Dell EMC Ready Architecture
for Red Hat OpenStack Platform

Architecture Guide
Version 16.1

DELLTechnologies

Dell EMC Telco Service Provider Solutions

Contents

List of Figures

List of Tables

Trademarks

Intel® and Xeon® are registered trademarks of Intel Corporation. Oracle® and Java® are registered trademarks of Oracle Corporation and/or its affiliates.

Mellanox, the Mellanox logo, ConnectX, CORE-Direct, and GPUDirect are registered trademarks of Mellanox Technologies, Ltd. Mellanox Multi-Host is a trademark of Mellanox Technologies, Ltd. All rights reserved.

DISCLAIMER: The OpenStack® Word Mark and OpenStack Logo are either registered trademarks/ service marks or trademarks/service marks of the OpenStack Foundation, in the United States and other countries, and are used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation or the OpenStack community.

Notes, cautions, and warnings

A **Note** indicates important information that helps you make better use of your system.

A **Caution** indicates potential damage to hardware or loss of data if instructions are not followed.

A **Warning** indicates a potential for property damage, personal injury, or death.

This document is for informational purposes only and may contain typographical errors and technical inaccuracies. The content is provided as is, without express or implied warranties of any kind.

Chapter

1

Overview

Topics:

- *Dell EMC Ready Architecture Guide 16.1*
- *Hardware options*
- *Networking and network services*
- *JetPack automation toolkit*
- *OpenStack Architecture*

An OpenStack® based cloud is now a common need by many organizations and Dell Technologies with Red Hat have worked together to build a jointly engineered and validated architecture that details software, hardware, and integration points of all solution components. The architecture provides prescriptive guidance and recommendations for:

- Hardware design
  - Infrastructure nodes
  - Compute nodes
  - Storage nodes
  - Hyper-Converged Infrastructure (HCI) nodes (Compute & Storage combined)
  - Distributed Compute Nodes (DCN)
- Network design
- Software layout
- Offers suggestion for other system configurations

Dell EMC Ready Architecture Guide 16.1

This Architecture Guide is built on Red Hat OpenStack Platform 16.1 which is based on the 20th OpenStack release codename *Train*. Red Hat OpenStack Platform 16.1 is a containerized version offering greater scalability, resiliency and user experience.

> **Note:** Please see *https://www.redhat.com/en/technologies/linux-platforms/openstack-platform* for more information.

The Red Hat OpenStack Platform provides the foundation to build a private or public Infrastructure-as-a-Service (IaaS) cloud on Red Hat Enterprise Linux. It offers a massively scalable, fault-tolerant platform for the development of cloud-enabled workloads or Telco Cloud for Service Providers.

It is packaged so that available physical hardware can be turned into a private, public, or hybrid cloud platform including:

- Fully distributed object storage
- Persistent block-level storage
- Virtual-machine provisioning engine and image storage
- Authentication and authorization mechanism
- Integrated networking

The Red Hat OpenStack Platform is implemented by a collection of interacting services that control its computing, storage, and networking resources.

New features

- Added support for Dell EMC PowerEdge XE2420 edge servers

  > **Note:** Please see *https://www.delltechnologies.com/en-us/servers/specialty-servers/poweredge-xe-servers.htm* for more information.
- Added Red Hat OpenStack Platform OVS off-loading using Mellanox ConnectX-5 network adapters
- Added automation support using JetPack for Distributed Compute Nodes (DCN) using multi stacks
- Support for the latest release of Red Hat OpenStack Platform 16.1 including the latest updates
- Support for RHEL 8.2 including the latest updates

Key benefits

The Dell EMC Ready Architecture for Red Hat OpenStack Platform offers several benefits to help service providers and high-end enterprises rapidly implement Dell EMC hardware and Red Hat OpenStack Platform software:

- Ready-to-use solution: The ready architecture has been fully engineered, validated, tested in Dell EMC laboratories and documented by Dell EMC. This decreases your investment and deployment risk, and it enables faster deployment time.
- Long lifecycle deployment: Dell EMC PowerEdge R-Series and Dell EMC PowerEdge XE2420 servers, recommended in the architecture, include long-life Intel® Xeon® processors which reduces your investment risk and protects your investment for the long-term.
- World-class professional services: The ready architecture includes Dell EMC professional services that spans consulting, deployment, and design support to guide your deployment needs.
- Customizable solution: The architecture is prescriptive, but it can be customized to address each customer's unique virtual network function (vNF) or other workload requirements.
- Co-engineered and Integrated: OpenStack depends upon Linux for performance, security, hardware enablement, networking, storage, and other primary services. The Red Hat OpenStack Platform delivers an OpenStack distribution with the proven performance, stability, and scalability of RHEL 8.2 enabling you to focus on delivering the services your customers want, instead of focusing on the underlying operating platform.

- Deploy with confidence, as the Red Hat OpenStack Platform provides hardened and stable branch releases of OpenStack and Linux. The Red Hat OpenStack Platform is a long life release product supported by Red Hat for a four (4) year "production phase" life cycle, well beyond the six-month release cycle of unsupported, community OpenStack. Red Hat OpenStack Platform life cycle support policies can be found at *https://access.redhat.com/support/policy/updates/openstack/platform*
- Take advantage of broad application support. Red Hat Enterprise Linux, running as guest virtual machines, provides a stable application development platform with a broad set of ISV certifications. You can therefore rapidly build and deploy your cloud applications.
- Avoid vendor lock-in by moving to open technologies, while maintaining your existing infrastructure investments.
- Benefit from the world's largest partner ecosystem: Red Hat has assembled the world's largest ecosystem of certified partners for OpenStack compute, storage, networking, ISV software, and services for Red Hat OpenStack Platform deployments. This ensures the same level of broad support and compatibility that customers enjoy today in the Red Hat Enterprise Linux ecosystem.
- Upgrade of Red Hat OpenStack Director-based installations.
- Bring security to the cloud. Rely upon the SELinux military-grade security and container technologies of Red Hat Enterprise Linux to prevent intrusions and protect your data, when running in public or private clouds. For more information regarding security in the Red Hat OpenStack Platform 16.1 release, see *https://access.redhat.com/documentation/en-us/red_hat_openstack_platform/16.1/html/ security_and_hardening_guide/index*. Additional information pertaining to Red Hat vulnerabilities can be found at *https://access.redhat.com/security/*.

Hardware options

To reduce time spent on specifying hardware for an initial system, this Architecture Guide offers a full solution using validated Dell EMC PowerEdge server hardware designed to allow a wide range of configuration options, including optimized configurations for:

- Compute nodes (Both Core and DCN)
- Hyper-Converged Infrastructure nodes
- Infrastructure nodes
- Storage nodes

Dell EMC recommends starting with OpenStack software using components from this Architecture Guide - Version 16.1 because the hardware and operations processes comprise a flexible foundation upon which to expand as your cloud deployment grows, so your investment is protected.

As noted throughout this Architecture Guide - Version 16.1, Dell EMC constantly adds capabilities to expand this offering, and other hardware may be available.

Networking and network services

Network configuration is based upon using the Neutron-based options supported by the Red Hat OpenStack Platform code base, and does not rely upon third-party drivers. This reference configuration is based upon the Neutron networking services using the ML2 drivers for Open vSwitch with the `VLAN` option.

Networking includes:

- Core and layered networking capabilities
- Network Function Virtualization (NFV)
- 25GbE and 100GbE networking
- NIC bonding
- Redundant trunking top-of-rack (ToR) switches into core routers

This enables the Dell EMC Ready Architecture for Red Hat OpenStack Platform to operate in a full production environment.

See *Network architecture* on page 27 for guidelines. Detailed designs are available through Dell EMC consulting services.

JetPack automation toolkit

Dell EMC has updated its opensourced JetPack automation toolkit[1], which is an innovative automation package that is used to configure the infrastructure hardware and OpenStack software in a fully automated fashion. The toolkit includes programs from Dell EMC that work with Red Hat OpenStack Director to automate the deployment process, not only saving time, but also ensuring the process is reliable and repeatable.

The JetPack automation toolkit was used to deploy Red Hat OpenStack Platform 16.1 on the documented hardware in this Architecture Guide. The JetPack automation toolkit is used by Dell EMC, Red Hat professional service teams, and system integrator partners, and is available on GitHub to customers who prefer to do self-deployment.

This release uses the concept of profiles. There are two different validated profiles, CSP or xSP, that can be used for a deployment. The CSP profile is designed for Telecommunications Service Providers, Cable TV Operators, Satellite TV, Internet Service Providers, etc. whereas the xSP profile is designed for Business & IT Services Providers such as Hosting Service Providers, Cloud Service Providers, Software-as-a-Service/Platform-as-a-Service Providers, Application Hosting Service Providers and Private Managed Cloud Service Providers.

OpenStack Architecture

While OpenStack has many configurations and capabilities, the primary components for the Dell EMC Ready Architecture for Red Hat OpenStack Platform 16.1 is defined as a containerized version.

Note: For a complete overview of the OpenStack software, visit *Red Hat OpenStack Platform* and the *OpenStack Project*.

---

[1] The toolkit can be found at *https://github.com/dsp-jetpack/JetPack/tree/JS-16.1*

Chapter

2

BIOS and firmware compatibility

Topics:

- *Tested BIOS and firmware*

This chapter documents the versions of BIOS and firmware that are used to create this Dell EMC Ready Architecture for Red Hat OpenStack Platform.

