compute nodes: Dell PowerEdge R660<br>or<br>Dell PowerEdge R760 | PowerEdge R660 Rack Server<br>or<br>PowerEdge R760 Rack Server |
| 2-4 | Power distribution unit:<br>APC metered rack PDU 17.2 kW | APC Metered Rack PDU AP8867 - 0U - 208V 3Phase IEC 309 60A 3P+PE Input / (30) C13 Output |

**Note:** Rack enclosures and power distribution units are site-specific. Review the physical dimensions and power requirements in a site survey.

## Performance monitoring, and logging

Cloud service providers typically require the ability to monitor and report on system use. OpenShift Container Platform 4.14 includes Prometheus system monitoring, which provides the capability for extensive data logging. For more information about cluster resource consumption, see the following Red Hat documentation:

- About cluster monitoring
- Managing Metrics
- About cluster logging

The cluster monitoring operator controls the monitoring components that are deployed, and the Prometheus operator controls Prometheus and Alert Manager instances. The platform monitors the following stack components:

- CoreDNS
- Elasticsearch (if Logging is installed)
- Etcd
- Fluentd (if Logging is installed)
- HAProxy
- Image registry
- Kubelets
- Kubernetes API server
- Kubernetes controller manager
- Kubernetes scheduler
- OpenShift API server
- OpenShift controller manager
- Operator Lifecycle Manager (OLM)

# Validated hardware configuration options

## Overview

The Dell OpenShift team used various server configurations for validation testing of the DVD for OpenShift Container Platform 4.14. Dell Technologies recommends selecting server configurations that are known to provide a satisfactory deployment experience and meet or exceed Day-2 operating experience expectations. This chapter provides guidelines for processor selection, memory configuration, local (on-server) disk storage, and network configuration.

## Server processors

The Intel® 4th Generation Xeon® Gold Processor family provides performance, advanced reliability, and hardware-enhanced security for demanding compute, network, and storage workloads. Dell Technologies recommends Intel Gold 6454S Processors for PowerEdge servers. While many sites prefer to use a single-server configuration for all node types, that option is not always cost-effective or practical. When selecting a processor, take account of the following requirements:

- **Processor core count:** The processor core count must be sufficient to ensure satisfactory performance of the workload operations and base services that are running on each node.
- **Thermal design power (TDP):** The CPU must be suitable for the amount of heat that is removed from the server through the heat sinks and cooling airflow.
- **Compute node configurations:** The design of compute nodes for use as part of your OpenShift Container Platform cluster can have many compute node configurations. The processor architecture and core count per node selection can significantly affect the acquisition and operating cost of the cluster that is required to run your organization's application workload.

When ordering and configuring your PowerEdge servers, see:

- Dell PowerEdge R660 Technical Guide
- Dell PowerEdge R760 Technical Guide

For CPU information, see Intel® Xeon® Gold Processors.

## Per-node memory configuration

Modify the memory configuration as necessary to meet your budgetary constraints and operating needs. Also consult OpenShift architectural guidance and consider your own observations from running your workloads on OpenShift Container Platform 4.14.

## Disk drive capacities

Disk drive performance significantly limits the performance of many aspects of OpenShift cluster deployment and operation. The DVD drive selection is based on a comparison of cost per GB of capacity divided by observed performance criteria such as cluster deployment time and application deployment characteristics and performance. Over time, users gain insight into the capacities that best enable them to meet their requirements.

## Network controllers and switches

Your selection of switches for the OpenShift Container Platform cluster infrastructure must take account of the network switches, the overall balance of I/O pathways within server nodes, and the NICs for your cluster. When you choose to include high-I/O bandwidth drives as part of your platform, ensure that sufficient network I/O is available to support high-speed, low-latency drives, as follows:

- **Hard drives:** Lower throughput per drive. You can use 10 GbE for this configuration.
- **SATA/SAS SSD drives:** High I/O capability. SATA SSD drives operate at approximately four times the I/O level of a spinning hard drive. SAS SSDs operate at up to 10 times the I/O level of a spinning hard drive. With SSD drives, configure your servers with 25 GbE.
- **NVMe SSD drives:** High I/O capability, up to three times the I/O rate of SAS SSDs. We populated each node with four 25 GbE NICs or two 100 GbE NICS to provide optimal I/O bandwidth.

