Dell EMC Ready Architecture for Red Hat OpenShift Container Platform v3.11

April 2019

H17679

Architecture Guide

Abstract

This architecture guide describes how to design a Red Hat OpenShift Container Platform solution on Dell EMC infrastructure, including Dell EMC PowerEdge Servers and Dell EMC Networking switches, for an on-premises deployment.

Dell EMC Solutions

Copyright

Contents

Chapter 1 Introduction

This chapter presents the following topics:

Ready Architecture overview

Digital transformation includes processes that enable an organization to move from the use of monolithic software applications to a more agile environment that can benefit from cloud-native-style microservice applications running in an appropriate execution environment.

Applications are at the core of modern businesses. To support faster application development and deployment, many IT organizations are turning to cloud- and container-based technologies. Containers bring numerous benefits, including mechanisms to run multiple isolated applications on a single host with a smaller footprint. Using a container platform to manage containers across the data center can simplify, accelerate, and orchestrate application development and deployment. Containers can be used with virtual machines (VMs) or on bare-metal operating system deployments. In both cases, containers share the host operating system kernel and decoupled libraries to significantly reduce platform maintenance costs. With VMs, the hypervisor abstracts the hardware on which VMs run. Containers depend on a container run-time engine, Linux control groups, the kernel NameSpace, and SELinux security settings. The container run-time engine abstracts the operating system on which the application code runs.

Not every IT organization has the time or resources to research, integrate, and test all the components that are required to deploy a customized container infrastructure. Together, Dell EMC and Red Hat take the guesswork and risk out of container platform deployment (Day 1) and operations (Day 2) with a complete, integrated reference architecture. This reference architecture guide shows you how to design and specify the hardware components that are needed to build a private cloud environment by using Red Hat OpenShift Container Platform on Intel-based Dell EMC infrastructure.

Red Hat OpenShift Container Platform can be deployed as a hyperconverged infrastructure, where servers combine compute and storage on the same nodes. Customers can conveniently deploy OpenShift Container Platform with discrete disaggregated compute and storage resources. Servers must be carefully chosen and grouped so that compute, memory, storage, and networking resources are effectively and efficiently utilized to ensure run-time availability and continuity for all hosted business applications. For applications that are designed to take advantage of (predominantly) horizontal scaling, containerization technologies such as Kubernetes permit rapid application scaling that keeps pace with changing workload demands.

The Dell EMC Ready Architecture for Red Hat OpenShift Container Platform is a proven design to help companies accelerate their container deployments and lead to cloud-native adoption. Dell EMC delivers tested, validated, and documented design guidance to help customers rapidly deploy Red Hat OpenShift on Dell EMC infrastructure and minimize the time and effort needed to get up and running.

Key benefits

Containerization provides four main benefits that are key to the rapid adoption of cloud-native microservice applications:

- **Encapsulation—**Containers solve a major application deployment problem—ensuring the runtime integrity of software applications when running under a

shared OS. Containers provide a layered framework for modular packaging and the deployment of "just-enough" code to include services or functions as part of an entire software solution stack. The runtime immutability of containers helps ensure the integrity of the microservices and is the basis of the Red Hat OpenShift Container Platform.

- **Distribution—**Businesses increasingly need to distribute the deployment of the software applications on which they depend. The emerging workflow methodology known as DevOps was inspired by this trend. Containers are best suited to distributed deployment because of the modularity of the package within which they can be executed. No other packaging model matches the dexterity of containerization to enable the efficient manageability of microservices. Containers are the ideal enabler for continuous development, deployment and integration, and therefore reduce the time from concept to delivery of application functionality.
- **Portability—**Containers and container orchestration engines such as Kubernetes enable agile software application developers to build an application and run it anywhere in their distributed DevOps platform environment.
- **Agility—**Docker provided the first broadly consumable containerization implementation, while Kubernetes has been an enabling factor for wider adoption. Kubernetes provides highly reusable, modular, shared software application services (such as service discovery, load balancing, and the use of distributed firewalls) and the means to operate them reliably and securely. Above all, Kubernetes can lead to faster application deployment and reduced upgrade times while potentially improving application availability.

Containerization of applications results in smaller, more agile application packages that can be easily deployed when another instance is needed. The key to the efficient use of containers lies in their orchestration. Red Hat OpenShift Container Platform is a secure enterprise implementation of open source Kubernetes, the leading orchestration framework for containers.

Ready Architecture benefits

This section focuses on why a joint customer of Dell EMC and Red Hat would choose the combination of Dell EMC products and Red Hat OpenShift technology.

Customers who are building a DevOps environment to use an incremental application development methodology combined with application deployment and management face a choice: selecting what they consider to be the appropriate components for that environment and either integrating those components themselves or using a pre-integrated tool set. OpenShift belongs to the latter category.

OpenShift has evolved from a web service offering a Platform as a Service (PaaS) for developers into a fully featured on-premises (or cloud-based) DevOps environment. This DevOps environment uses industry-leading Open Source technologies, such as Kubernetes or Docker format containers integrated with toolchains such as Prometheus, with technologies developed by Red Hat, such as JBoss AS. This combination offers the customer a fully supported environment while accelerating the time to production of a DevOps environment.

Combining OpenShift software technology with the selection, preparation, and deployment methodologies for the Dell EMC Ready Architecture platform, storage, and networking configuration reduces the time needed to move into the production phase of a deployment. This choice further reduces the predeployment investment in setup time and effort. Customers can feel confident of a successful and complete hardware and software platform for their application development.

Document purpose

This ready architecture guide describes the infrastructure necessary for deployment and operation of the application deployment platform, providing the information that you need to facilitate readiness for both initial and ongoing operations. Topics include the architectural requirements that you must meet to successfully deploy:

- **Red Hat OpenShift Container Platform v3.11—**An enterprise Kubernetes-based container application deployment platform
- **Red Hat OpenShift Container Storage—**Software-defined storage that is integrated with and optimized for Red Hat OpenShift Container Platform and that runs anywhere OpenShift runs
- **Dell EMC PowerEdge R640 and R740xd servers—**The industry choice for dependable and durable compute and storage needs
- **Dell EMC Networking S-5200 series switches—**The network enablement platform of choice for providing a highly agile communications environment that is well suited to container cloud environments
- **Dell EMC Networking S-3000 series switches—**The devices used for the out of band (OOB) management of the cluster infrastructure.

This guide includes:

- An overview of the container ecosystem design
- Hardware requirements to support Red Hat OpenShift Container Platform node roles
- Hardware platform configuration requirements
- Network architecture and switch configuration details
- Hardware bill of materials for all components required to assemble the OpenShift cluster
- Rack-level design and power configuration considerations

For information about the manual installation and deployment of Red Hat software products, see Product Documentation for OpenShift Container Platform 3.11.

Audience

This reference architecture is for system administrators and system architects. Some experience with Docker and Red Hat OpenShift Container Platform technologies is helpful, but is not required.

We value your feedback

Dell EMC and the authors of this document welcome your feedback on the solution and the solution documentation. Contact the Dell EMC Solutions team by email or provide your comments by completing our documentation survey. Alternatively, contact the Dell EMC OpenShift team at openshift@dell.com

Authors: John Terpstra, Stephen Wanless, Scott Powers, Aighne Kearney

Note: The OpenShift Container Platform Info Hub for Ready Solutions space on the Dell EMC Communities website provides links to additional documentation for this solution.

Chapter 2 Solution Overview

This chapter presents the following topics:

Overview

This chapter describes the architecture of the Red Hat Openshift Container Platform. The Red Hat Openshift Container Platform handles cloud-native and traditional applications on a single platform. With this platform, you can containerize and manage your existing applications, modernize on your own timeline, and work faster with new cloud-native applications.

New features

Dell EMC published the *Red Hat Intel Architecture and Deployment Guides for Red Hat OpenShift Container Platform v3.10* in November 2018. This document updates the networking and storage design of Openshift Container Platform v3.10 to align with current practices. Version 3.11 of OpenShift Container Platform has the following new features:

- **An administrator-oriented console—**An operator is available to install Prometheus with default alerts for the cluster. Grafana dashboards provide deep insight into operating metrics. This console provides advanced technical control with:
  - A CaaS environment that tightly integrates with Kubernetes
  - AppDev/PaaS experience with standard OpenShift user experience
  - Credentials that are shared across consoles, but not sessions
  - The administrator-oriented console and the CaaS environment which are hosted on the cluster and are accessible from the openshift-console and openshift-webconsole namespaces
  - Expanded node status event reporting assistance with diagnosis of resource utilization
  - Protection using RBAC so that metrics are not publicly available
- **Application services—**Operator-framework driven application services have been added to the OpenShift Container Platform. The Helm operator provides expanded ecosystem content management. Ansible play book automation and execution has been integrated with Ansible Galaxy and now enables viewing, editing, and deletion of the full range of Kubernetes objects:
- **Networking—**Routes and ingress
- **Storage—**Persistent Volumes (PVs), Persistent Volume Claims (PVCs), and Storage Classes
- **Admin—**Projects and Namespaces, Nodes, Roles and RoleBindings, CRDs

Red Hat OpenShift Container Platform v3.11 provides a consistent, reliable, and Cloud Native Compute Foundation (CNCF)-verified enterprise Kubernetes container platform.

OpenShift platform architecture

The Red Hat OpenShift Container Engine provides an enterprise Kubernetes platform for hybrid cloud deployments.

The following figure shows the architecture of the OpenShift platform.