Tested BIOS and firmware

Table 1: Dell EMC PowerEdge R640/Dell EMC PowerEdge R740xd tested BIOS and firmware versions on page 15 and *Table 2: Dell EMC PowerEdge XE2420 tested BIOS and firmware versions* on page 15 tables lists the server BIOS and firmware versions that were tested for the Dell EMC Ready Architecture for Red Hat OpenStack Platform.

Table 3: Dell EMC SC, Unity and PowerMax tested software and firmware versions on page 16 lists the Dell EMC Storage Center, SC series storage software and Dell EMC Unity firmware versions that were tested for the Dell EMC Ready Architecture for Red Hat OpenStack Platform.

Table 4: Dell EMC Network Switches tested firmware versions on page 16 lists the default S3048-ON, and optional S4048-ON management switch firmware versions that were tested for the Dell EMC Ready Architecture for Red Hat OpenStack Platform.

The leaf switches tested in this release are the S5232F-ON firmware and the S5224F-ON firmware

⚠ **CAUTION:** The versions listed below are what were used during the development of this Ready Architecture. Newer versions may be available since the time of this release. Ensure that the firmware on all servers, storage devices, and switches are at minimum up to date with the versions listed below. Otherwise, unexpected results may occur.

Table 1: Dell EMC PowerEdge R640/Dell EMC PowerEdge R740xd tested BIOS and firmware versions

Product	Version
BIOS	2.8.1
iDRAC with Lifecycle controller	4.22.00.00
Intel® XXV710 NIC	19.5.12
Mellanox ConnectX-4	14.26.60.00
Mellanox ConnectX-5	16.26.60.00
PERC H730P Mini controller (Dell EMC PowerEdge R640)	25.5.6.0009
PERC H740P Mini controller (Dell EMC PowerEdge R640)	50.9.4-3025
HBA330 Mini (Dell EMC PowerEdge R740xd)	16.17.01.00
BOSS-S1 (Dell EMC PowerEdge R740xd)	2.5.13.3024

Table 2: Dell EMC PowerEdge XE2420 tested BIOS and firmware versions

Product	Version
BIOS	1.1.7
iDRAC with Lifecycle controller	4.00.129.00
Intel® XXV710 NIC	19.5.12
PERC H740P Adapter	50.9.4-3025

Table 3: Dell EMC SC, Unity and PowerMax tested software and firmware versions

Product	Version
Dell EMC SC storage center software	2016 R2 Build 16.2.1.228
SC series storage firmware	6.6.11.9
Dell EMC Unity 380F firmware	5.0.0.0.5.116
Dell EMC PowerMax 2000 firmware	5978.221.221

Table 4: Dell EMC Network Switches tested firmware versions

Product	Version
S3048-ON firmware	Dell EMC Networking OS10 Enterprise OS Version: 10.5.1.4 Build Version: 10.5.1.4.249
S4048-ON firmware (optional)	Dell EMC Networking OS10 Enterprise OS version: 10.5.1.4 Build version: 10.5.1.4.249
S5232F-ON firmware	Dell EMC Networking OS10 Enterprise OS version: 10.5.1.4 Build version: 10.5.1.4.249
S5224F-ON firmware	Dell EMC Networking OS10 Enterprise OS version: 10.5.1.4 Build version: 10.5.1.4.249

Chapter

3

Server options

Topics:

- *Dell EMC PowerEdge XE2420 server*
- *Dell EMC PowerEdge R640 server*
- *Dell EMC PowerEdge R740xd servers*

The base validated solution supports the Dell EMC PowerEdge XE2420, Dell EMC PowerEdge R640, and Dell EMC PowerEdge R740xd server lines.

Note: Please contact your Dell EMC sales representative for detailed parts lists.

Dell EMC PowerEdge XE2420 server

The Dell EMC PowerEdge XE2420 is a configurable, dual-socket 2U rack server that delivers powerful 2S performance in a short-depth form-factor. With a scalable rack option, it is ideal for low-latency, large storage edge applications. Its performance can be further boosted by its support of up to 4 accelerators. A wide and flexible range of storage options (SSD, NVMe) and large capacity up to 92TB* gives it the flexibility to tackle a diverse set of demanding workloads at the edge.

With the Dell EMC PowerEdge XE2420 is a specialty edge server that is specifically engineered to deliver powerful performance for harsh environments. It is a dual-socket, 2U, short-depth, front accessible server that is designed to support demanding edge applications such as streaming analytics, manufacturing logistics, 5G cell processing.

Dell EMC PowerEdge R640 server

The Dell EMC PowerEdge R640 is the ideal dual-socket, 1U platform for dense scale-out cloud computing. The scalable business architecture of the Dell EMC PowerEdge R640 is designed to maximize application performance and provide the flexibility to optimize configurations based on the application and use case.

With the Dell EMC PowerEdge R640 you can create an NVMe cache pool and use either 2.5" or 3.5" drives for data storage. Combined with up to 24 DIMM's, 12 of which can be NVDIMM's, you have the resources to create the optimum configuration to maximize application performance in only a 1U chassis. This can simplify and speed up deployments of the Red Hat OpenStack Platform.

Dell EMC PowerEdge R740xd servers

The Dell EMC PowerEdge R740xd delivers a perfect balance between storage scalability and performance. The 2U two-socket platform is ideal for software defined storage. The R740xd versatility is highlighted with the ability to mix any drive type to create the optimum configuration of SSD and HDD for either performance, capacity or both.

The Dell EMC PowerEdge R740xd is the platform of choice for Red Hat Ceph storage for this Architecture Guide - Version 16.1. This platform has also been validated for compute and HCI roles.

Chapter

4

Configuring PowerEdge R-Series hardware

Topics:

- *Configuring the SAH node*
- *Configuring Overcloud nodes*

This section describes manually configuring PowerEdge R-Series server hardware for the Dell EMC Ready Architecture for Red Hat OpenStack Platform:

- IPMI configuration
- BIOS configuration
- RAID configuration

Configuring the SAH node

The Solution Admin Host (SAH) is a physical server that hosts one VM, the Red Hat OpenStack Director as the Undercloud.

The SAH BIOS configuration must be configured prior to OpenStack deployment. Refer to *SAH BIOS specification* on page 20

> **Note:** The SAH node must have the Boot Mode set to BIOS Boot instead of UEFI Boot prior to running the Open Source hardware configuration toolkit (OS-HCTK) tool. All other nodes will be configured to use UEFI Boot by OpenStack Ironic.

The SAH is configured using the Open Source hardware configuration toolkit (OS-HCTK) tool.

Please contact your Dell EMC Sales or Professional Services representative for access to the toolkit.

Open Source hardware configuration toolkit

The Open Source hardware configuration toolkit (OS-HCTK) is a configuration utility with sample scripts and configuration files that is used to automate the setup and configuration of BIOS and RAID settings for Dell EMC servers used for OpenStack open source software solutions.

> **Note:** The OS-HCTK ISO is run **only** on the system that will be configured as the SAH node. All other systems will be configured by OpenStack Ironic.

The OS-HCTK allows a user to create a USB key from which a Dell EMC PowerEdge R640 can be booted and the BIOS and RAID settings can be applied.

Topics discussed include:

- *iDRAC settings* on page 20
- *SAH BIOS specification* on page 20

iDRAC settings

> **Note:** The following settings need to be set on the SAH nodes BIOS configuration prior to a deployment.

Table 5: iDRAC Specification for SAH Nodes on page 20 lists and describes iDRAC default `racadm` settings that will be set by the OS-HCTK.

Table 5: iDRAC Specification for SAH Nodes

Menu Choice	iDRAC Setting
iDRAC.IPMILan.Enable	Enabled
iDRAC.IPMILan.PrivLimit	4
iDRAC.IPv4.Enable	Enabled
iDRAC.Users.2.Enable	Enabled
iDRAC.Users.2.IpmiLanPrivilege	4
iDRAC.Users.2.Privilege	0x1ff
iDRAC.WebServer.Enable	Enabled

SAH BIOS specification

Table 6: SAH BIOS Specification on page 21 lists and describes the default BIOS settings for the SAH that will be set by the OS-HCTK.

Table 6: SAH BIOS Specification

Display Name	Attribute	Settings
Boot Mode	BootMode	BIOS
Boot Sequence Retry	BootSeqRetry	Enabled
DCU IP Prefetcher	DcuIpPrefetcher	Enabled
DCU Streamer Prefetcher	DcuStreamerPrefetcher	Enable
Logical Processor Idling	DynamicCoreAllocation	Disabled
Integrated RAID Controller	IntegratedRaid	Enabled
Internal SD Card	InternalSdCard	Off
I/OAT DMA Engine	IoatEngine	Enabled
Logical Processor	LogicalProc	Enabled
Memory Operating Mode	MemOpMode	OptimizerMode
System Memory Testing	MemTest	Disabled
Node Interleaving	NodeInterleave	Disabled
OS Watchdog Timer	OsWatchdogTimer	Disabled
Adjacent Cache Line Prefetch	ProcAdjCacheLine	Enabled
Number of Cores per Processor	ProcCores	all
Hardware Prefetcher	ProcHwPrefetcher	Enabled
CPU Power Management	ProcPwrPerf	MaxPerf
Turbo Mode	ProcTurboMode	Enabled
Virtualization Technology	ProcVirtualization	Enabled
CPU Interconnect Bus Speed	CpuInterconnectBusSpeed	MaxDataRate
SR-IOV Global Enable	SriovGlobalEnable	Enabled
System Profile	SysProfile	PerfOptimized

Configuring Overcloud nodes

Configuring server network settings

The Overcloud hardware configurations and the settings for the Intelligent Platform Management Interface, (IPMI) are done through Ironic which is included in the *JetPack automation toolkit* on page 13.