The following table provides guidance for ensuring adequate I/O bandwidth and taking advantage of available disk I/O bandwidth:

**Table 6. NIC and storage selection to optimize I/O bandwidth**

| NIC selection | Compute node storage device type |
|---|---|
| 2 x 10 GbE | Spinning magnetic media (hard drive) |
| 2 x 25 GbE or 4 x 25 GbE | SATA or SAS SSD drives |
| 4 x 25 GbE or 2 x 100 GbE | NVMe SSD drives |

True network HA fail-safe design requires that each NIC is duplicated, permitting a pair of ports to be split across two physically separated switches. A pair of PowerSwitch S5248F-ON switches provides 96 x 25 GbE ports, enough for approximately 20 servers. This switch is cost-effective for a compact cluster. For a larger cluster, consider using PowerSwitch S5232F-ON switches. You can also choose to add another pair of S5248F-ON switches to scale the cluster to a full rack.

The PowerSwitch S5232F-ON provides 32 x 100 GbE ports. When used with a four-way QSFP28 to SFP28, a pair of these switches provides up to 256 x 25 GbE endpoints, more than enough for a rackful of servers in the cluster before more complex network topologies are required.

## Power configuration

Dell Technologies strongly recommends that all servers are equipped with redundant power supplies and that power cabling provides redundant power to the servers. Configure each rack with pairs of power distribution units (PDUs). For consistency, connect all right-most power supply units (PSUs) to a right-side PDU, and all left-most PSUs to a left-side PDU. Use as many PDUs as you need, in pairs. Each PDU must have an independent connection to the data center power bus.

The following figure shows an example of the power configuration that is designed to ensure a redundant power supply for each cluster device:

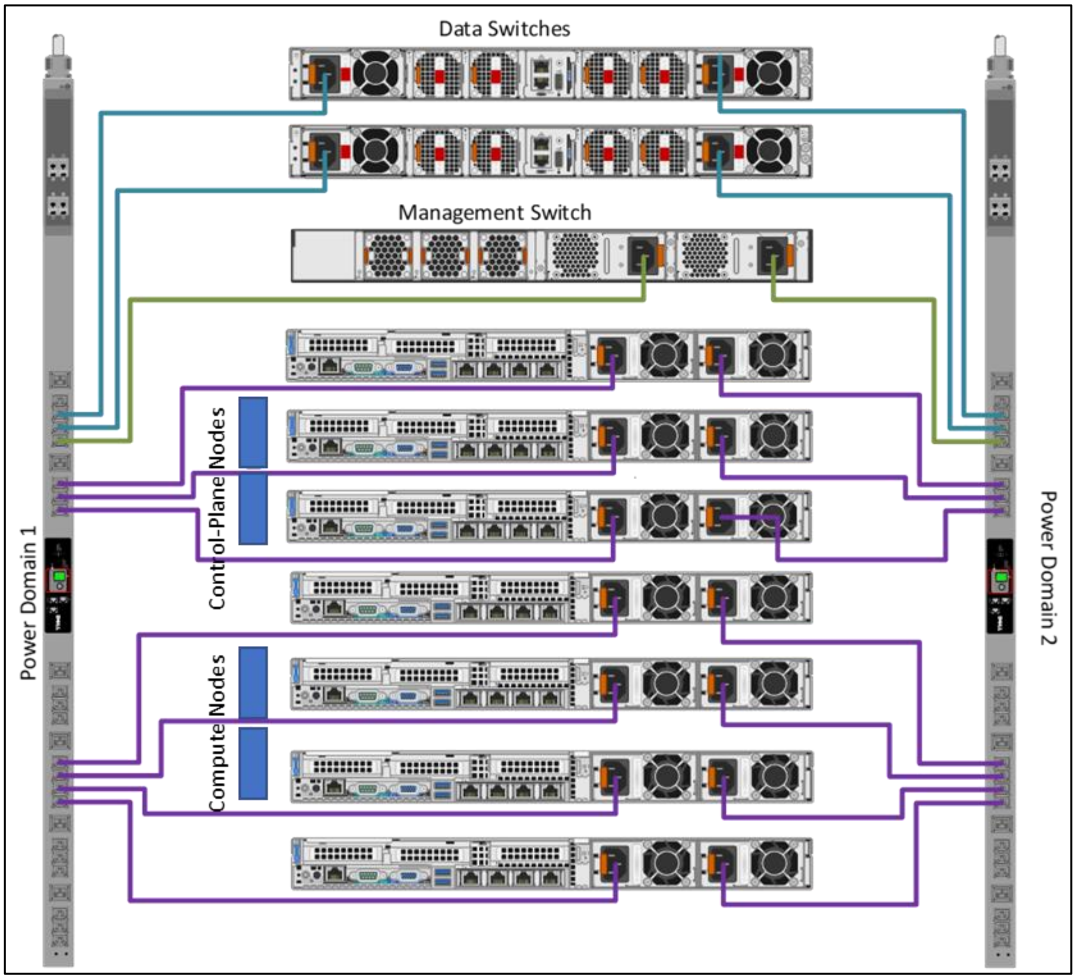


**Figure 7. PSU to PDU power template**

# Chapter 6 References

This chapter presents the following topics:

# Dell Technologies documentation

The following Dell Technologies documentation provides additional information. If you do not have access to a document, contact your Dell Technologies representative.

- Dell PowerEdge R660 Technical Guide
- Dell PowerEdge R760 Technical Guide
- Dell PowerMax
- Dell PowerScale
- Dell PowerStore
- Dell PowerFlex
- Dell Unity XT
- PowerProtect Data Manager Protecting OpenShift Workloads
- Dell Technologies Container Storage Modules

**Note**: For additional documentation, see the Dell Technologies Info Hub for Red Hat OpenShift Container Platform.

# Red Hat documentation

The following Red Hat documentation provides additional information:

- Installing on Bare Metal
- What are Operators?
- Recommended practices for scaling the cluster
- Understanding the monitoring stack
- About logging
- Machine management
- Planning your environment according to object maximums
- Red Hat OpenShift AI
- The HyperShift component of the OpenShift Container Platform

# Other resources

The following resources provide additional information:

- Intel® Xeon® Gold Processor
- Operating etcd clusters for Kubernetes
- NVIDIA® H100 Tensor Core GPU

# Appendix A Hardware Configuration

This appendix presents the following topics:

# Overview

The tables in this appendix show the Bill of Materials (BOM) for each node that is used in the DVD for OpenShift Container Platform 4.14. The memory, CPU, NIC, and drive configurations that are shown in the tables are preferred, but not mandated.

**Note:** When orders are placed, the Dell Technologies ordering center adds new SKUs and substitutes the SKUs that are shown in the tables with the current local SKUs.

# Dell PowerEdge R660 BOM

The following table shows the PowerEdge R6625 baseline configurations that the DVD uses:

**Table 7. PowerEdge R660 BOM**

| Quantity | SKU | Description |
|---|---|---|
| 1 | [210-BEQQ] | PowerEdge R660 server |
| 1 | [379-BEID] | 10 x 2.5 front storage |
| 1 | [379-BDSW] | SAS/SATA/NVMe capable backplane |
| 1 | [321-BHRZ] | 2.5" chassis with up to 10 hard drives (SAS/SATA), 2CPU, PERC12 |
| 1 | [338-CHTD] | Intel® Xeon® Gold 6438Y+ 2G, 32C/64T, 16GT/s, 60M Cache, Turbo, HT (205W) DDR5-4800 |
| 1 | [338-CHTD], [379-BDCO] | Intel® Xeon® Gold 6438Y+ 2G, 32C/64T, 16GT/s, 60M Cache, Turbo, HT (205W) DDR5-4800 |
| 1 | [412-ABCG] | Heatsink for 2 CPU configuration (CPU more than or equal to 250W) |
| 1 | [370-AHCL] | 4800MT/s RDIMMs |
| 8 for the control-plane node | [370-AGZO] | 16 GB RDIMM, 4800MT/s single rank |
| 12 for the data-plane node | [370-AGZR] | 64 GB RDIMM, 4800MT/s dual rank |
| 1 | [405-ABDN], [750-ADWX] | PERC H965i with rear load bracket |
| • 2 for the control-plane node<br>• 6 for the data-plane node | [345-BEOZ] | 1.6 TB SSD SAS mixed use up to 24Gbps 512e 2.5in hot-plug 3WPD, AG drive |
| 2 (for the data-plane node) | [400-BKGF] | 1.6 TB, Enterprise, NVMe, mixed use, U2, G4, P5600 with carrier |
| 1 | [450-AKLF] | Dual, hot-plug, power supply fault tolerant redundant (1+1), 1100W MM (100-240Vac) titanium, NAF |

| Quantity | SKU | Description |
|---|---|---|
| 1 | [528-CTID], [528-CTZH] | iDRAC9 data center 16G with OpenManage Enterprise Advanced Plus |
| 1 | [540-BDJL] | Nvidia ConnectX-6 Lx dual port 10/25GbE SFP28, OCP NIC 3.0 |
| 1 | [540-BDJO] | Nvidia ConnectX-6 Lx dual port 10/25GbE SFP28 Adapter, PCIe low profile |
| 1 | [406-BBTM] | Emulex LPe35002 dual port FC32 fibre channel HBA, PCIe low profile |
| 1 | [403-BCRZ], [470-AFMG] | BOSS-N1 controller card + with 2 M.2 960 GB (RAID 1) |