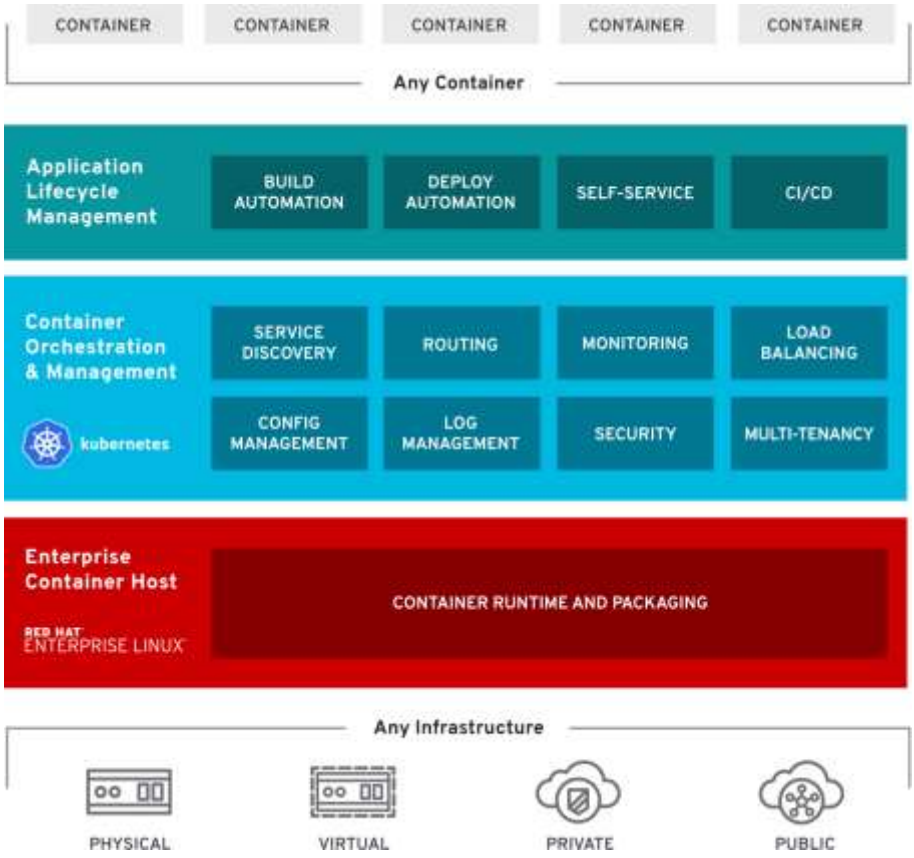

Figure 1. OpenShift system architecture

The lowest layer consists of the hosting platform, which comprises Dell EMC PowerEdge servers and Dell EMC Networking switches in this reference architecture. Dell EMC storage products may be integrated into this layer to create a comprehensive platform for hosting the Openshift Container Platform, as shown in the first layer in Figure 1.

Container ecosystem platform

Openshift Container Platform provides enterprises with full control over the Kubernetes environment. OpenShip Container Platform includes:

- Self-service dashboard access to control application and development
- A pluggable architecture that gives you a choice of supported container runtimes, networking, storage, and continuous integration and continuous delivery (CI/CD) solutions
- Built-in automation integrated into an orchestrator framework
- Unified operations across the OpenShift environment

- On-demand services that seamlessly integrate configuration, deployment, and consumption of services from the OpenShift Service Catalog
- A platform that is part of the CNCF-certified Kubernetes program, ensuring a maximum level of portability and interoperability for your container workloads

Container ecosystem infrastructure

The container ecosystem software infrastructure described in Chapter 4 consists of five key parts, each of which is deployed on servers that provide a functional service:

- A bastion node.
- Three master enterprise-grade nodes. This number of nodes is sufficient for several hundred workload/application nodes and many thousands of containers.
- Three or more infrastructure nodes.
- Four or more workload/application nodes.
- Storage nodes. The storage node type depends on the type of storage that is provided within the cluster design. This Ready Architecture makes use of Red Hat OpenShift Container Storage.

Container ecosystem storage

Red Hat OpenShift Container Storage, which is based on GlusterFS, can be used with Red Hat OpenShift Container Platform. Both products require separate subscription support. Red Hat OpenShift Container Storage is available in Standard or Premium subscription levels.

The deployment example in this guide uses integrated GlusterFS storage for infrastructure. The sample deployment uses persistent volumes on the infrastructure nodes for the deployment of containers onto the workload/application nodes and potentially also for storage of workload/application data. Carefully map out your storage needs and how you want to implement this within your OpenShift cluster. It is important that your cluster design includes an appropriate level of data protection and backup.

Chapter 3 Hardware Infrastructure

This chapter presents the following topics:

Overview

The Red Hat OpenShift Container Platform is built around a core of containers run by the Docker container runtime. Orchestration and management are provided by Kubernetes and powered by Intel Xeon Gold processors and solid-state drives (SSDs). Dell EMC PowerEdge servers deliver compute and storage resources for the application environment and Dell EMC Networking switches provide connectivity between nodes and to external networks.

Hardware components

This Dell EMC Ready Architecture requires 15 server nodes and includes the following hardware:

- 11 Dell EMC PowerEdge R640 servers that are distributed among three master nodes, three infrastructure nodes, four application nodes, and one bastion node
- 4 Dell EMC PowerEdge R740xd servers that are used as storage nodes
- 2 Dell EMC Networking S5248F-ON or S5232F-ON switches that are used as the Top-of-Rack (ToR) data switches
- 1 Dell EMC Networking S3048-ON switch that is used as a management switch

Dell EMC PowerEdge servers

PowerEdge R640 servers

The Dell EMC PowerEdge R640 server is a 2-socket, 1U platform designed for dense scale-out data center computing. With Intel Xeon scalable processors and support for up to 24 DIMMs, 12 of which can be non-volatile DIMMS (NVDIMMs), the scalable architecture enables you to customize the configuration to optimize your workload performance. With the PowerEdge R640, you can maximize storage performance with a non-volatile memory express (NVMe) cache pool of up to 10 NVMe drives or an array of twelve 2.5 in. drives. For more information, see Dell EMC Online Support.

PowerEdge R740xd servers

The Dell EMC PowerEdge R740xd server is a 2-socket, 2U platform designed for scalable, high-performance, software-defined storage. The versatile system architecture of the R740xd server allows you to mix any drive types to create the optimum configuration of NVMe, SSD, and hard disk drive (HDD) to meet your storage needs. With support for up to 24 SAS SSD drives or up to 12x NVMe drives, you can ensure that the storage performance scales to meet application demands. For more information, see Dell EMC Online Support.

Dell EMC Networking ToR switches

Data plane switches

Data plane networking switches carry all OpenShift cluster ingress/egress traffic, all internal traffic, and all storage traffic. Dell EMC engineers chose base-line switches for the Ready Architecture design to handle anticipated maximum data flow rates over a 25 GbE connection. For an economy of scale or to provide a longer switch lifetime, you might decide to use 25 GbE switches or 100 GbE switches to provision the network infrastructure.

Dell EMC Networking S5248F-ON switch

The Dell EMC Networking S5248F-ON switch is a multi-rate ToR data center switch that delivers the architectural agility and flexibility that is needed to deploy software-defined infrastructure so that you can easily deploy cost-effective, high-capacity network fabrics. The S5248F-ON switch provides optimum flexibility and cost-effectiveness for demanding compute and storage traffic environments. This switch features 48 x 25 GbE SFP28 ports, 4 x 100 GbE QSFP28 ports, and 2 x 100 GbE QFSP28-DD ports.

The S5248F-ON switch also supports Open Network Install Environment (ONIE) for zero-touch installation of network operating systems. Line-rate performance delivered by non-blocking switch fabrics is 2.0 Tbps on the S5248F-ON switch.

Dell EMC Networking S5232F-ON switch

The Dell EMC Networking S5232F-ON switch is a suitable ToR choice that supports a configuration in a rack of up to 30 nodes. The S5232F-ON switch is also suitable for building out a multirack cluster infrastructure.

The S5232F-ON switch enables you to build a high-performance, cost-efficient data center leaf/spine fabric with a switch featuring 32 x 100 GbE QSFP28 ports. The S5232F-ON switch also supports ONIE for zero-touch installation of network operating systems. Non-blocking switch fabrics provide a line-rate performance of 3.2 Tbps S5232F-ON switches.

For more information, see the Dell EMC Networking S5200 Series Switches Document Library.

Dell EMC Networking management switches

Dell EMC Networking S3048-ON switch

The Dell EMC Networking S3048-ON switch is designed for reliable server aggregation and cost-effective deployment. It is used here as the OOB management switch. With 48 x 1 GbE and 4 x 10 GbE ports, a dense 1U design, and up to 260 Gbps performance, the S3048-ON switch delivers low latency and high density with hardware and software redundancy.

For more information, see Dell EMC Networking S-Series 1GbE switches.

Chapter 4 Software Infrastructure

This chapter presents the following topics:

Overview

This chapter describes the software components in the Dell EMC Ready Architecture for Red Hat OpenShift Container Platform v3.11.

The ready architecture includes the following software:

- Red Hat Enterprise Linux as the operating system for all nodes
- Red Hat Openshift Container Platform as the container application platform based on Kubernetes and Docker
- Red Hat OpenShift Container Storage based on Red Hat Gluster Storage to provide a containerized, distributed storage platform for both persistent volume storage and the container registry
- Red Hat Ansible Automation to provision resources, deploy applications, and configure and manage infrastructure
- Integrated Dell Remote Access Controller 9 (iDRAC9) Enterprise for remote server administration

For details, see Software components.