Configuring server network settings

1. Set the iDRAC IP address source:

    a. If the Overcloud nodes were ordered with the iDRACs configured for DHCP, or are currently configured for DHCP, then no further configuration is necessary.

    b. If you wish to use static IP addresses, then configure the Overcloud nodes' DRAC IP address, subnet mask, default gateway IP, and default VLAN (ID = *110*, if required) using the iDRAC GUI.

Chapter

5

Storage Options

OpenStack has several storage services, including:

*   Cinder
*   Glance
*   Manila
*   Ephemeral
*   Swift [2]

Together these services provide virtual machines (VMs) with block, image, file share, and object storage. In turn, the services employ block, file share, and object storage subsystems. Since the service design has a mechanism to replace some or all of the implementation of these services, this solution can provide alternate implementations of these services that better serve your needs.

---

[2] Available through a Ceph cluster with the Swift API enabled or through a custom Professional Services engagement.

Storage options overview

Cinder virtualizes storage enabling VMs to use persistent block storage through Nova. OpenStack consumers should write data that must exist beyond the lifecycle of the guest to Cinder volumes. The volume can be accessed afterwards by a different guest.

Glance provides images to VMs. Generally, the images are block devices containing DVDs or virtual machines. VMs can be booted from these images or have the images attached to them. Glance storage now can use Red Hat Ceph storage, Dell EMC Unity or Dell EMC SC series storage.

Note: Dell EMC Unity or SC series storage provides the Glance support through Cinder.

Dell EMC Manila driver framework (EMCShareDriver) delivers a shared filesystem in OpenStack. The plugin-based Dell EMC Manila driver design is compatible with various plugins to control Dell EMC storage products.

The Dell EMC Unity plugin manages the Dell EMC Unity Storage System for shared filesystems.

The Dell EMC Unity driver is a REST API. Unity Storage System Manila backends are a one-to-one managed storage system. Each Manila backend configures a Unity Storage System. See *https:// docs.openstack.org/manila/train/admin/emc_unity_driver.html*

Swift provides an object storage interface to VMs and other OpenStack consumers. Unlike block storage where the guest is provided a block device of a given format and is accessible within the cluster, object storage is not provided through the guest. Object storage is generally implemented as a HTTP/HTTPS-based service through a web server. Client implementations within the guest or external OpenStack clients would interact with Swift without any configuration required of the guest other than providing the requisite network access. For example, a VM within OpenStack can put data into Swift, and later external clients could pull that data for additional processing.

Note: "Swift" in this document refers to the Swift interfaces, not the Swift implementation, of the protocol.

As with other OpenStack services, there are client and server components for each storage service. The server component can be modified to use a particular type of storage rather than the default. For example, Cinder uses local disks as the storage back-end by default. The Dell EMC Ready Architecture for Red Hat OpenStack Platform modifies the default configuration for these services.

All virtual machines will need a virtual drive that is used for the OS. Two options are available:

- Ephemeral disks
- Boot from volume or snapshot, hosted on Red Hat Ceph storage, Dell EMC Unity storage or SC series storage arrays.

Ephemeral disks are virtual drives that are created when a VM is created, and destroyed when the VM is removed. The virtual drives can be stored on the local drives of the Nova host or on a shared file system, such as Ceph Rados Block Device *(RBD)*. During the planning process, decisions can be made to place ephemeral on local or shared storage, it is recommended that a shared backend is used which will allow live migration.

Boot from volume/snapshot will use one of the Cinder backends.

The Dell EMC Ready Architecture for Red Hat OpenStack Platform includes alternate implementations of Cinder that enable the cluster to fit many needs. Cinder has been validated using each of the back-ends independently, and multi back-ends utilizing multiple Storage back-ends consisting of two or all of:

- *Local storage* on page 24
- *Red Hat Ceph storage* on page 24
- *Optional Dell EMC Unity storage*

Local storage

When using local storage, each Compute node will host the ephemeral volumes associated with each virtual machine. Cinder will utilize LVMs that are shared by NFS for independent volumes, utilizing the local storage subsystem. With the hardware configuration for the Compute nodes see *Table 23: Compute node Dell EMC PowerEdge R640* on page 66 using the eight (8) 600GB disks in a RAID 10, there will be approximately 2 TB of storage available.

Red Hat Ceph storage

The Dell EMC Ready Architecture for Red Hat OpenStack Platform includes Red Hat Ceph storage, which is a scale-out, distributed, software-defined storage system. Red Hat Ceph storage is used as backend storage for Nova, Cinder and Glance. Storage nodes run the Red Hat Ceph storage software, and Compute and Controller nodes run the Red Hat Ceph storage block client.

Red Hat Ceph storage also provides object storage for OpenStack VMs and other clients external to OpenStack. The object storage interface is an implementation of:

• The OpenStack Swift RESTful API (basic data access model)
• The Amazon S3 RESTful API

The object storage interface is provided by Red Hat Ceph storage RADOS Gateway software running on the Controller nodes. Client access to the object storage is distributed across all Controller nodes in order to provide HA and IO load balancing.

Red Hat Ceph storage is containerized in RHOSP16.1 (Mon, Mgr, Object Gateway, and Object Storage Daemon). Each OSD has an associated physical drive where the data is stored, and a journal where write operations are staged prior to being committed.

• When a client reads data from the Red Hat Ceph storage cluster the OSDs fetch the data directly from the drives.
• When a client writes data to the storage cluster the OSDs write the data to their journals prior to committing the data.

OSD journals can be located in a separate partition on the same physical drive where the data is stored, or they can be located on a separate high-performance drive, such as an SSD optimized for write-intensive workloads. For the Architecture Guide, a recommended ratio of one (1) SSD to four (4) hard disks is used to achieve optimal performance. It should be noted that as of this writing, using a greater ratio will result in server performance degradation.

In a cost-optimized solution, the greatest storage density is achieved by forgoing separate SSD journal drives, and populating every available physical drive bay with a high-capacity HDD.

In a throughput-optimized solution, a few drive bays can be populated with high performance SSDs that will host the journals for the OSD HDDs. For example, in an Dell EMC PowerEdge R740xd system with 20 drive bays available for Red Hat Ceph storage, 4 bays are used for SSD journal drives and 16 bays for HDD data drives. This is based upon the current industry guideline of a 4:1 HDD to SSD journal ratio.

Dell EMC Unity storage

One of the supported and validated Dell EMC storage options is:

• Dell EMC Unity storage which can provide Glance (Image), Cinder (Block) and Manila (Shared FileSystems) Storage support.

> **Note:** Other storage options are available upon request provided by Professional Services.

The Dell EMC Unity platforms accelerate application infrastructure with All-Flash unified storage platforms that simplify operations while reducing cost and datacenter footprint.

The completely refreshed, next generation, Dell EMC Unity XT family -- the no compromise midrange storage, built new from the ground up to deliver unified storage speed and efficiency and built for a multi-cloud world

- Dell EMC Unity 350F, 450F, 550F, 650F (This RA has validated the Unity 350F and the 550F platforms)
- Dell EMC Unity XT 380F,480F,680F,880F (This RA has validated the Unity XT 380F platform)

Please see *https://www.delltechnologies.com/en-us/storage/unity.htm* for more information about Dell EMC Unity storage.

Please see *https://access.redhat.com/ecosystem/software/3172821* for more information about Red Hat OpenStack Platform certifications for Unity.

Please see *https://docs.openstack.org/cinder/train/configuration/block-storage/drivers/dell-emc-unity-driver.html* for more information about Dell EMC Unity Cinder driver.

Please see *https://docs.openstack.org/manila/train/admin/emc_unity_driver.html* for more information about the Dell EMC Unity Manila driver.

Dell EMC SC series storage

Another one of the supported and validated Dell EMC storage options is:

- Dell EMC SC series storage which now supports Glance (Image) storage via cinder in addition to Cinder (Block) Storage

    Note: Other storage options are available upon request provided by Professional Services.

Please see *https://access.redhat.com/documentation/en-us/red_hat_openstack_platform/16.1/html/dell_storage_center_back_end_guide/index* for more information.

Please see *https://access.redhat.com/ecosystem/software/1610473* for more information about Red Hat OpenStack Platform certification for Dell EMC SC series storage

Dell EMC PowerMax storage

One of the supported and validated Dell EMC storage options is:

- Dell EMC PowerMax storage NVMe Data Storage which can provide Glance (Image) and Cinder (Block) Storage support.

    Note: Other storage options are available upon request provided by Professional Services.

With end-to-end NVMe, storage class memory (SCM) for persistent storage, real-time machine learning and up to 350GB per second, PowerMax features high-speed smarts to power your most critical workloads.

- VMAX2 Series: VMAX 40K, 20K, 10K
- VMAX3 Series: VMAX 400K, 200K, 100K
- VMAX All Flash Series: VMAX 250F, 450F and 850F
- PowerMax Series 2000 and 8000 (This RA has validated the PowerMax Series 2000 platforms)

Please see *https://www.delltechnologies.com/en-us/storage/powermax.htm* for more information about Dell EMC PowerMax storage NVMe Data Storage.

Please see *https://access.redhat.com/ecosystem/software/1572113* for more information about Red Hat OpenStack Platform certifications for Dell EMCVMAX and PowerMax.