## Dell PowerEdge R760 BOM

The following table shows the PowerEdge R760 baseline configurations that the DVD uses:

**Table 8. PowerEdge R760 BOM**

| Quantity | SKU | Description |
|---|---|---|
| 1 | [210-BDZY] | PowerEdge R760 server |
| 1 | [379-BDTF] | Chassis with up to 24x2.5" drives |
| 1 | [379-BDSW] | SAS/SATA/NVMe capable backplane |
| 1 | [404-BBEM] | 2.5" chassis with up to 24 drives (16 SAS/SATA + 8 NVMe Direct), Front PERC 12, MaxIO, 2 CPU |
| 1 | [338-CHTD] | Intel® Xeon® Gold 6438Y+ 2G, 32C/64T, 16GT/s, 60M cache, Turbo, HT (205W) DDR5-4800 |
| 1 | [338-CHTD], [379-BDCO] | Intel® Xeon® Gold 6438Y+ 2G, 32C/64T, 16GT/s, 60M cache, Turbo, HT (205W) DDR5-4800 |
| 1 | [412-ABCQ] | Heatsink for 2 CPU with GPU configuration |
| 1 | [370-AHCL] | 4800MT/s RDIMMs |
| 12 | [370-AGZR] | 64 GB RDIMM, 4800MT/s dual rank |
| 1 | [405-ABDN], [750-ADWO] | PERC H965i with rear load bracket |
| 10 | [345-BEOZ] | 1.6TB SSD SAS mixed use up to 24 Gbps 512e 2.5in Hot-Plug 3WPD, AG drive |
| 2 | 400-BKGF] | 1.6TB Enterprise NVMe mixed use AG Drive U.2 Gen4 with carrier |
| 1 | [450-AJEV] | Dual, Hot-Plug, Fully Redundant Power Supply (1+1), 2400W, Mixed Mode |
| 1 | [528-CTID], [528-CTZH] | iDRAC9 data center 16G with OpenManage Enterprise Advanced Plus |

| Quantity | SKU | Description |
|---|---|---|
| 1 | 540-BDJL] | Nvidia ConnectX-6 Lx Dual Port 10/25GbE SFP28, OCP NIC 3.0 |
| 1 | 540-BDJO] | Nvidia ConnectX-6 Lx Dual Port 10/25GbE SFP28 adapter, PCIe Low Profile |
| 1 (optional for GPU intensive workloads) | [490-BJBZ] | NVIDIA Hopper H100, PCIe, 300W-350W, 80 GB passive, double wide, GPU |
| 1 | [403-BCRZ], [470-AFMF] | BOSS-N1 controller card + with 2 M.2 960 GB (RAID 1) |

# Dell PowerStore 5000T BOM

The following table shows the Dell PowerStore 5000T baseline configurations that the DVD uses:

**Table 9. Dell PowerStore 5000T BOM**

| Quantity | SKU | Description |
|---|---|---|
| 1 | [210-ASTS] | PowerStore 5000T base dell customer racked |
| 1 | [406-BBOO] | 25GBE Optical 4-port card pair |
| 1 | [565-BBJS] | 32 GB FC 4-Port I/O module pair |
| 12 | [400-BGGM] | 3.84 TB NVMe SSD |
| 1 | [343-BBMR] | PowerStore field install kit |
| 1 | [450-AIOM] | Dual 1800 1800 W (200-240V) power supply. Includes C13/C14 power cables |
| 4 | [470-ADUJ] | 3M OM4 fiber cable QTY2 |
| 8 | [470-ADUI] | 3M passive 25G Twinax cable QTY 2 |
| 4 | [407-BCGC] | 32G FC Multi Mode Optical SFP Pair |
| 1 | [370-AEZR] | 1152 GB appliance DIMM (576 GB per node) |
| 1 | [379-BEIQ] | 4:1 |
| 1 | [528-BTZK] | PowerStore base SW |
| 2 | [450-AIOH] | C19 power cable pair NEMA5-15 125V 10A 2Meter |
| 1 | [800-BBQV] | Clustered storage |
| 1 | [210-ATXO] | AppSync for PowerStore |
| 1 | [528-BYHF] | AppSync Str Pk for PowerStore |