The Openshift Container Platform architecture also makes use of:

- The Docker runtime to build, ship, and run containerized applications
- Kubernetes to orchestrate and manage containerized applications
- Etcd, a key-value store for the Red Hat OpenShift Container Platform cluster
- Open vSwitch (OVS) to provide software-defined networking (SDN)-specific functions in the Red Hat OpenShift Container Platform environment
- HAProxy for routing and load balancing purposes
- Keepalived, which uses the VRRP protocol to automate failover to assist with virtual IP management of HAProxy instances

The following figure shows the logical environment of the Red Hat OpenShift Container Platform system architecture:

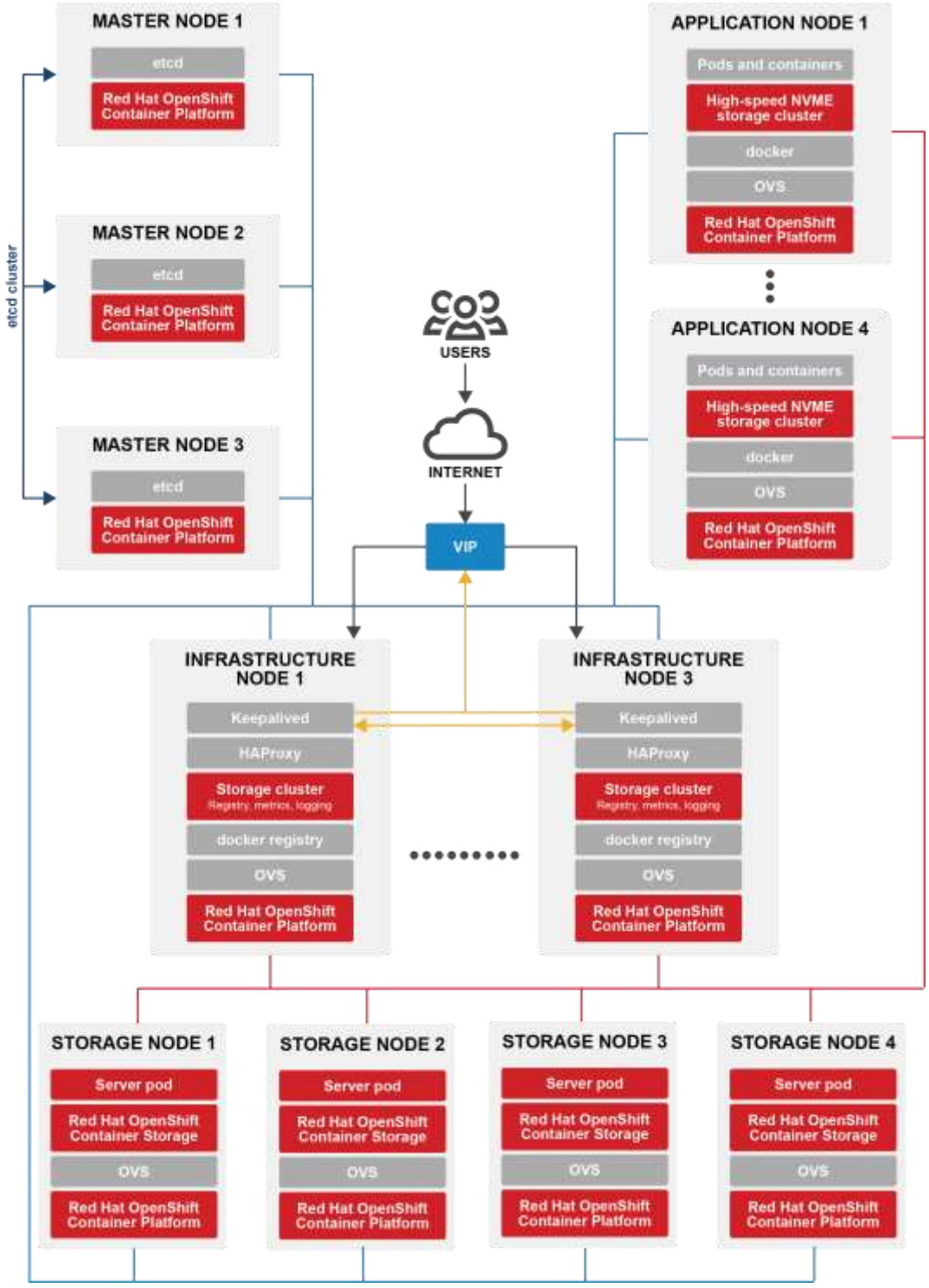

Figure 2. Red Hat OpenShift Container Platform system architecture - logical environment

Software components

Openshift Container Platform includes the following components:

- **Red Hat Enterprise Linux**—The stable, reliable operating system for all nodes in the system. Security controls and functions increase system protection, while integrated management capabilities help you maintain your systems quickly and easily. For more information, see Red Hat Enterprise Linux.
- **Red Hat OpenShift Container Platform**—An enterprise-grade application environment that supports all aspects of the development process in one consistent platform across multiple infrastructures. Red Hat OpenShift Container Platform integrates the architecture, processes, platforms, and services needed to build applications. Using the Kubernetes container orchestrator, Red Hat Openshift Container Platform manages applications across clusters of systems that are running the Docker container runtime. For more information, see Red Hat OpenShift.
- **Red Hat OpenShift Container Storage**—Based on Red Hat Gluster storage, Red Hat OpenShift Container Storage provides high-performance, persistent storage for container environments with granular control over every component of the storage landscape. Through integration with Red Hat OpenShift Container Platform, Red Hat OpenShift Container Storage stores application data as well as the container registry, logs, and metrics. With Red Hat OpenShift Container Storage, both application and storage containers can reside on the same server, which helps reduce costs and simplify management. For more information, see Red Hat OpenShift Container Storage.
- **Red Hat Ansible Automation**—Simple, agentless IT automation technology that can be used to provision resources, deploy applications, and configure and manage infrastructure. Ansible Automation provides a visual dashboard that enables you to manage all aspects of your automated tasks, including role-based access control (RBAC), job scheduling, and real-time job status updates. The automation used in this reference architecture is based largely on Ansible Playbooks because of their simplicity and extensibility. For more information, see Red Hat Ansible Automation.

Supported software versions

The following table shows the recommended minimum software component versions for Openshift Container Platform:

Table 1. Minimum supported software component versions

Component	Version
Red Hat Enterprise Linux	7.6
Red Hat OpenShift Container Platform	3.11
Red Hat OpenShift Container Storage	3.11
Docker Engine	1.13.1
Ansible	2.6.15
Etcd	3.2.22

Component	Version
Open vSwitch	2.9.0
Keepalived	1.2.13

Node types

This reference architecture uses five node types: bastion, master, infrastructure, application, and storage.

- **Bastion node**—The bastion node serves as the main deployment and management server for the Red Hat OpenShift cluster. It is used as the login node for cluster administrators to perform the system deployment and management operations across separate zones in the cluster at the same time—for example, the bastion node runs DNS services and a HTTP/TFTP for server provisioning. The bastion node also provides external connectivity such as code fetching and container image pulling for internal application hosts. The bastion node is an integral part of the solution platform. After the cluster installation is started with openshift-ansible, those individual nodes all pull containers from the Red Hat registry at registry.redhat.io. For more information, see the Red Hat Container Catalog.
- **Master nodes**—Master nodes perform control functions for the entire cluster environment. They are responsible for the creation, scheduling, and management of all objects specific to Red Hat OpenShift, including the API, controller management, and scheduler capabilities. The master nodes host the API server, Etcd, the Controller Manager, and the HAProxy. Etcd requires fully redundant deployment with load balancing. The API Server is managed by the HAProxy. A single instance of the Controller Manager server is elected as the customer lead at all times, while Etcd requires an odd number of hosts for quorum.

  To achieve a low-latency link between etcd and Red Hat OpenShift master nodes, you can install an etcd key-value store on the master nodes. Dell EMC recommends that you run both Red Hat OpenShift masters and etcd in highly available environments. Do this by running at least three Red Hat OpenShift master nodes in conjunction with an external active-passive load balancer and etcd clustering functions.

  The web console is started as part of the master node as a OpenShift binary with static content required for running the web console. The OpenShift Container Platform web console is a UI managed from a web browser. The web console provides a set of user options such as visualizing, browsing, and managing the contents of projects.
- **Infrastructure nodes**—The infrastructure nodes execute a range of services, including an internal service, the OpenShift Container registry, the HAProxy router, and the Heketi service. The OpenShift Container registry stores application images in the form of containers. HAProxy provides routing functions for Red Hat OpenShift applications and supports HTTP(S) traffic and Transport Layer Security (TLS)-enabled traffic via Server Name Indication (SNI). Heketi provides the OpenShift Container Storage volume life cycle management for

configuring persistent storage via REST. You can deploy additional applications and services on infrastructure nodes to meet your specific requirements.

For more information, see the OpenShift Container Registry.

- **Application nodes**—Application nodes run containerized workloads. They contain a single binary of Red Hat OpenShift node components and are used by Red Hat OpenShift master nodes to schedule and control containers.
- **Storage nodes**—The storage nodes provide persistent storage for the environment. Storage classes can create persistent volumes manually or automatically. These nodes can be configured to run in converged mode, providing both storage and compute services, which means that they are capable of running user-facing, containerized applications.