Please see *https://docs.openstack.org/cinder/train/configuration/block-storage/drivers/dell-emc-powermax-driver.html* for more information about Dell EMC Unity Cinder driver.

Chapter

6

Network architecture

Topics:

- *Network architecture overview*
- *Infrastructure layouts*
- *Network components*

This Ready Architecture Guide supports consistency in rapid deployments through minimal network configuration.

Network architecture overview

The Dell EMC Ready Architecture for Red Hat OpenStack Platform uses 2 x S5232F-ON as Leaf switches and the S3048-ON (S4048-ON optional) as the management switch.

Infrastructure layouts

The network consists of the following major network infrastructure layouts:

- **Core Network Infrastructure** - The connectivity of aggregation switches to the core for external connectivity.
- **Data Network Infrastructure** - The server NICs, Leaf switches, and the aggregation switches.
- **Management Network Infrastructure** - The BMC management network, consisting of iDRAC ports and the out-of-band management ports of the switches, is aggregated into a 1-rack unit (RU) S3048-ON switch in one of the three racks in the cluster. This 1-RU switch in turn can connect to one of the aggregation or core switches to create a separate network with a separate VLAN.

Network components

The data network is primarily composed of the ToR and the aggregation switches. The following component blocks make up this network:

- *Server nodes* on page 28
- *Leaf Switches* on page 29
- *Spine switches* on page 29
- *Core*
- *Layer-2 and Layer-3 Switching* on page 29
- *VLANs* on page 30
- *Management network services* on page 30
- *Dell EMC OpenSwitch solution* on page 30

Server nodes

In order to create a highly-available solution, the network must be resilient to loss of a single network switch, network interface card (NIC) or bad cable. To achieve this, the network configuration uses bonding across the servers and switches.

There are several types (or modes) of bonding, but only one is recommended for the Solution. The OpenStack Controller, Compute nodes, Red Hat Ceph storage nodes, and Solution Admin Host can use:

- `802.3ad` or LACP (mode = 4)

  Note: Other modes, such as `balance-rr` (mode=0), `balance-xor` (mode=2), `broadcast` (mode=3), `balance-tlb` (mode=5), and `balance-alb` (mode=6), are not supported. Please check with your technician for current support status of `active-backup` (mode = 1).

All nodes' endpoints are terminated to switch ports that have been configured for LACP bonding mode across two S5232F-ON ToR switches for 25GbE/100GbE configured with a Virtual Link Trunking interconnect (VLTi) across them.

Please contact your Dell EMC sales representative for other viable options.

Table 7: Bonding nodes supported

	Bonding type
Node Type	**802.3ad (LACP mode 4)**
Solution Admin Host	Yes (solution default)
OpenStack Controller nodes	Yes (solution default)
OpenStack Compute nodes	Yes (solution default)
Red Hat Ceph storage nodes	Yes (solution default)
OpenStack Hyper-Converged Infrastructure nodes	Yes (solution default)

A single port is an option when bonding is not required. However, it is neither used nor validated in the Dell EMC Ready Architecture for Red Hat OpenStack Platform. The need to eliminate single points of failure is taken into consideration as part of the design, and this option has been eliminated wherever possible.

Please contact your Dell EMC sales representative for other cofigurations.

Leaf Switches

Dell EMC's recommended architecture uses VLT for HA between the two Leaf switches, which enables the servers to terminate their Link Aggregation Group (LAG) interfaces (or bonds) into two different switches instead of one. This configuration enables active-active bandwidth utilization and provides redundancy within the rack if one Leaf switch fails or requires maintenance. Dell EMC recommended Leaf switch is:

• 10/25/100GbE connectivity – S5232F-ON

The Leaf switches are responsible for providing the different network connections such as tenant networks, storage networks etc, between the different nodes of the OpenStack deployment such as compute, controller and storage nodes.

Spine switches

> **Note:** Please contact your Dell EMC sales representative for aggregation switch recommendation.

Layer-2 and Layer-3 Switching

The layer-2 and layer-3 boundaries are separated at the aggregation layer.

The Architecture Guide uses layer-2 as the reference up to the Leaf layer, which is why VLT is used on the Leaf switches.

The three network links - Provisioning, Storage, and Management - can have uplinks to a gateway device. The Provisioning network can use the Red Hat OpenStack Director as a proxy for pulling packages from a subscription server, or a gateway can be added. The Red Hat Ceph storage, Dell EMC Unity, SC series storage or the Dell EMC PowerMax storage arrays on the Storage network may need access:

• From metrics and monitoring tools
• To enable management and updates

> **Note:** For Dell EMC Unity Shared File Systems Storage (Manila), the Dell EMC Unity storage array requires additional access to networks. Options are External Network VLAN (Floating Network) or Internal Network VLAN for Tenants (Tenant Network) based on the use case. To support this functionality, administrators need to configure the Dell EMC Networking S5232F-ON or Dell EMC Networking S5224F-ON ethernet ports to allow access for the Storage (untagged) and Floating/ Tenant (tagged) VLANs.

There are many tools for Out of Band management, (OOB) for the iDRAC, which you can use after first adding the gateway to the network, and then updating the iDRAC.

The OpenStack Controllers are connected to a gateway device, usually a router or firewall. This device will handle routing for all networks external to the cluster. The required networks are:

• The floating IP range used by virtual machines
• A network for all external Public API and Graphical User Interface access

VLANs

This Architecture Guide implements at a minimum eight (8) separate Layer 2 VLANs:

• **External Network VLAN for Tenants**- Sets up a network that will support the *floating IPs* and default external gateway for tenants and virtual machines. This connection is through a router external to the cluster.
• **Internal Networks VLAN for Tenants** - Sets up the backend networks for Nova and the VMs to use.
• **Management/Out of Band (OOB) Network** - iDRAC connections can be routed to an external network. All OpenStack HA Controllers need direct access to this network for IPMI operations.
• **Private API Network VLAN** - Used for communication between OpenStack Controllers, the Red Hat OpenStack Director, and Compute nodes for Private API and cluster communications.
• **Provisioning Network VLAN** - Connects a NIC from all nodes into the fabric, used for setup and provisioning of the OpenStack servers.
• **Public API Network VLAN** - Sets up the network connection to a router that is external to the cluster. The network is used by the front-end network for routable traffic to individual VMs, access to the OpenStack API, RADOS Gateway, Red Hat Ceph Storage Dashboard, and the Horizon GUI. Depending upon the network configuration these networks may be either shared or routed, as needed. The Red Hat OpenStack Director requires access to the Public API Network.
• **Storage Clustering Network VLAN** - Used by all Storage nodes for replication and data checks (Red Hat Ceph storage clustering).
• **Storage Network VLAN** - Used by all nodes for the data plane reads/writes to communicate to OpenStack Storage; setup, and provisioning of the Red Hat Ceph storage cluster; and when included, the Dell EMC Unity storage or SC series storage arrays.
• **Tenant Tunnel Network VLAN** - Used by Tenants for encapsulated networks such as, GRE or VXLAN tunnels, in place of the **Internal Networks VLAN for Tenants**.

Management network services

The management network and the provisioning network for all the servers and switches aggregate into a Dell EMC Networking S3048-ON switch.

> **Note:** Both the default management switch, S3048-ON, and the optional S4048-ON switch is valid for use in the Dell EMC Ready Architecture for Red Hat OpenStack Platform Ready Architecture Guide Version 16.1.

The Management network services is used for several functions:

• The highly available software uses it to reboot and partition servers.
• An uplink to a router and an iDRAC configure a gateway monitoring the servers and gathering metrics.

> **Note:** Discussion of this topic is beyond the scope of this document.

Dell EMC OpenSwitch solution

In addition to the Dell EMC switch-based Architecture Guide, Dell EMC provides an open standard that enables you to choose other brands and configurations of switches for your OpenStack environment.

The following list of requirements will enable other brands of switches to properly operate with Dell EMC's required tools and configurations:

> **Note:** You are expected to ensure that the switches conform to these requirements, and that they are configured according to this Architecture Guide's guidelines. These requirements are the minimums needed for a successful deployment and should be treated as such. Performance of the solution is directly affected by the performance of the Network.

- Support for IEEE 802.1Q VLAN traffic and port tagging
- Support for using one untagged, and multiple tagged VLANs, on the same port
- Support for using bonded interfaces as a single interface for TFTP/DHCP booting.
- Ability to provide a minimum of 96 x 25GbEs Ethernet ports in a non-blocking configuration within the Provisioning VLAN

    - Configuration can be a single switch or a combination of stacked switches to meet the additional requirements
- The ability to create LAGs with a minimum of two physical links in each LAG
- If multiple switches are stacked:

    - The ability to create a LAG across stacked switches
    - Full-bisection bandwidth
    - Support for VLANs to be available across all switches in the stack
- 250,000 packets-per-second capability per switch
- A managed switch that supports SSH and serial line configuration
- SNMP v3 support

Chapter

7

Network Function Virtualization (NFV) support

Topics:

- *NUMA Optimization and CPU Pinning*
- *Hugepages*
- *OVS-DPDK*
- *SR-IOV*
- *SR-IOV OVS-Hardware-Offload with VF-LAG*
- *Distributed Virtual Router (DVR)*

This Architecture Guide supports the following Network Function Virtualization (NFV) features:

- NUMA
- Huge Pages
- OVS-DPDK
- SR-IOV
- SR-IOV with Open vSwitch (OVS) offload
- DVR

Note: All of the above NFV features can be enabled by using the JetPack version 16.1 automation toolkit.