# Dell PowerScale H7000 BOM

The following table shows the Dell PowerScale H7000 baseline configurations that the DVD uses:

**Table 10. Dell Isilon H7000 BOM**

| Quantity | SKU | Description |
|---|---|---|
| **4** | | **PowerScale H7000** |
| 1 | [210-AZWP] | H7000 - 320 TB (20x16TB)/2x3.2TB SSD |
| 1 | [800-BBQV] | Two 3.2TB cache SSDs included |
| 1 | [590-TFHH] | 2x100GbE (QSFP28) back-end W/O OPTICS |
| 1 | [590-TFHF] | 2x100GbE (QSFP28) W/O OPTICS |
| 1 | [709-BDXM] | Parts only warranty 36 months, 36 months |
| 1 | [199-BJZJ] | ProSupport and 4Hr Mission Critical, 36 month(s) |
| 1 | [683-24553], [706-12470] | ProDeploy Plus for PowerScale node |
| 1 | [800-BBQV] | VP software |
| **1** | | **PowerScale Accessories** |
| 1 | [210-AYYS] | Accessories Virtual Base - VI |
| 1 | [800-BBQV] | Ethernet back-end I/O |
| 2 | [450-AJHN] | PWCRD KIT for deep chassis |
| **1** | | **PowerScale Services** |
| 1 | [210-AYWM] | PowerScale services |
| 1 | [519-10872] | ProDeploy Plus add-on for PowerScale Advanced Bundle |
| **1** | | **PowerScale Chassis** |
| 1 | [210-BBKQ] | Base chassis - Deep H-Series |
| **1** | | **DataIQ** |
| 1 | [210-BCNB] | DataIQ perpetual base - VP |
| 1 | [142-BBDM] | DataIQ Dell base Lic=IF |
| **1** | | **Isilon Rack** |
| 1 | [210-AUVO] | RACK ASSY TITAN-HD 42U FULL HGT PDU SINGLE PHASE |
| 1 | [450-AJHQ] | PWCRD, 6, SINGLE PHASE G2/3 15F L6-30P/NORTH AMER |
| **4** | | **PowerScale OE software** |
| 1 | [210-BBWD] | PowerScale hybrid OESW virtual base |
| 1 | [800-BBQV] | H7000 - 320 TB (20x16TB)/2x3.2TB SSD |

| Quantity | SKU | Description |
|---|---|---|
| 1 | [800-BBQV] | OneFS |
| 1 | [149-BBGS] | OneFS base license H7 12TB+ tier 3 |
| 320 | [149-BBGN] | OneFS capacity H7 tier 3 |
| **4** | | **PowerScale Additional Software** |
| 1 | [210-BCEP] | PowerScale hybrid ADDSW virtual base |
| 1 | [800-BBQV] | H7000 - 320TB (20x16TB)/2x3.2TB SSD |
| 1 | [151-BBGV], [151-BBHE], [151-BBHF], [151-BBHW], [151-BBIH], [151-BBIS] | Enterprise Advanced Bundle H7 Tier 3 |
| 320 | [151-BBGY], [151-BBHG], [151-BBHH], [151-BBJH], [151-BBJR], [151-BBLW] | Enterprise Advanced Bundle Capacity H7 Tier 3 |
| 1 | [151-BBIW] | SmartConnect Base License H7 Tier 3 |
| 1 | [151-BBHD] | SmartDedupe Base License H7 Tier 3 |
| 1 | [151-BBIV] | SmartLock Base License H7 Tier 3 |
| 1 | [151-BBHU] | SmartPools Base License H7 Tier 3 |
| 1 | [151-BBGW] | SmartQuotas Base License H7 Tier 3 |
| 1 | [151-BBIR] | SnapshotIQ Base License H7 Tier 3 |
| 1 | [151-BBIK] | SyncIQ Base License H7 Tier 3 |
| 1 | [151-BBEE] | SWIFT for OneFS ($0.00) |
| 1 | [151-BBDR] | STIG Hardening for OneFS ($0.00) |
| 1 | [151-BBES] | CloudPools 0.00 Dell Base License |
| 1 | [151-BBDM] | CloudPools for Third Party 0.00 Base License |
| 320 | [151-BBKB] | SmartConnect Capacity License H7 Tier 3 |
| 320 | [151-BBJO] | SmartDedupe Capacity License H7 Tier 3 |
| 320 | [151-BBHS] | SmartLock Capacity License H7 Tier 3 |
| 320 | [151-BBGX] | SmartPools Capacity License H7 Tier 3 |
| 320 | [151-BBJG] | SmartQuotas Capacity License H7 Tier 3 |
| 320 | [151-BBIC] | SnapShotIQ Capacity License H7 Tier 3 |
| 320 | [151-BBHQ] | SyncIQ Capacity License H7 Tier 3 |
| **1** | | **PowerScale CloudPools Capacity** |
| 1 | [210-BCEO] | PowerScale Hybrid CloudPools Virtual Base |
| 1 | [151-BBJQ] | SmartConnect Base License A3 Tier 3 L3 |
| 1 | [151-BBDX] | CloudPools for Dell EMC Capacity License (TB tiered to Cloud) |