Storage class

This reference architecture uses one storage class consisting of NVMe drives only. Although the use of mixed drive types is possible, do not mix drive types in an OpenShift Container Storage cluster. In other words, each storage node must have only one drive type and a sufficient number of drives to create a full storage cluster with the same drive type. Three is the minimum supported number of drives but we recommend using more than four. While you can have one drive in a storage node, one drive is not suitable with bare-metal servers. Note that the recommendation for four drives relates to the nodes, not drives specifically.

You can disable workload scheduling capabilities if storage performance is expected to be critical. See the 'Managing Nodes' chapter in the OpenShift Container Platform 3.11 documentation.

The following figure shows the Openshift Container Platform node roles.

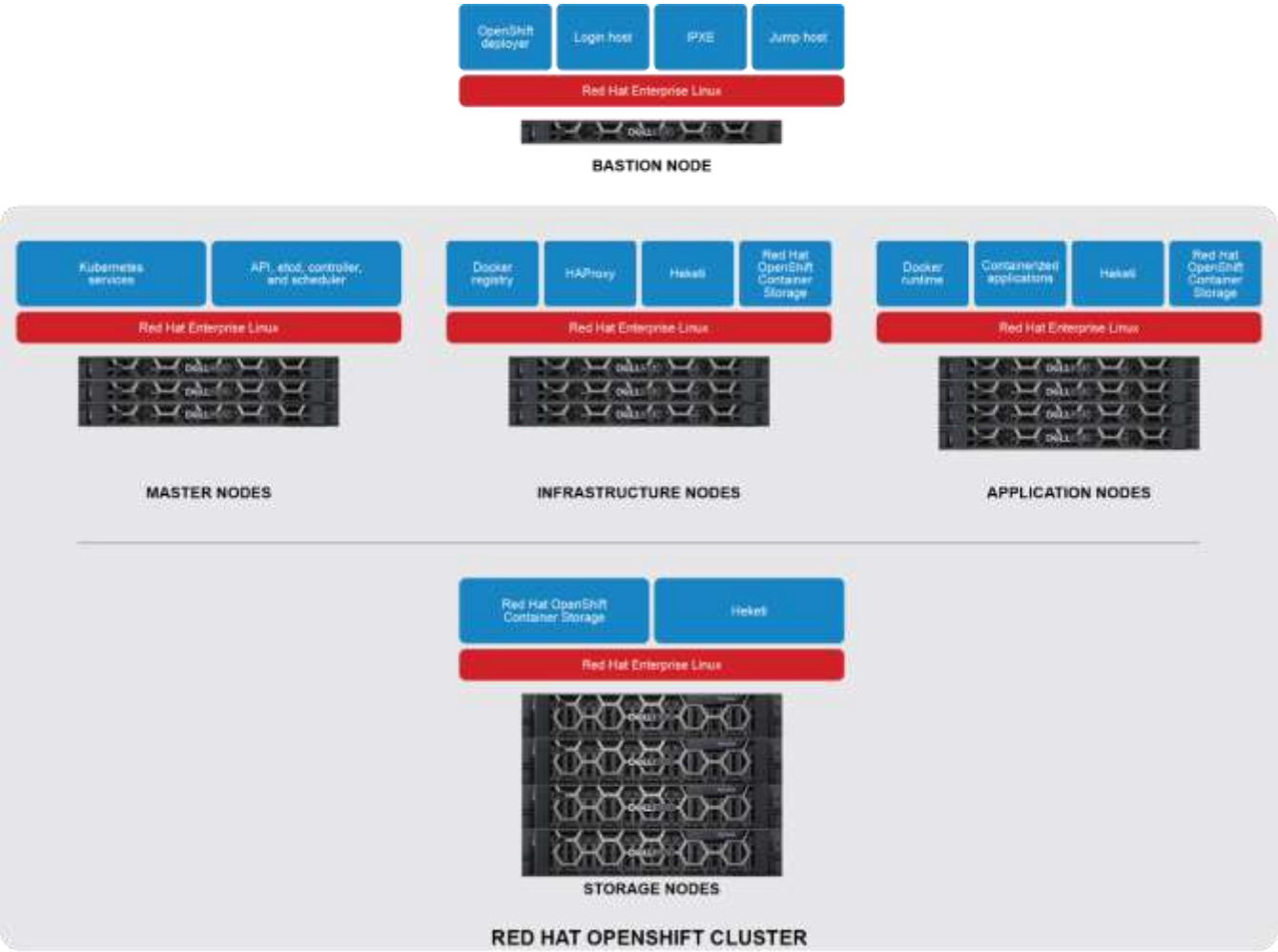

Figure 3. Red Hat OpenShift Container Platform node roles

Integrated cluster storage

Red Hat OpenShift Container Storage can be configured to provide persistent storage and dynamic provisioning for Openshift Container Platform. Gluster storage can be containerized within Openshift Container Platform (converged mode) or non-containerized on its own nodes (independent mode).

Converged mode

With converged mode, Red Hat Gluster Storage runs containerized directly on Openshift Container Platform nodes. This mode allows for compute and storage instances to be scheduled and run from the same set of hardware. Converged mode is available starting with the Red Hat Gluster Storage 3.1 update 3. For more information, see Red Hat OpenShift Container Storage for OpenShift Container Platform.

The following figure shows the converged mode of cluster storage:

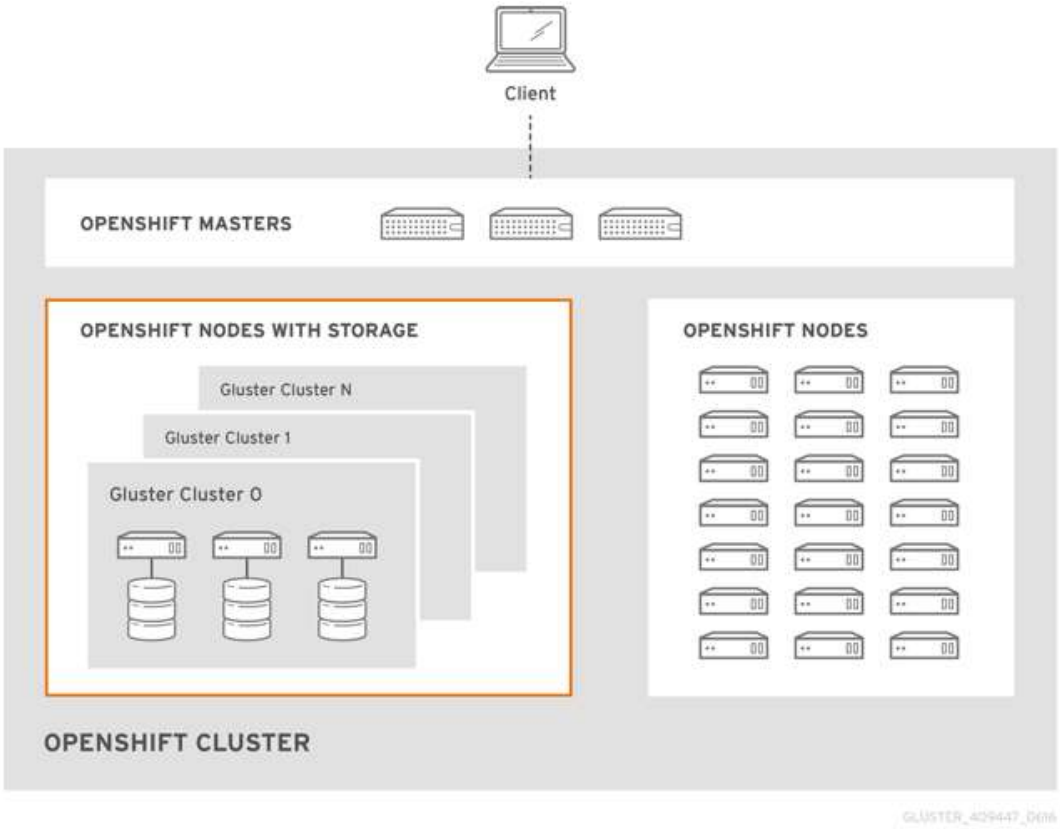

Figure 4. Cluster storage - converged mode

GlusterFS volumes

GlusterFS volumes present a POSIX-compliant file system. These volumes consist of one or more bricks across one or more nodes in their cluster. A brick is a directory on a given storage node, typically the mount point for a block storage device. GlusterFS handles distribution and replication of files across a given volume's bricks for that volume's configuration.

Dell EMC recommends using heketi for most common volume management operations such as create, delete, and resize. Openshift Container Platform expects heketi to be present when using the GlusterFS provisioner. By default, heketi creates volumes that are three-way replicas; that is, volumes where each file has three copies across three different nodes. Therefore it is required that any Red Hat Gluster storage clusters that will be used by heketi have at least three nodes available.

GlusterFS volumes can be provisioned either statically or dynamically. Static provisioning is available with all configurations. Only converged mode and independent mode support dynamic provisioning.

Gluster-block volumes

The Gluster-block volumes are volumes that you can mount over iSCSI. You can create a file on an existing GlusterFS volume and then present that file as a block device over an iSCSI target. Such GlusterFS volumes are called block-hosting volumes.

Because Gluster-block volumes are consumed as iSCSI targets, these volumes can only be mounted by one node/client at a time. This limitation is in contrast to GlusterFS volumes, which can be mounted by multiple nodes/clients. As files on the backend, Gluster-block volumes allow for operations that are typically costly on GlusterFS volumes (such as metadata lookups) to be converted to operations that are typically much faster on GlusterFS volumes, such as reads and writes. This leads to potentially substantial performance improvements for certain workloads.