Note: Red Hat OpenStack Platform version 16.1 has Open Virtual Network (OVN) controller as the default software defined networking solution for supplying network services to OpenStack. However, OVN has limited NFV feature support at this time. JetPack version 16.1 automation toolkit installs Red Hat OpenStack Platform version 16.1 with Open vSwitch (OVS) controller as default, to fully support all the NFV features.

NUMA Optimization and CPU Pinning

The Dell EMC Ready Architecture for Red Hat OpenStack Platform version 16.1 provides the ability to enable NUMA optimization and CPU pinning support on all compute nodes at the core or the edge site(s) in the solution.

Non-Uniform Memory Access or NUMA allows multiple CPUs to share L1, L2, L3 caches, and main memory. When running workloads on NUMA hosts, it is important that the vCPUs executing processes are on the same NUMA node as the memory used by these processes. This ensures all memory accesses are local to the node and thus do not consume the limited cross-node memory bandwidth, adding latency to memory accesses. Similarly, large pages are assigned from memory and benefit from the same performance improvements as memory allocated using standard pages.

For real-time workloads, it is beneficial to control which host CPUs are bound to an instance's vCPUs. This process is known as CPU pinning. No instance with pinned CPUs can use the CPUs of another pinned instance, thus preventing resource contention between instances.

Note: The NUMA feature can be enabled by the JetPack version 16.1 automation toolkit and enables both NUMA Optimization and vCPU Pinning.

Hugepages

The Dell EMC Ready Architecture for Red Hat OpenStack Platform version 16.1 provides the ability to enable hugepage support on the compute nodes at the core or the edge site(s) in the solution.

If a program or a VM instance uses large data structures, then the use of small (i.e. 4KB) pages leads to the fragmentation of those data structures and loss of data locality. With Hugepages of sizes 2MB and 1GB, programs and VM instances can improve data locality, thus leading to a higher performing and lower latency architecture

Note: Hugepages can be enabled by the JetPack version 16.1 automation toolkit.

OVS-DPDK

The Dell EMC Ready Architecture for Red Hat OpenStack Platform version 16.1 provides the ability to enable OVS-DPDK support based on 2 ports or 4 ports on the compute nodes at the core or the edge site(s) in the solution.

Open vSwitch (OVS) is a multilayer software/virtual switch used to interconnect virtual machines in the same host and between different hosts. OVS makes use of the kernel for packet forwarding through a data path known as "fastpath" which consists of a simple flow table with action rules for the received packets. Exception packets or packets with no corresponding forwarding rule in the flow table are sent to the user space (slowpath). Switching between two memory spaces creates a lot of overheard, thus making the user space "slowpath". User space makes a decision and updates the flow table in the kernel space accordingly so they can be used in the future.
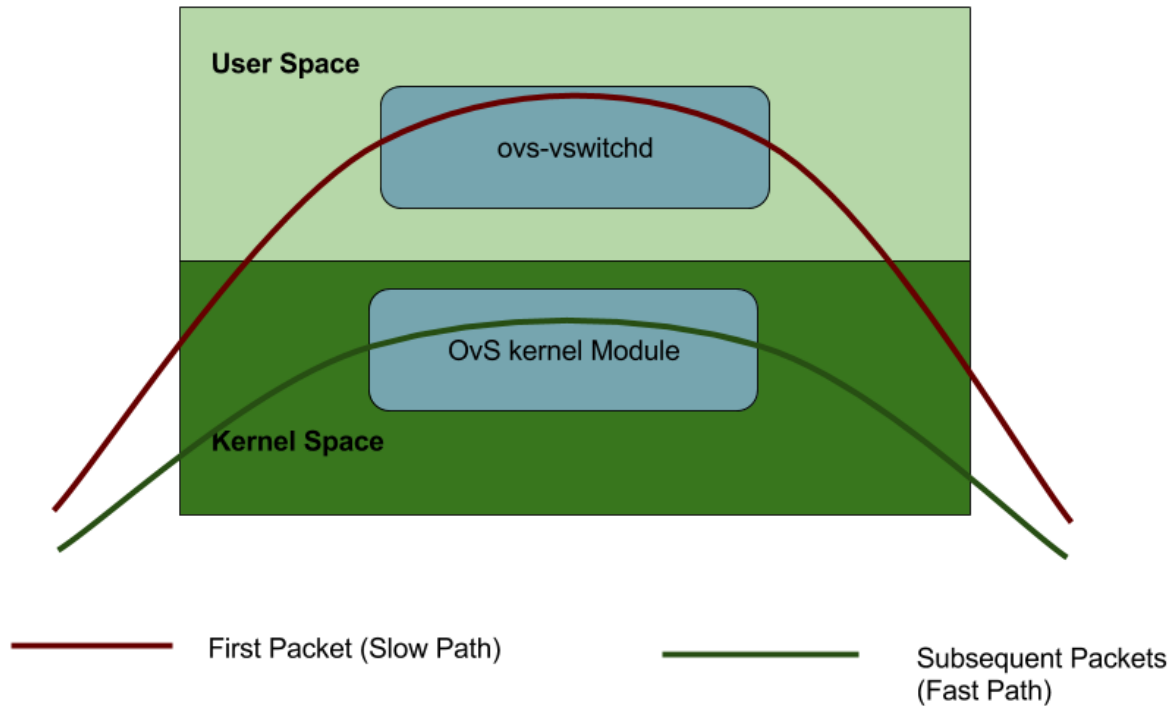
Figure 1: OVS Architecture

As can be seen in the Figure 1, the OVS kernel module acts as a cache for the user space. And just like a cache, its performance increases as the number of rules increase in the user space.

DPDK (Data-Plane Development Kit) eliminates packet buffer copies. It does this by running a dedicated poll-mode driver, and allocating hugepages for use as a packet buffer, then passing pointers to the packets. The elimination of copies leads to higher performance. OVS, when enabled to use DPDK-controller physical NIC interfaces, experiences a tremendous boost to packet delivery performance. It is also advantageous that both OVS and DPDK can operate in userspace, thus reducing kernel switches and improving packet processing efficiencies.

> **Note:** Enabling OVS-DPDK requires both Hugepages and NUMA (CPU Pinning) to be enabled.

> **Note:** OVS-DPDK can be enabled by the JetPack version 16.1 automation toolkit.

SR-IOV

The Dell EMC Ready Architecture for Red Hat OpenStack Platform version 16.1 provides the ability to enable SR-IOV support based on 2 ports and/or 4 ports on the compute nodes at the core or the edge site(s) in the solution.

Single root I/O virtualization (SR-IOV) is an extension to the PCI Express (PCIe) specification. SR-IOV enables a single PCIe device to appear as multiple, separate virtual devices. Traditionally in a virtualized environment, a packet has to go through an extra layer of the hypervisor, that results in multiple CPU interrupts per packet. These extra interrupts cause a bottleneck in high traffic environments. SR-IOV enabled devices have the ability to dedicate isolated access to its resources among various PCIe hardware functions. These functions are later assigned to the virtual machines which allow direct memory access (DMA) to the network data.

By default, SR-IOV is **not enabled** in the Dell EMC PowerEdge R-Series system BIOS. When SR-IOV is deployed, virtual functions **are not created** on the NIC interfaces.

Use the following steps to set the virtualization mode to SR-IOV and VF count to 64 in the Dell EMC PowerEdge R-Series system BIOS device settings for Mellanox NICs:

1. Enter system BIOS during boot by selecting F2.
2. Select `Device Settings`.
3. Select the Network Interface Card.

   > **Note:** SR-IOV has been validated with Mellanox ConnectX-5 / Intel XXV 710 NICs. Below is an example of SR-IOV configuration with Mellanox NIC

4. Select the `Device Level Configuration` option.

Figure 2: Mellanox configuration

5. Change the virtualization mode from None to SR-IOV and the `PCI Virtual Functions Advertised` from the default setting eight to the desired value.

   > **Note:** Dell EMC has tested with a value of 64.

6. Repeat steps three through five for each Mellanox ConnectX-5 100GbE adapter where SR-IOV will be enabled.
7. Save BIOS settings and reboot the system.

   > **Note:** SR-IOV can be enabled for both Mellanox ConnectX-5 and Intel XXV 710 NICs by the JetPack version 16.1 automation toolkit.

SR-IOV OVS-Hardware-Offload with VF-LAG

The Dell EMC Ready Architecture for Red Hat OpenStack Platform version 16.1 provides the ability to enable SR-IOV offloading with VF-LAG (bonding of PFs) functionality based on 2 ports on all the compute nodes at the core in the solution.

Open vSwitch hardware offload is a RHOSP feature which takes advantage of single root input/output virtualization (SR-IOV). An OVS software based solution is CPU intensive, affecting system performance and preventing full utilization of available bandwidth. This bottleneck can be addressed by making use of the OVS hardware offload capability, where the OVS data plane is moved to the underlying offloading capable smart NIC, while keeping the OVS control plane unmodified. This enables higher OVS performance without the associated CPU load.