| Quantity | SKU | Description |
|---|---|---|
| 1 | [151-BBEW] | CloudPools for Third-Party Capacity License (TB tiered to Cloud) |
| 1 | [800-BBQV] | CloudPools |
| **4** | | **Dell Network Switches** |
| 1 | [210-BCUZ] | S5232F Dell Networking Switch ROW |
| 1 | [528-CSYU] | OS10 Enterprise Software, S5232F |
| 1 | [750-ACVZ] | Dell Switch Field Installation Kit for S5232 |
| 1 | [800-BBQV] | Standard offer |
| 1 | [709-BDXY] | Parts only warranty 36 months |
| 1 | [199-BJZJ] | ProSupport and 4Hr mission critical, 36 months |

## Dell PowerMax BOM

The following table shows the PowerMax array baseline configurations that are used in this DVD:

**Table 11. Dell PowerMax BOM**

| Quantity | SKU | Description |
|---|---|---|
| 1 | EH-SYS1-3D | PowerMax 2000 SYS BAY1 3D |
| 1 | EH-PSNT-3D | PowerMax 2000 SYS |
| 1 | EH-SKINS | PowerMax 2000 SIDE PANELS |
| 1 | EH-VBX-1024 | PowerMax 2000 PRO BASE 1024 GB |
| 8 | QNDN3192051 | PowerMax 2000 RAID 5(3+1) 1.92 TB |
| 1 | EH-PCBL3DHR | PWR CBL HBL-RSTOL 3D |
| 2 | EH-ACON3P-50 | ADPTR AC 3 PH 50A W3-4IN CONDUIT ADPTR |
| 1 | EH-VBBASE-KIT | PowerMax 2000 PB base install kit |
| 2 | EH-DE24 | PowerMax 2000 DIRECT 24 SLT DR ENCL |
| 1 | EHX-BEDIR | PowerMax 2000 PRO DIR |
| 1 | EH-1024BASE | PowerMax 2000 BASE 1024 GB |
| 1 | EH-FE00800T | PowerMax 2000 8MM 10GIGE |
| 1 | EH-COMPRESS | PowerMax 2000 HDW compression |
| 1 | QNDN31920S1 | PowerMax 2000 1.92 TB spare |
| 2 | EH-1600MOD | PowerMax 2000 FLASH MOD 1600 |
| 8 | E-GE-ISCSI | VMAX VG GIGE ISCSI port tracking model |
| 3000 | E-DRR | PowerMax 2/8K data reduction reservation |

| Quantity | SKU | Description |
|---|---|---|
| 1 | E-Q118E | ELM tracking model |
| 13 | E-OPROVISION | Oprovision factor tracking model |
| 1 | WKPROFILE-BAL | VMAX VG work profile balanced |
| 1 | EH-MGMT | Embedded management PowerMax 2000 TRK |
| 1 | M-PSM-HW-020 | ProSupport 4HR/MC hardware support |
| 11 | 458-002-223 | PowerMax PRO Suite OS 1 TB=CC |
| 1 | M-PSM-SW-020 | ProSupport 4HR/MC software support |
| 1 | 450-001-644 | PowerMax PRO SUITE=IC |
| 1 | M-PSM-SW-020 | ProSupport 4HR/MC software support |
| 11 | 450-001-645 | PowerMax PRO SUITE 1 TB=CC |
| 1 | M-PSM-SW-020 | ProSupport 4HR/MC software support |
| 1 | PS-PD-PMAX2DP | PD for PowerMax 2000 |
| 1 | PSINST-ESRS | Zero Dollar Esrs Install |
| 1 | CE-VALPAKPMAXSADM | PowerMax/VMAX AF storage admin valuepak |