Although GlusterFS supports other modes of presentation for the storage it can manage, Red Hat at this time recommends using Gluster-block volumes only for OpenShift Logging and OpenShift Metrics storage. While application workloads can use this integrated storage starting from Red Hat OpenShift Container Platform v3.11, this use has not been validated in production at the time of writing. As a best practice, we recommend using dedicated enterprise-grade storage technologies for application use and locating this storage outside of the OpenShift Container Ecosystem cluster.

For more information, see Complete Example Using GlusterFS.

Supported storage technologies

When it is deployed on-premises, the Red Hat OpenShift Container Platform supports several types of storage, providing you with a significant choice. The following table shows the storage choices that are available:

Table 1 - Supported on-premises storage types for PVs

Volume plug-in	ReadWriteOnce	ReadOnly	ReadWriteMany
Ceph RBD	✓	✓	-
Fibre Channel	✓	✓	-
GlusterFS	✓	✓	✓
HostPath	✓	-	-
iSCSI	✓	✓	-
NFS	✓	✓	✓
Openstack Cinder	✓	-	-
Local	✓	-	-

Storage is required for internal infrastructure use and to store and preserve application data. Be sure to take your application requirements into account when you select your storage platform technology.

Note: A later Ready Architecture for Red Hat OpenShift will provide design and deployment information for Dell EMC storage products and services.

Storage recommendations

The following table summarizes the recommended configurable storage technologies for the Red Hat OpenShift Container Platform cluster application.

Table 2. Storage configuration for the Openshift Container Platform cluster

Storage type	ROX	RWX	Registry	Scaled registry[1]	Metrics	Logging	Apps
Block	✔	No	Configurable	Not configurable	Recommended	Recommended	Recommended
File	✔	✔	Configurable	Configurable	Configurable	Configurable	Recommended
Object	✔	✔	Recommended	Recommended	Not configurable	Not configurable	Not configurable

[1] An OpenShift Container Platform registry in which three or more pod replicas are running

Chapter 5 System Architecture

This chapter presents the following topics:

Overview

The following figure shows the hardware used at the rack level in this reference architecture. In addition to servers and switches, the rack includes power distribution units (PDUs) and the cables that are required for network connectivity.

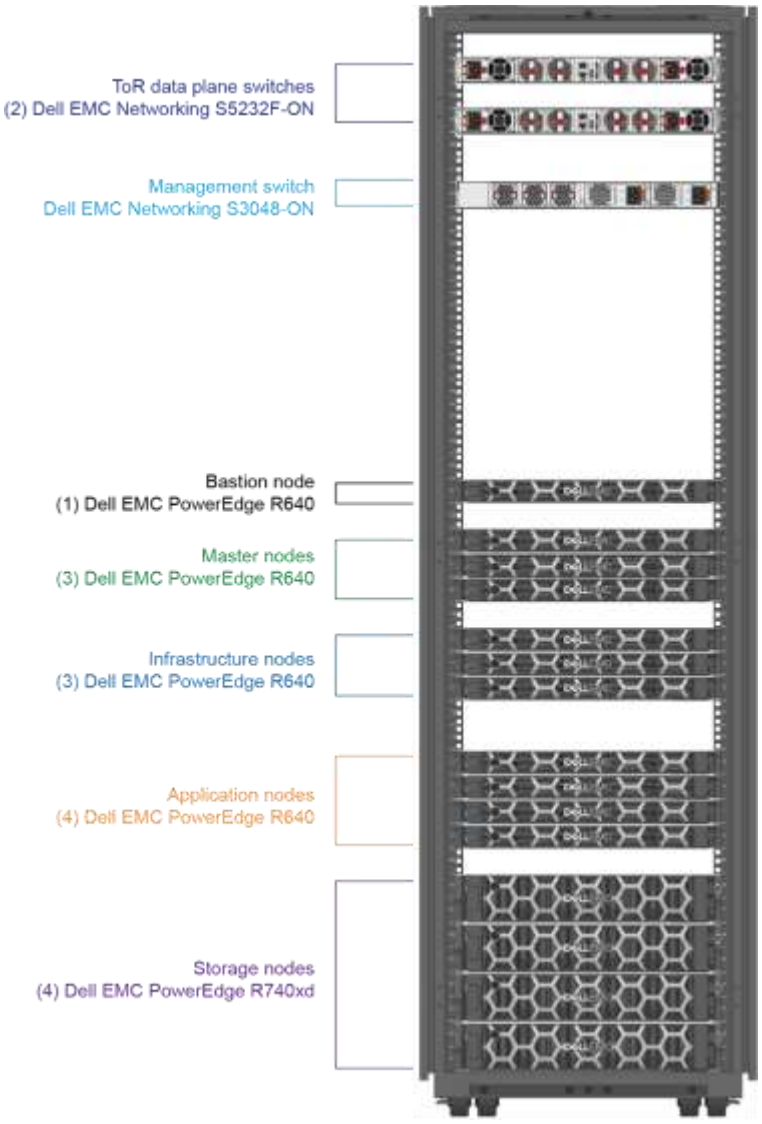

Figure 5. Ready Architecture for OpenShift Platform rack diagram

Hardware platform design and ordering information

DevOps teams who manage large numbers of servers typically purchase servers of almost identical configuration to simplify the management of cluster nodes and the movement of servers from rack to rack, from cluster to cluster, or within a cluster. Dell EMC recognizes the benefits of uniformity in server node configuration.

This ready architecture uses a single base configuration for the PowerEdge R640 nodes that are used in the cluster role nodes: master, infrastructure, application/workload, and storage. Infrastructure nodes require the addition of two NVMe drives to this baseline configuration.

We applied the same concept to the configuration for the larger 2U PowerEdge R740xd storage node servers. The following table shows the PowerEdge Server R640 baseline configurations used in the reference architecture:

Important: At the time of ordering, new SKUs will be added and local current SKUs will replace those shown in the table.

Table 3. PowerEdge R640 baseline server configuration

Qty	SKU	Description
1	210-AKWU	PowerEdge R640 Server
1	329-BDKC	PowerEdge R640 motherboard
1	321-BCQQ	2.5 in. chassis with up to 10 hard drives, 8 NVMe drives, and 3 PCIe slots, 2 CPU only
1	338-BLMH	Intel Xeon Gold 6138 2.0 G, 20C/40T, 10.4 GT/s, 27 M Cache, Turbo, HT (125W) DDR4-2666
1	374-BBOC	Intel Xeon Gold 6138 2.0 G, 20C/40T, 10.4GT/s, 27 M Cache, Turbo, HT (125W) DDR4-2666
1	370-ABWE	DIMM blanks for system with 2 processors
2	412-AAIQ	Standard 1U Heatsink
1	370-ADNU	2,666 MT/s RDIMMs
1	370-AAIP	Performance-optimized
12	370-ADNF	32 GB RDIMM 2,666 MT/s dual rank
1	405-AAJU	HBA330 12 Gbps SAS HBA Controller (NON-RAID), minicard
1	403-BBPZ	BOSS controller card + with 2 M.2 Sticks 240 G (RAID 1), LP
1	385-BBKT	iDRAC9, Enterprise
1	379-BCQV	iDRAC Group Manager, enabled
1	379-BCSG	iDRAC, legacy password
1	385-BBLG	Static IP
1	330-BBGN	Riser Config 2, 3 x 16 LP
1	406-BBLG	Mellanox ConnectX-4 Lx Dual Port 25 GbE SFP 28 rNDC

Qty	SKU	Description
1	406-BBLD	Mellanox ConnectX-4 Lx dual port 25 GbE SFP28 network adapter, low profile
1	429-AAIQ	No internal optical drive
1	384-BBQI	8 performance fans for the R640 server
1	450-ADWS	Dual, hot-plug, redundant power supply (1+1), 750W
2	492-BBDH	C13 to C14, PDU Style, 12 AMP, 2 ft. (.6m) power cord, North America
1	800-BBDM	UEFI BIOS boot mode with GPT partition
1	770-BBBC	ReadyRails sliding rails without cable management arm
1	366-0193	Std Bios setting power management* - maximum performance
2	400-AWLD	Dell 1.6TB, NVMe, Mixed Use Express Flash, 2.5 SFF Drive, U.2, P4610 with Carrier, CK

The following table shows the PowerEdge Server R740xd baseline configurations used in the reference architecture:

Important: At the time of ordering, new SKUs will be added and local current SKUs will replace those shown in the table.