In this Dell EMC Ready Architecture for Red Hat OpenStack Platform 16.1, Dell EMC has validated this feature enabling 2 port SR-IOV offload capability with VF-LAG functionality using Mellanox 100GbE ConnectX-5 network adapters for tenant networks with bonding. Below is a logical diagram showing 2 port SR-IOV OVS Offload connections to the Dell EMC switches.
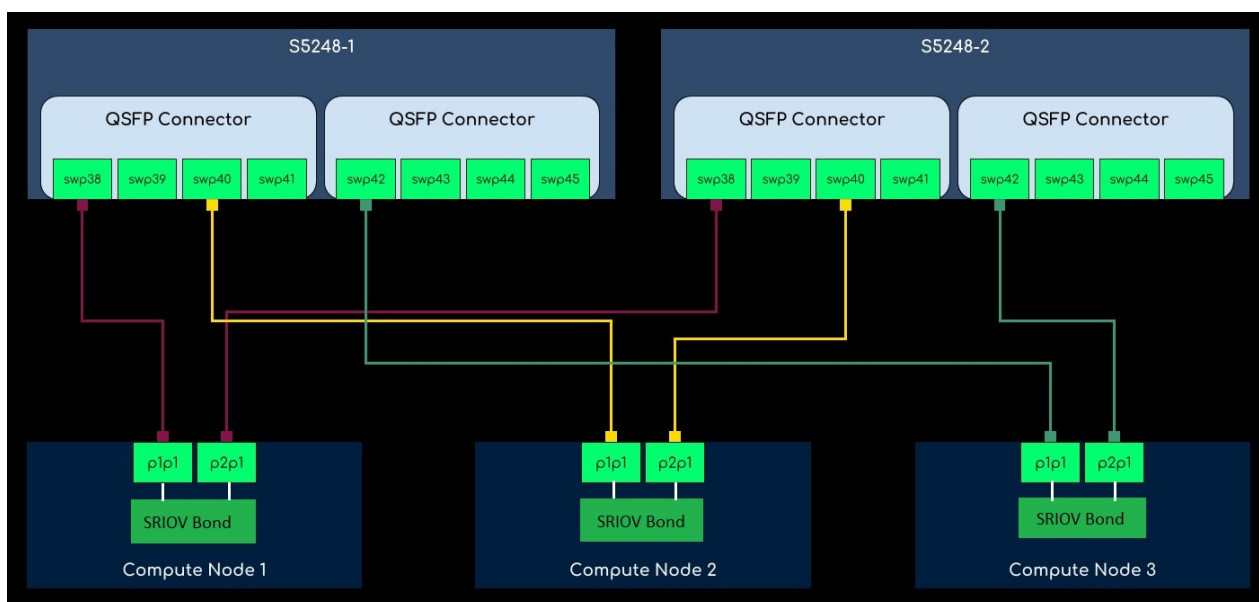
Figure 3: 2 Port SR-IOV OVS Offload

Note: SR-IOV offloading has been tested with Mellanox ConnectX-5 100G NICs only.

Note: For SR-IOV Offload with VF-LAG functionailty to work, both ports of the bond should originate from a single Mellanox ConnectX-5 NIC.

Note: SR-IOV offloading can be enabled by the JetPack version 16.1 automation toolkit.

Distributed Virtual Router (DVR)

The Dell EMC Ready Architecture for Red Hat OpenStack Platform version 16.1 provides the ability to enable Distributed Virtual Routing (DVR) on all all the compute nodes in the solution.

Distributed Virtual Routing (DVR) offers an alternative routing design to the centralized routing model. It intends to isolate the failure domain of the Controller node and optimize network traffic by deploying the L3 agent and schedule routers on every Compute node. By eliminating a centralized layer 3 agent, the routing that was performed by single node (Primary controller), is now distributed across the compute nodes using the local L3 agent. DVR follows the below routing flow rules:

- East-West traffic is routed directly on the Compute nodes in a distributed fashion.
- North-South traffic with floating IP is distributed and routed on the Compute nodes. This traffic requires the external network connects to each Compute node.
- North-South traffic without floating IP allocates and still needs a dedicated Controller node.
- The L3 agent on the Controller node is configured with a new dvr_snat mode so that the node serves only SNAT traffic.
- The Neutron metadata agent distributes and deploys on all Compute nodes. The metadata proxy service hosts all the distributed routers.

Note: DVR can be enabled by the JetPack version 16.1 automation toolkit.

Chapter

8

Additional Red Hat/OpenStack features

Topics:

- *Barbican*
- *Octavia*
- *Satellite*

This Ready Architecture Guide supports the following additional Red Hat OpenStack features:

- Barbican
- Octavia
- Satellite Support

Note: These features can be enabled by using the JetPack version 16.1 automation toolkit.

Barbican

Barbican is the secrets manager for Red Hat OpenStack Platform 16.1. You can use the barbican API and command line to centrally manage the certificates, keys, and passwords used by OpenStack services.

Symmetric encryption keys include -

• Block Storage (cinder) volume encryption
• Ephemeral disk encryption
• Object Storage (swift) encryption

Asymmetric keys and certificates include -

• Glance image signing and verification

Note: Barbican can be enabled by the JetPack version 16.1 automation toolkit.

Octavia

Octavia is the OpenStack Load-balancing as a Service (LBaaS) version 2 implementation for the Red Hat OpenStack platform. It accomplishes its delivery of load balancing services by managing a fleet of virtual machines, collectively known as amphorae, which it spins up on demand.

Load Balancing Methods -

• Round robin - Rotates requests evenly between multiple instances.
• Source IP - Requests from a unique source IP address are consistently directed to the same instance.
• Least connections - Allocates requests to the instance with the least number of active connections.

Note: Octavia can be enabled by the JetPack version 16.1 automation toolkit.

Satellite

JetPack support for Red Hat Satellite 6.5 gives the user the ability to deploy from a satellite instance. All nodes within the Dell EMC Ready Architecture for Red Hat OpenStack Platform can register and pull required packages and container images from the satellite instance, rather than the Red Hat Content Delivery Network (CDN).

Note: Satellite support can be enabled by the JetPack version 16.1 automation toolkit.

Chapter

9

Distributed Compute Nodes (DCN)

Topics:

An OpenStack® based cloud is now a common need by many organizations and Dell EMC with Red Hat have worked together to build a jointly engineered and validated edge architecture, using Red Hat OpenStack Platform's Distributed Compute Nodes (DCN) capability, that details software, hardware, and integration points of all solution components. The architecture provides prescriptive guidance and recommendations for:

- Edge compute nodes
- Network design
- Software layout
- Offers suggestion for other system configurations

Overview

DCN, as part of Red Hat OpenStack Platform 16.1, leverages OpenStack features like Availability Zones and provisioning over routed L3 networks with Ironic, to enable deployment of compute nodes to remote locations. For example, a service provider may deploy several DCN sites to scale out a virtual Radio Access Network (vRAN) implementation.

DCN has several caveats that must be considered when planning remote compute site deployment(s):

- Only Compute can be run at an edge site, other services such as persistent block storage are not supported at this time
- Image considerations - Overcloud images for bare-metal provisioning of the remote compute nodes are pulled from the undercloud. Also, instance images for VMs running on edge nodes will initially be fetched from the control plane the first time they are used, subsequent instances will use the locally cached image. Images are large files, implying a fast, reliable connection to the Red Hat Director node and control plane is required.
- Networking:

    - Latency - a round-trip between the control plane and remote site must be under 100ms or stability of the system could become compromised
    - Drop-outs - If an edge site temporarily loses its connection to the control plane, then no OpenStack control plane API or CLI operations can be executed until connectivity is restored to the site. For example, existing workloads will continue to run, but no new instances can be started until the connection is restored. Any control functions like snapshotting, live migration etc. cannot occur until the link between the central cloud and edge site is restored, as all control features are dependent on the control plane being able to communicate with the edge site.

        Note: Connectivity issues are DCN site specific, losing connection to one DCN site does not effect other DCN sites.

    - This guide recommends using Provider networks for DCN workloads at this time. Depending on the type of workloads you are running on the edge nodes, and existing networking policies, there are several ways for configuring instance IP addressing:

        - Static IPs using config-drive in conjunction with cloud-init - Utilizing config-drive as the metadata API server leverages the virtual media capabilities of Nova, which means there are no Neutron Metadata agents or DCHP relay required to assign a IP address to edge instances.
        - DHCP relay - Forwards DHCP requests to Neutron at central site.

            Note: A separate DHCP relay instance is required for each provider network.

        - External DCHP server at edge site - in this case instance IP addresses are not managed by neutron.

    - Inter-compute node awareness - A limitation of Neutron is that it is not able to identify individual compute nodes as remote or local. Therefore each compute node, across all DCN sites, including the central cloud, will have a list of every other compute node. Depending on your networking configuration this can happen by:

        - When using VXLAN - First, the same Neutron networks must be configured at every site, then every Compute node will build a VXLAN tunnel (through the control plane) to all Controllers and compute nodes regardless if they are remote or local.
        - VLAN only - This method requires that identical network bridges and VLANs are used across all sites.

DCN topology example

It is possible with DCN to deploy any number of remote compute sites, each with it's own Availability Zone (AZ). Once the site is provisioned, specifying the site's AZ when launching an instance is all that is required to launch an instance, for example:

```
openstack server create --flavor tiny --image cirros \
--network edge-management-vxlan \
--security-group edge-management \
--availability-zone AZ_DCN_2 \
edge_vm_1
```

**Central OpenStack Deployment
(Director, HA Control Plane & Storage)**

**L3
Routing**

**DCN Site 1
(Availability Zone: AZ_DCN_1)**

**DCN Site 2
(Availability Zone: AZ_DCN_2)**

**DCN Site N
(Availability Zone: AZ_DCN_N)**
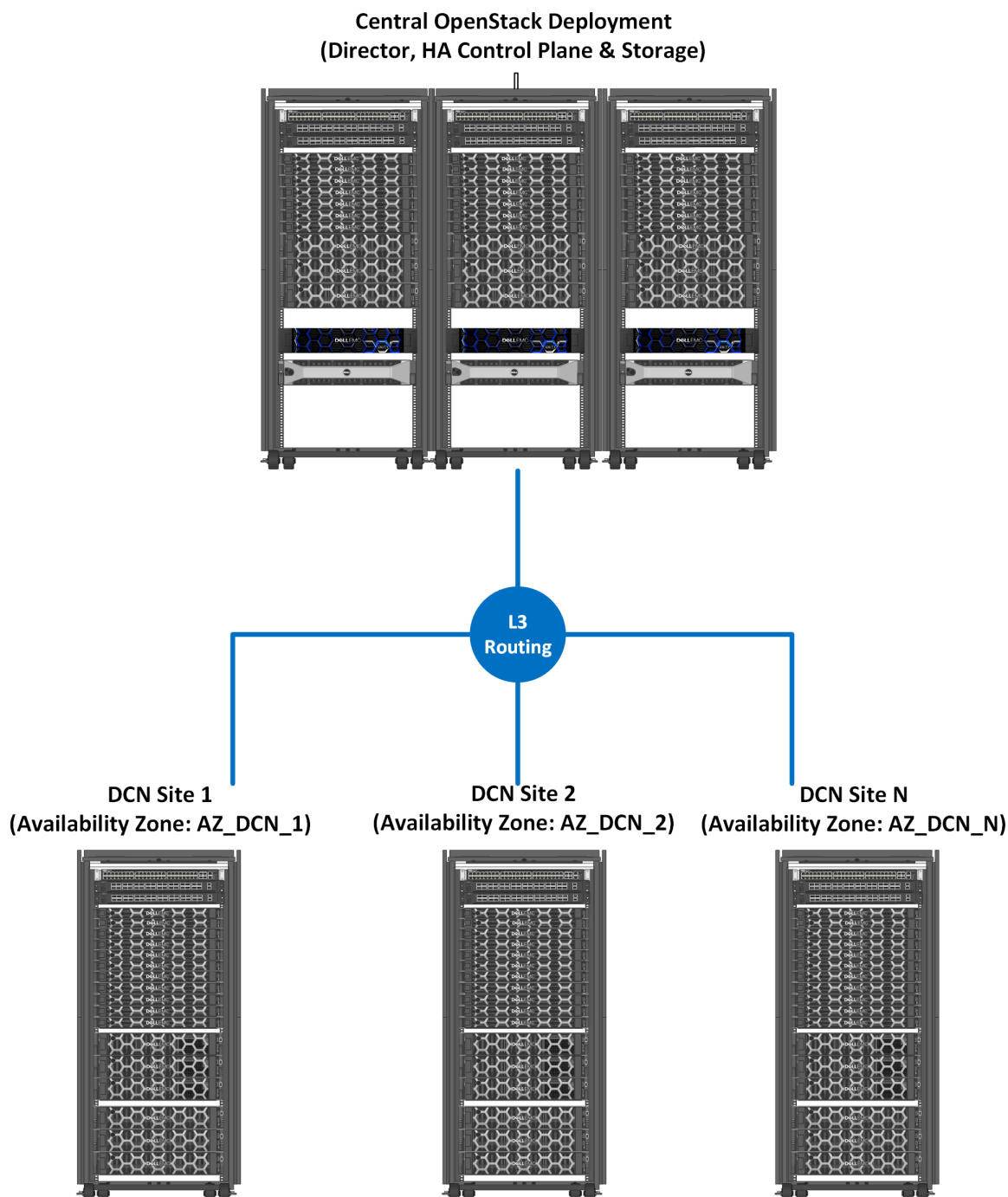
Figure 4: DCN topology

Red Hat OpenStack Platform and DCN further reading

For more on Red Hat OpenStack Platform 16.1 and DCN please see:

- *Red Hat OpenStack Platform 16.1*
- *Deploying Distributed Compute Nodes to Edge Sites*
- *Routed L3 Networks (Spine Leaf Networking)*

Hardware options

To reduce time spent on specifying hardware for an initial edge deployment, this Architecture Guide offers a full solution using validated Dell EMC PowerEdge server hardware designed to allow a wide range of configuration options, including optimized configurations for edge compute nodes.

Dell EMC recommends starting an edge site deployment using components from this Architecture Guide - Version 16.1 because the hardware and operations processes comprise a flexible foundation upon which to expand as your edge site deployment(s) grow, so your investment is protected.

As noted throughout this Ready Architecture Guide - Version 16.1, Dell EMC constantly adds capabilities to expand this offering, and other hardware may be available at the time of this reading. Please contact your Dell EMC sales representative for more information on new hardware releases.

Service layout

During a DCN site deployment, a subset of OpenStack services will be installed each Compute node.

Table 8: Edge compute node services

Hypervisor (KVM)
Nova compute services
Neutron (ovs/ml2 agent)
Heat agent

Deployment overview

This is an overview of the DCN deployment process that can be utilized for planning purposes:

Note: The DCN feature can be enabled by the JetPack version 16.1 automation toolkit and can enable multiple DCN sites.

1. **Hardware setup:**
   - Rack and stack
   - Cabling
   - iDRAC setup
   - PXE NIC configuration
   - Server BIOS and RAID configuration
   - Switch configuration
2. **Software setup at each DCN site:**
   - Install a DCHP relay(s) used to forward DCHP traffic to Neutron running on the central cloud control plane
     - At least one DCHP relay is required for provisioning compute nodes at the DCN site
     - Where the DHCP relay(s) are installed is up to the customer
   - Discover edge site nodes
   - Import discovered nodes into Red Hat OpenStack Director
   - Configure Overcloud files for each DCN site
   - Provision DCN site
   - Validate DCN nodes' networking
3. **Environment tests**

- Once Red Hat OpenStack Director has completed a DCN site deployment, a typical set of validation tasks will include:
  - Create a flavor, with the supporting metadata to target the DCN site when launching an instance, for example:

```
$ openstack flavor create --property dcn-site-1=true dcn-site-1-
flavor
+---------------------------+-------------------------------------+
| Field                     | Value                               |
+---------------------------+-------------------------------------+
| id                        | 4bc6c3e7-42be-4620-ae84-819459baf496 |
| name                      | dcn-site-1-flavor                   |
| os-flavor-access:is_public | True                               |
| properties                | dcn-site-1='true'                   |
+---------------------------+-------------------------------------+
```

  - Launch an instance using the new flavor and DCN site provider network, for example:

```
$ openstack server create --flavor dcn-site-1-flavor \
--network dcn-site-1-provider-net dcn-site-1-vm-1 ...
+----------+-------------------------------------+
| Field    | Value                               |
+----------+-------------------------------------+
| flavor   | dcn-site-1-flavor (4bc6c3e7-42be-...) |
| hostId   | 2ef6c3e7-32ae-2120-ae84-819459baf488 |
| name     | dcn-site-1-vm-1                     |
| addresses | dcn-site-1-provider-net=192.168.1.123 |
| ...      | ...                                 |
+----------+-------------------------------------+
```

  - Validate instance is running and that networking and addressing is correct

DCN solution architecture

Solution bundle 25GbE/100GbE network configuration

The network for the Dell EMC Ready Architecture for Red Hat OpenStack Platform has been designed to support production-ready servers *with a highly available network configuration*.

> **Note:** See figure *Figure 10: 25GbE cluster network logical architecture for Edge profile* on page 60 for a logical architectural layout.

Solution expansion

The Dell EMC Ready Architecture for Red Hat OpenStack Platform can be expanded by:

- Adding compute nodes to an existing DCN site deployment(s). No more than 20 compute nodes per site are supported at this time.
- Deploy additional DCN site(s)

> **Note:** Currently, Red Hat OpenStack Platform 16.1 supports up to a total of 700 compute nodes across all sites, including the core OpenStack installation. For other expansion details, please speak with your Dell EMC sales representative.

Chapter
10

Operational notes

Topics:

- *High availability*
- *Service layout*
- *Deployment overview*

This section provides a basic overview of several important system aspects.

High availability

In order for the solution to be ready for production, different systems need to be fault-tolerant. The Architecture Guide design utilizes both hardware-based and software-based redundancy. This includes, but is not limited to:

- Operating Systems are hosted on either a RAID 1 or RAID 10 hard drive set.
- Critical network connections from server to switch utilize network bonding.
- Multiple Controllers host the control plane services. Minimally, 3 Controller nodes are required.
- Control plane services are made highly available utilizing *ha-proxy*, *Corosync, Pacemaker*, and/or native resiliency.
- Red Hat Ceph storage utilizes a minimum of three (3) servers.
- Red Hat Ceph storage is used with either replication or erasure coding.
- Optional: Instance High Availability

    This validated option utilizes remote *pacemaker* to monitor the Compute nodes. If preset criteria are met, the process of migrating instances off of the failing Compute nodes to others begins. If a Compute node completely fails, *pacemaker* can be configured to start the failed instances on different Compute nodes.

- Optional: Dell EMC Unity storage , SC series storage , and Dell EMC PowerMax storage arrays are highly available.

> **Note:** The Solution Admin Host, and the server hosted on it (Red Hat OpenStack Director), are not fault tolerant, but are not required for continued functionality of the OpenStack cluster.