Table 4. PowerEdge R740xd baseline server configuration

Qty	SKU	Description
1	210-AKZR	PowerEdge R740XD Server
1	329-BDKH	PowerEdge R740/R740XD motherboard
1	321-BCRC	Chassis up to 24 x 2.5 in. hard drives including 12 NVME drives, 2 CPU configuration
1	338-BLMH	Intel Xeon Gold 6138 2.0G, 20C/40T, 10.4GT/s, 27 M Cache, Turbo, HT (125W) DDR4-2666
1	374-BBOC	Intel Xeon Gold 6138 2.0G, 20C/40T, 10.4GT/s, 27 M Cache, Turbo, HT (125W) DDR4-2666
1	412-AAIQ	Standard 1U Heatsink
1	370-ADNU	2,666 MT/s RDIMMs
12	370-ADNF	32GB RDIMM 2,666 MT/s dual rank
1	780-BCDI	No RAID
1	405-AANK	HBA330 controller adapter, low profile
1	365-0354	CFI, standard option not selected

Qty	SKU	Description
1	403-BBPT	BOSS controller card + with 2 M.2 Sticks 240G (RAID 1), FH
1	385-BBKT	iDRAC9, Enterprise
1	379-BCQV	iDRAC Group Manager, enabled
1	379-BCSG	iDRAC, legacy password
1	385-BBLG	Static IP
1	330-BBHD	Riser Config 6, 5 x 8, 3 x1 6 slots
1	406-BBLG	Mellanox ConnectX-4 Lx Dual Port 25 GbE SFP28 rNDC
1	406-BBLE	Mellanox ConnectX-4 Lx Dual Port 25 GbE SFP28 network adapter
1	384-BBPZ	6 performance fans for R740/740XD
1	450-ADWM	Dual, hot-plug, redundant power supply (1+1), 1100W
1	492-BBDH	C13 to C14, PDU Style, 12 AMP, 2 ft (.6m) power cord, North America
1	325-BCHU	PowerEdge 2U standard bezel
1	800-BBDM	UEFI BIOS Boot Mode with GPT partition
1	770-BBBQ	ReadyRails sliding rails without cable management arm
1	366-0193	Std Bios setting power management - maximum performance
Select ONE of the rows below		
24	Check part at order time	800 GB, 1.92 TB, or 3.84 TB SSD SAS Mix Use 12 Gbps 512e 2.5 in. hot-plug AG drive, 3 DWPD, 4380 TBW, CK
12	Check part at order time	Dell 1.6 TB, 3.2 TB or 6.4 TB, NVMe, mixed use express flash, 2.5 SFF drive, U.2, P4610 with Carrier, CK

The following table describes the hardware configuration that is required to build the cluster design that we used for our validation work:

Table 5. Cluster configuration

Node name	Qty.	Configuration
Bastion	1	Dell EMC PowerEdge R640 baseline server configuration
Master	3	Dell EMC PowerEdge R640 baseline server configuration
Infrastructure	3	Dell EMC PowerEdge R640 baseline server configuration Add to each node: 2 x Dell 1.6 TB, NVMe, mixed use express flash, 2.5 in. SFF drive, U.2, P4610 with Carrier, CK (SKU 400-AWLD)
Application	4 or more	Dell EMC PowerEdge R640 baseline server configuration

Node name	Qty.	Configuration
Storage	4 or more	Dell EMC PowerEdge R740xd baseline server configuration

The following table provides additional cluster configuration information:

Table 6. Additional cluster configuration reference information

Quantity	Description	Dell EMC reference
1*	Rack enclosure: APC AR3300 NetShelter SZ 42U	APC AR3300 NetShelter SZ 42U
1**	Management switch: Dell EMC Networking S3048-ON	Dell Networking S-Series 1GbE switches
2	Data switch: Dell EMC Networking S5248F-ON	Dell Networking S-Series 25/40/50/100GbE switches
11	Bastion, infrastructure, master, application nodes: Dell EMC PowerEdge R640	PowerEdge R640 Rack Server
4	Storage nodes: Dell EMC PowerEdge R740xd	PowerEdge R740xd Rack Server
	Dell EMC PowerEdge R640**	PowerEdge R640 Rack Server
2-4*	Power distribution unit APC metered rack PDU 17.2 kW	APC metered rack PDU 17.2 kW

*Rack enclosures and power distribution units are site-specific. The physical dimensions and power requirements must be reviewed during a site survey.

**The configuration of the Dell EMC PowerEdge Server R640 equipped with 10 x NVMe SSD drives serving as storage nodes has been validated.

Validated hardware configuration options

We validated Ready Architecture for Red Hat OpenShift Contain Platform using different node configurations. The design described in this document took into account overall cluster performance across multiple dimensions, including the time taken to deploy the cluster, workload launch characteristics, and application runtime performance. Because time and productivity have value and a cost, the design of a container platform must consider risks as well as time opportunity costs. The following guidance is provided for possible specification changes.

Selecting a processor

The Intel Xeon Gold processor family is optimized to provide performance, advanced reliability, and hardware-enhanced security for demanding compute, network, and storage workloads. Up to 22 cores and 6 memory channels deliver high performance and scalability for compute- and memory-intensive workloads, while 48 lanes of PCIe 3.0 bandwidth and throughput provide support for demanding input/output (I/O)-intensive workloads. A near-zero encryption overhead enables higher performance on all secure data transactions. For more information, see Intel Xeon Gold Processors.

Dell EMC recommends selecting Intel Xeon processors in the range of the Intel Gold series CPUs of the 6126 to 6152 models.

When selecting a processor, customers must take account of the following:

- Sufficient core count to ensure adequate performance of all workload operations
- Power consumption and sizing of the power supply units (PSUs) in each server
- Ability to dissipate heat. During validation work with high core-count high TDP processors, the thermal delta (air discharge temperature minus air intake temperature) across a server was recorded at 65°F. A high temperature air discharge (egress) from the server might lead to a premature component or system failure.

Per-node memory configuration

For memory configuration, refer to Red Hat OpenShift architectural guidance and your own observations from running your workloads on the Red Hat Openshift Container Platform. The Dell EMC engineering team chose 192, 384, or 768 GB RAM as representing the best choices based on memory usage, DIMM module capacity for the current cost, and likely obsolescence during the server life cycle. We chose a mid-range memory configuration of 384 GB RAM to ensure that the memory for each CPU has multiples of three banks of DIMM slots populated to ensure maximum memory access cycle speed. You might choose to alter the memory configuration to meet your budgetary constraints and your operating needs.

Disk drive capacities

The performance of disk drives significantly limits the performance of the many aspects of OpenShift cluster deployment and operation. The Dell EMC engineering team validated deployment and operation of Red Hat OpenShift Container Platform with magnetic storage drives (spinners), SATA SSD drives, SAS SSD drives, and NVMe SSD drives. The selection of all NVMe SSD drives was based on a comparison of cost per GB of capacity divided by observed performance criteria such as deployment time for the cluster, application deployment characteristics, and application performance. There are no

universal guidelines, but over time you will gain insight into the capacities that will best enable you to meet your requirements.

Optionally, you can deploy the cluster with only HDD disk drives. This configuration has been tested and shown to have few adverse performance consequences.

Network controllers and switches

When selecting the switches to include in the Red Hat OpenShift Container Platform cluster infrastructure, give careful consideration to the overall balance of I/O pathways within your compute and storage nodes, the network switches, and the NICs for your cluster. Key considerations include:

- **HDD drives**—These drives have lower throughput per drive. It is acceptable to use 10 Gb ethernet for this configuration.
- **SATA/SAS SSD drives**—These drives have high I/O capability. SATA SSDs drives operate at approximately four times the I/O level of a spinning HDD. SAS SSDs operate at up to ten times the I/O level of a spinning HDD. With SSD drives, it is highly recommended to configure your servers with 25 GbE.
- **NVMe SSD drives**—These drives have very high I/O capability, up to three times the I/O rate of SAS SSDs. We chose to populate each node with 4 x 25 GbE NICs to provide additional I/O bandwidth.

True high availability (HA) fail-safe design demands that each NIC is duplicated, permitting a pair of ports to be split across two physically separated switches. A pair of Dell EMC Networking S5248F-ON switches provides 96 x 25 GbE ports, enough for a total of approximately 20 servers. This switch is cost-effective for a compact cluster. While it is possible to add an additional pair of S5248F-ON switches to scale the cluster to a full rack, you might want to consider using Dell EMC Networking S5232F-ON switches for a larger cluster.

The Dell EMC Networking S5232F-ON switch provides 32 x 100 Gbe ports. When used with a 4-way QSFP28 to SFP28, a pair of these switches provides up to 256 x 25 GbE end-points, more than enough for a rack full of servers in the cluster before more complex network topologies are required.

Switch choice for OpenShift in an NFV environment

There is a quest for very low latency in all aspects of container ecosystem design for application deployment in NFV-centric data centers. This requirement means that you must give particular attention to selecting low-latency components throughout the OpenShift cluster. Dell EMC highly recommends using only NVMe drives, NFV-centric versions of Intel CPUs, and, at a minimum, the Dell EMC Networking switch model S5232F-ON. Consult the Dell EMC Service Provider organization support team for specific guidance.

Power configuration

For HA operation, each server must be equipped with redundant power supplies. Each rack is configured as pairs of Power Distribution Units (PDUs). For consistency, connect all right-most 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. The following figure shows an example.

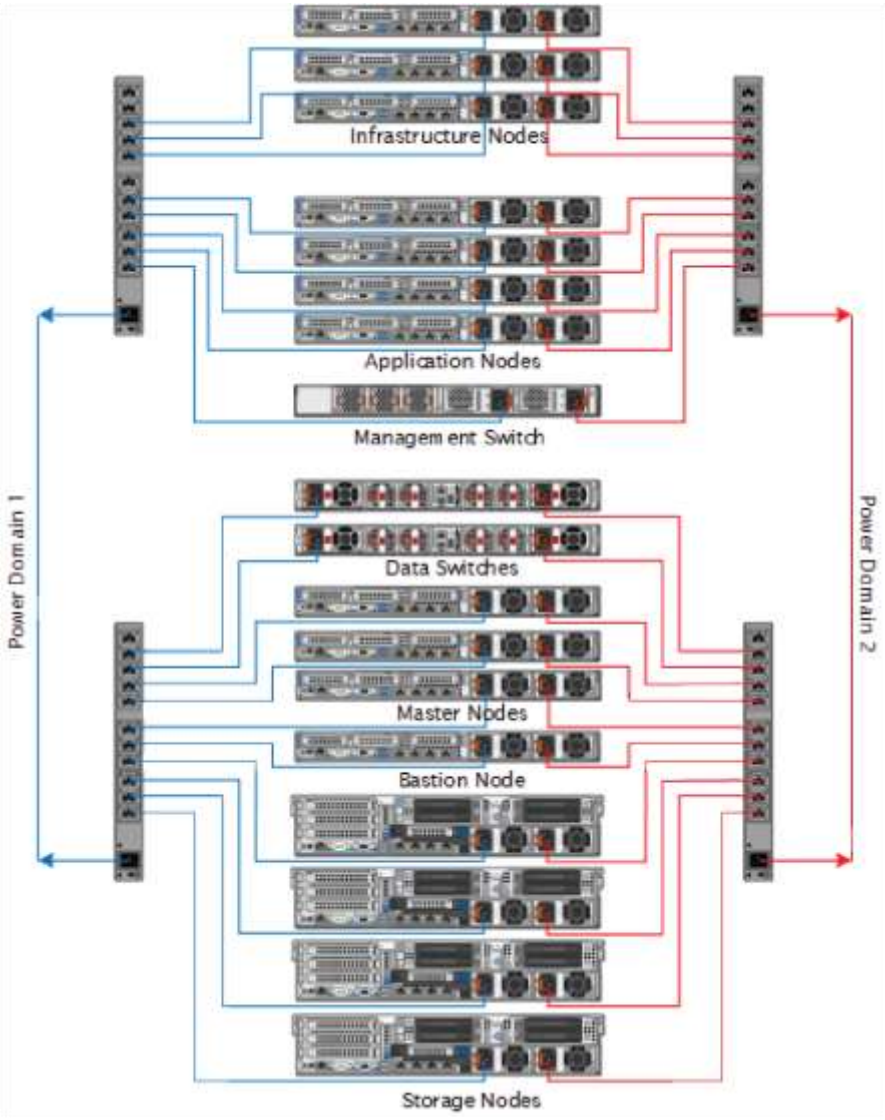

Figure 6. PSU to PDU power template

Cluster scaling guidance

Cluster scaling and sizing initiatives must account for the Openshift Container Platform cluster limits. For more information, see the Red Hat Cluster Limits documentation. The following table summarizes the limits:

Table 7. Openshift Container Platform cluster limits

Limit type	OpenShift v3.11 limit
Number of nodes	2,000

Limit type	OpenShift v3.11 limit
Number of pods	150,000
Number of pods per node	250
Number of pods per core	There is no default value. The maximum supported value is the number of pods per node.
Number of namespaces	10,000
Number of builds: Pipeline Strategy	10,000 (default pod RAM 512Mi)
Number of pods per namespace	3,000
Number of services	10,000
Number of services per namespace	5,000
Number of back-ends per service	5,000
Number of deployments per namespace	2,000

The guidelines in the following table apply to the deployment and scaling of Red Hat OpenShift Container Storage:

Table 8. OpenShift container storage limits

Limit type	OpenShift Container Storage v3.11 limit
Persistent volumes backed by the file interface	1,000
Persistent volumes backed by block-based storage	300
Storage cluster size	4 nodes at a minimum

When you design and specify the configuration of workload/application nodes, Dell recommends carefully sizing the cluster for a minimum number of nodes as a reference point for the total cluster cost. As a second data point relating to cost, you can size the nodes using the most cost-effective CPU and memory configuration. The optimal node configuration in many workload deployment sites lies somewhere between these options. As the cluster size and number of nodes increase, industry practice favors keeping the cluster size below four racks to avoid unnecessary complexity in network management and switch topology.

Even though deploying extremely large OpenShift container ecosystem clusters is possible, there are some disadvantages to managing such designs. In most container deployment sites, the largest cluster does not span more than two or three racks. As the need for container ecosystem clusters becomes established, it is often helpful to deploy two or more clusters so that workloads can be migrated from one cluster to another. An additional benefit of multiple clusters is the ability to release a cluster for reconfiguration or redeployment of the container infrastructure, although this is rare.

Chapter 6 Networking

This chapter presents the following topics:

Overview

This chapter describes the networking requirements for the Ready Architecture for OpenShift Container Platform and how to configure the Dell Networking switches that are used for an OpenShift deployment at various scales.

Network architecture

For the Dell EMC reference architecture, we provisioned each server with 4 x 25 Gbe NIC ports that are cross-wired to the network switches. A 25 GbE network is the preferred primary fabric for internode communication. Two S5248F-ON or S5232F-ON switches provide data layer communication, while one S3048-ON switch is used for OOB management.

The following figure shows the network architecture.

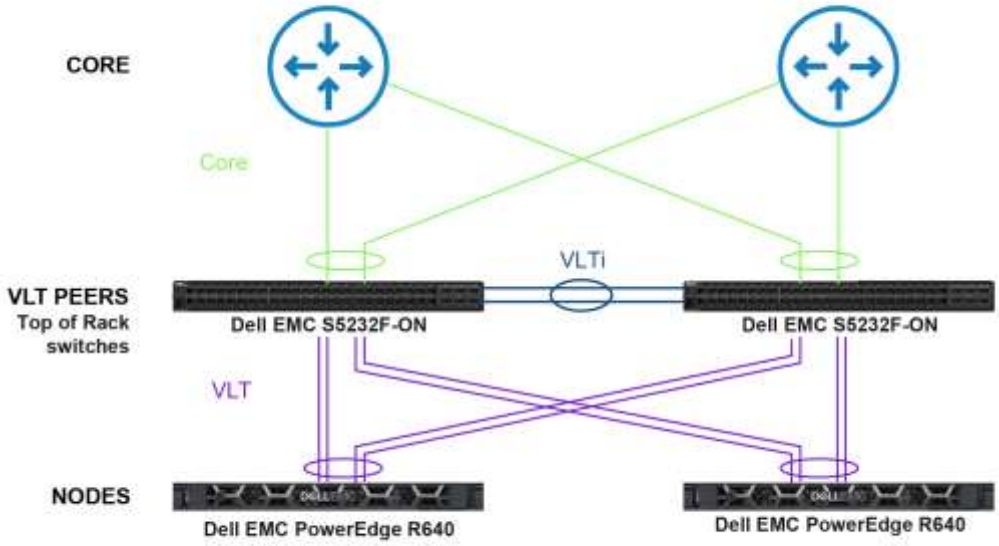

Figure 7. Virtual Link Trunking (VLT)

This reference architecture is designed to deliver maximum availability and enough network bandwidth that storage performance and compute performance are not limited by available network bandwidth. Each server has 4 x 25 Gbe ports. Two ports are provided on the Network Daughter Card (NDC) and two more from a dual port NIC. One port from each dual port NIC goes to ToR switch A, and the other port from each NIC is wired to ToR Switch B.

Virtual Link Trunking (VLT) is a layer-2 link aggregation protocol between end devices connected to two switches. VLT offers a redundant, load-balancing connection to the core network in a loop-free environment and eliminates the need to use a spanning tree protocol. VLT allows link connectivity between a server and the network over two different

switches. VLT can also be used for uplinks between access or distribution switches and core switches.

This reference architecture has three logical networks:

- **External**—The external network is used for the public API, the Red Hat OpenShift Container Platform web interface, and exposed applications.
- **Internal**—The internal network is the primary, non-routable network for cluster management, internode communication, and server provisioning using PXE and HTTP. DNS and DHCP services also reside on this network to provide deployment functionality. Communication with the internet is provided by network address translation (NAT) configured on the bastion node.
- **OOB/Intelligent Platform Management Interface (IPMI)**—The OOB/IPMI network is a secured and isolated network for switch and server hardware management, including access to the iDRAC9 module and Serial-over-LAN .

The following figure shows the network components of the Red Hat OpenShift Container Platform and their logical architecture.

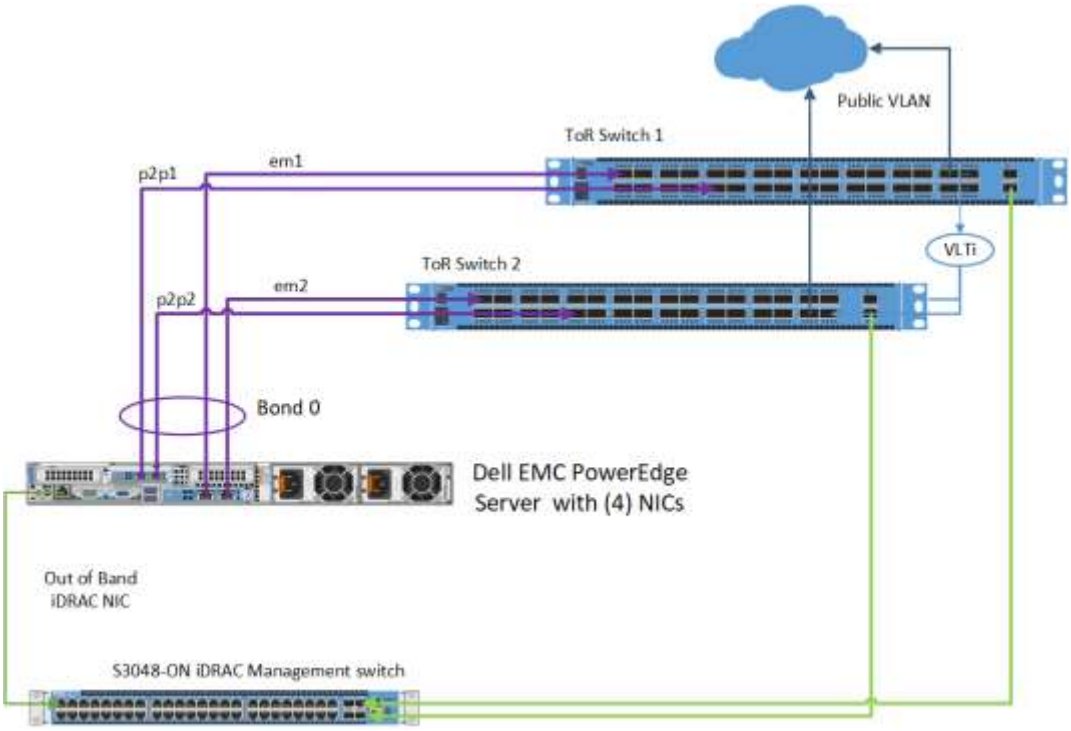

Figure 8. Red Hat OpenShift Container Platform logical/physical network connectivity

The four physical 25 GbE interfaces are bundled together in an 802.3ad link aggregation. The external network is presented as an 802.1Q VLAN tagged network from the upstream Dell EMC S5248F-ON switch pair, and the internal interface is presented as the untagged VLAN on the same interface pair to facilitate PxE booting. LACP fallback is enabled so that when the nodes are booting by PxE using the UEFI PxE module they can communicate with the provisioning system. This interface pair is labeled “Bond 0” in the diagram.

The recommended bonding options for the Linux are:

```
BONDING_OPTS=”mode=802.3ad miimon=100 xmit_has_policy=layer3+4
lacp_rate=1
```

All Red Hat OpenShift nodes are logically connected through the internal network, which means that they are all on the same layer-2 broadcast domain. In addition, Open vSwitch creates its own network for Red Hat OpenShift pod-to-pod communication. The OpenShift ovs-multitenant plugin allows only pods with the same project namespace to communicate. Keepalived manages a virtual IP address on three infrastructure hosts for external access to the Red Hat OpenShift Container Platform web console and applications.

If necessary, you can use an enterprise external load balancer (F5, NGINX, or other) as the ingress point for both the web console and OpenShift Router. An external load balancer may be needed for edge request routing and load balancing for applications deployed on the OpenShift cluster. The external load balancer might be already available in the target deployment location. For more information, see the Multiple Masters example in OpenShift Example Inventory Files.

Recommended NIC to switch cabling

To secure the maximum fail-safe high availability (HA) configuration, cable the NIC ports across HA ToR Switches, as shown in the following table:

Table 9. Recommended cabling

Server	ToR Switch-1	ToR Switch-2	Port channel
bastion	em1 p2p1	em2 p2p2	1
master1	em1 p2p1	em2 p2p2	2
master2	em1 psp1	em2 p2p2	3
master3	em1 psp1	em2 p2p2	4
infra1	em1 psp1	em2 p2p2	5
infra2	em1 psp1	em2 p2p2	6
infra3	em1 p2p1	em2 p2p2	7
app1	em1 p2p1	em2 p2p2	8
app2	em1 p2p1	em2 p2p2	9
app3	em1 p2p1	em2 p2p2	10
app4	em1 p2p1	em2 p2p2	11

Server	ToR Switch-1	ToR Switch-2	Port channel
stor1	em1 p7p1	em2 p7p2	12
stor2	em1 p7p1	em2 p7p2	13
stor3	em1 p7p1	em2 p7p2	14
stor4	em1 p7p1	em2 p7p2	15

Configuring Dell Networking switches

Overview

This section describes how to configure the Dell Networking switches used for an OpenShift deployment at various scales.

Single rack networking

The network architecture employs a VLT connection between the two ToR switches. In a non-VLT environment, redundancy requires idle equipment, which drives up infrastructure costs and increases risks. In a VLT environment, all paths are active, adding immediate value and throughput while still protecting against hardware failures. VLT technology allows a server or bridge to uplink a physical trunk into more than one Dell Networking switch by treating the uplink as one logical trunk. A VLT-connected pair of switches acts as a single switch to a connecting bridge or server. Both links from the bridge network can actively forward and receive traffic. VLT provides a replacement for Spanning Tree Protocol (STP)-based 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 LAG interfaces
- Active/Active design for seamless operations during maintenance events

Dell EMC Networking S5248F-ON switch

Each Dell Networking S5248F-ON switch provides six 100 GbE uplink ports.

- The Virtual Link Trunk Interconnect (VLTi) configuration in this architecture uses two 100 GbE ports from each ToR switch.
- The remaining four 100 GbE ports allow for high speed connectivity to spine switches or directly to the data center core network infrastructure. They can also be used to extend connectivity to other racks.

The remaining 48 ports of 25 GbE are used for server connectivity. An Openshift Container Platform cluster with up to 20 server nodes can easily be accommodated using a pair of S5248F-ON switches. Expansion of an Openshift Container Platform single-rack

cluster beyond 20 nodes is managed in one of two ways: add a pair of aggregated S5248F-ON switches or select Dell EMC Networking S5232F-ON switches.

Dell EMC Networking S5232F-ON switch

The S5232F-ON switch also supports ONIE for zero-touch installation of network operating systems. In addition to 100 GbE Spine/Leaf deployments, the S5232F-ON switch can also be used in high density deployments using breakout cables to achieve up to 128 x 10 GbE or 128 x 25 GbE ports.

- The VLTi configuration in this architecture uses two 100 GbE ports from each ToR switch.
- 100 GbE ports can also be used for high speed connectivity to spine switches or directly to the data center core network infrastructure. They can also be used to extend connectivity to other racks.

The following figure shows the switch connectivity.

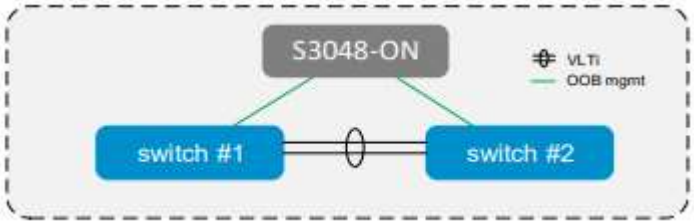

Figure 9. Switch connectivity

Configuring VLT

The VLT configuration involves the following high-level steps:

1. Enable Spanning Tree on the VLT peer switches. Spanning tree is enabled by default and is recommended to prevent loops in VLT domain. RPVST+ (the default) and RSTP modes are supported on VLT ports.
2. Create a VLT Domain and configure the VLT interconnect (VLTi).
3. Configure the VLT Priority, VLT MAC Address, and VLT Backup Link.
4. Configure the LAG for the connected device.
5. Verify and monitor the status of VLT and mismatches by using appropriate OS10 `show` commands.

Scaling the network solution

This section describes how to scale out the Dell EMC Ready Architecture for Red Hat OpenShift Container Platform v3.11.

Container solutions can be scaled by adding multiple application and storage nodes. Your solution might contain multiple racks of servers. To create a non-blocking fabric to meet the needs of the micro-services data traffic, we used a leaf-spine network.

Leaf-spine overview

The following concepts apply to layer 2 and layer 3 leaf-spine topologies:

- Each leaf switch connects to every spine switch in the topology.

- Servers, storage arrays, edge routers, and similar devices always connect to leaf switches, but never to spines.

Dell Networking uses two leaf switches at the top of each rack, configured as a VLT pair. VLT allows all connections to be active while also providing fault tolerance. As administrators add racks to the data center, two leaf switches configured for VLT 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. You can increase the bandwidth of the fabric by adding connections between leaves and spines as long as the spine layer has capacity for the additional connections.

Layer 3 leaf-spine

In a layer 3 leaf-spine network, traffic is routed between leaves and spines. The layer 3/layer 2 boundary is at the leaf switches. Spine switches are never connected to each other in a layer 3 topology. 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 leaves, as shown in the following figure.

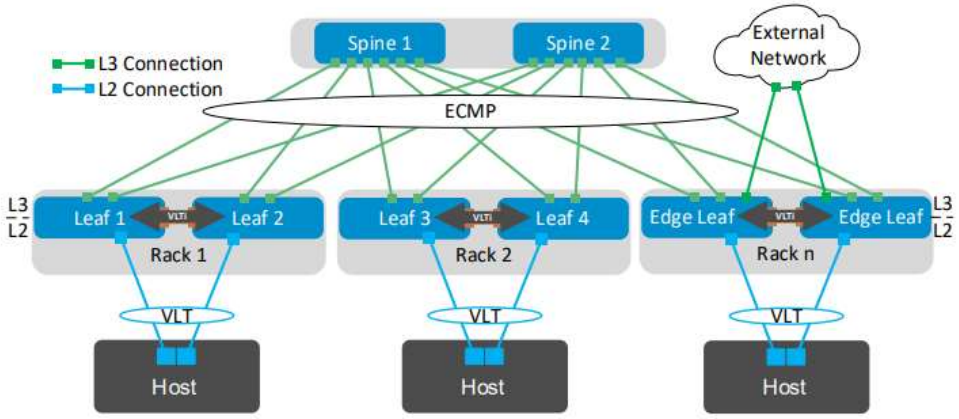

Figure 10. Leaf-spine network configuration

Chapter 7 References

This chapter presents the following topics:

Dell EMC documentation

The following Dell EMC documentation provides additional and relevant information. Access to these documents depends on your login credentials. If you do not have access to a document, contact your Dell EMC representative.

- OpenShift Container Platform Info Hub for Ready Solutions

Red Hat documentation

For additional information from Red Hat, see:

- *Red Hat OpenShift Container Platform*
- *Red Hat OpenShift Container Storage*