Service layout

During the deployment each service configured by the Dell EMC Ready Architecture for Red Hat OpenStack Platform needs to reside upon a particular hardware type. For each server platform, two types of nodes have been designed:

- Dell EMC PowerEdge R640 or Dell EMC PowerEdge R740xd for Computes, Controllers, Solution Admin Hosts, or Infrastructure hardware type
- Dell EMC PowerEdge R740xd for Storage nodes.

Red Hat OpenStack Director is designed for flexibility, enabling you to try different configurations in order to find the optimal service placement for your workload. *Table 9: Overcloud: Node type to services* on page 46 presents the recommended layout of each service.

The Red Hat OpenStack Director is deployed to the Solution Admin Host as an individual VM. This enables the VM to control its respective resources.

Table 9: Overcloud: Node type to services

Node to deploy	Service
Solution Admin Host (KVM)	Red Hat OpenStack Director
Solution Admin Host (KVM)	Red Hat Ceph Storage Dashboard
OpenStack Controllers	Cinder-scheduler
OpenStack Controllers	Cinder-volume
OpenStack Controllers	Database-server
OpenStack Controllers	Glance-Image
OpenStack Controllers	HAproxy (Load Balancer)

Node to deploy	Service
OpenStack Controllers	Heat
OpenStack Controllers	Keystone-server
OpenStack Controllers	Neutron-server
OpenStack Controllers	Nova-Controller
OpenStack Controllers	Nova dashboard-server
Three or more Compute Nodes	Nova-multi-compute
OpenStack Controllers	Pacemaker
OpenStack Controllers	RabbitMQ-server (Messaging)
OpenStack Controllers	Barbican (Secure manangement of secrets)
OpenStack Controllers	Octavia (LBaaS)
OpenStack Controllers	Red Hat Ceph storage RADOS Gateway
OpenStack Controllers	Red Hat Ceph storage Monitor [3]
Three or more Storage Servers	Red Hat Ceph storage (Block)[4]
Optional Services	
Dell EMC Unity storage	
Dell EMC PowerMax storage	
Dell EMC SC series storage	
Dell EMC SC series storage Enterprise Manager Server	

Deployment overview

This is an overview of the deployment process that can be utilized for planning purposes:

1. **Hardware Setup:**

   • Rack and stack
   • Cabling
   • iDRAC setup
   • PXE NIC configuration
   • Server BIOS and RAID configuration
   • Switch configuration

2. **Software Setup:**

   • Deploy Solution Admin Host for provisioning services:

     • Deploy Red Hat Ceph Storage Dashboard VM to the Solution Admin Host
     • Deploy Red Hat OpenStack Director Virtual Server VM to the Solution Admin Host
   • Discover nodes

---

[3] The number of OSDs that can be supported with 3 Controller nodes is listed at 1,000 (*https://access.redhat.com/articles/1548993*)

[4] Available through a custom Services engagement.

- Import discovered nodes into Red Hat OpenStack Director
- Configure Overcloud files
- Provision Overcloud
- Validate all nodes' networking
- Post-deployment, including but not limited to:
  - Enabling fencing
  - Enabling local storage for ephemeral

3. **Environment Tests**

- Tempest can be used to validate the deployment. At minimum the following tests should be performed:
  - Project Creation
  - User Creation
  - Network Creation
  - Image upload and launch
  - Floating IP Assignment
  - Basic network testing
  - Volume creation and attachment to VM
  - Object storage upload, retrieval and deletion
  - Deletion of all artifacts created during validation.

Chapter
11

Solution architecture

Topics:

* *Solution common settings*
* *Solution with 25GbE/100GbE networking overview*

This core architecture provides prescriptive guidance and recommendations, jointly engineered by Dell EMC and Red Hat, for deploying Dell EMC Ready Architecture for Red Hat OpenStack Platform version 16.1 with Dell EMC infrastructure.

The goals are to:

* Provide practical system design guidance and recommended configurations
* Develop tools to use with OpenStack for day-to-day usage and management
* Develop networking configurations capable of supporting your production system

The development of this architecture builds upon the experience and engineering skills of Dell EMC and Red Hat, and encapsulates best practices developed in numerous real-world deployments. The designs and configurations in this architecture have been tested in Dell EMC and Red Hat labs to verify system functionality and operational robustness.

The solution consists of the components shown in *Figure 5: Solution with 25GbE/100GbE, Red Hat Ceph storage cluster, optional Dell EMC Unity storage, optional Dell EMC PowerMax storage and optional SC series storage* on page 55, and represents the base upon which all optional components and expansion of the Dell EMC Ready Architecture for Red Hat OpenStack Platform are built.

Solution common settings

Many settings are common through the Solution. The configurations that are tested are outlined in this section.

Solution Admin Host (SAH) networking

The Solution Admin Host is configured for 25GbE with the server internal bridged networks for the Virtual Machines. It is physically connected to the following networks:

- **Management Network** — used by the Red Hat OpenStack Director for iDRAC control of all Overcloud nodes.
- **Private API Network** — Used by the Red Hat OpenStack Director to run Tempest tests against the OpenStack private API
- **Provisioning Network** — Used by the Red Hat OpenStack Director to service DHCP to all hosts, provision each host, and act as a proxy for external network access
- **Public API Network** — used for:

  - Inbound Access

    - HTTP/HTTPS access to the Red Hat OpenStack Director
    - HTTP/HTTPS access to the Red Hat Ceph Storage Dashboard VM
    - Optional - SSH Access to the Red Hat OpenStack Director and Red Hat Ceph Storage Dashboard VM
  - Outbound Access

    - HTTP/HTTPS access for Red Hat Ceph storage, RHEL, and RHOSP subscriptions.
    - Used by the Red Hat OpenStack Director to run Tempest tests using the OpenStack public API.
- **Storage Network** — Used by the Red Hat Ceph Storage Dashboard VM to monitor the Red Hat Ceph storage Cluster.

Node type 802.1q tagging information

The solution is designed with the idea that different network traffic should be segregated from other traffic. This is accomplished by utilizing 802.1q VLAN Tagging for the different segments. The tables *Table 10: OpenStack node type to network 802.1q tagging* on page 50, *Table 11: OpenStack Compute Node for xSP and CSP profile to network 802.1q tagging* on page 51, and *Table 12: Storage node type to network 802.1q tagging* on page 51 summarize this. This segregation is independent of network speed and used by 25GbE configuration.

Table 10: OpenStack node type to network 802.1q tagging

Network	Solution Admin Host	OpenStack controller	Red Hat Ceph storage
External Network VLAN for Tenants (Floating IP Network)	Not Connected	Connected, tagged	Not Connected
iDRAC physical connection to the Management/OOB VLAN	Connected, Untagged	Connected, Untagged	Connected, Untagged
Internal Networks VLAN for Tenants	Not Connected	Connected, Tagged	Not Connected
Management/OOB Network VLAN	Connected, Tagged	Not Connected	Not Connected
Private API Network VLAN	Connected, Tagged	Connected, Tagged	Not Connected

Network	Solution Admin Host	OpenStack controller	Red Hat Ceph storage
Provisioning VLAN	Connected, Tagged	Connected, Untagged	Connected, Untagged
Public API Network VLAN	Connected, Tagged	Connected, Tagged	Not Connected
Storage Clustering VLAN	Not Connected	Not Connected	Connected, Tagged
Storage Network VLAN	Connected, Tagged	Connected, Tagged	Connected, Tagged
Tenant Tunnel Network	Not Connected	Connected, Tagged	Not Connected

Table 11: OpenStack Compute Node for xSP and CSP profile to network 802.1q tagging

Network	xSP OpenStack compute	CSP - OpenStack compute NFV
External Network VLAN for Tenants (Floating IP Network)	Not Connected	Connected, Tagged
iDRAC physical connection to the Management/OOB VLAN	Connected, Untagged	Connected, Untagged
Internal Networks VLAN for Tenants	Connected, Tagged	Connected, Tagged
Management/OOB Network VLAN	Not Connected	Not Connected
Private API Network VLAN	Connected, Tagged	Connected, Tagged
Provisioning VLAN	Connected, Untagged	Connected, Untagged
Public API Network VLAN	Not Connected	Not Connected
Storage Clustering VLAN	Not Connected	Not Connected
Storage Network VLAN	Connected, Tagged	Connected, Tagged
Tenant Tunnel Network	Connected, Tagged	Connected, Tagged

Table 12: Storage node type to network 802.1q tagging

Network	Dell EMC Unity	Dell EMC SC series storage Enterprise Manager	Dell EMC SC series storage array
External Network for Tenants VLAN (Floating IP Network)	Connected, Tagged	Not Connected	Not Connected
iDRAC physical connection to the Management/OOB VLAN	Not Connected	Not Connected	Not Connected
Internal Networks VLAN for Tenants	Connected, Tagged	Not Connected	Not Connected
Management/OOB Network VLAN	Not Connected	Not Connected	Not Connected

Network	Dell EMC Unity	Dell EMC SC series storage Enterprise Manager	Dell EMC SC series storage array
Provisioning VLAN	Not Connected	Not Connected	Not Connected
Private API Network VLAN	Not Connected	Not Connected	Not Connected
Public API Network VLAN	Not Connected	Connected, Untagged	Not Connected
Storage Network VLAN	Connected, Untagged	Connected, Untagged	Connected, Untagged
Storage Clustering VLAN	Not Connected	Not Connected	Not Connected
Tenant Tunnel Network	Not Connected	Not Connected	Not Connected

Table 13: OpenStack node types for HCI profile to network 802.1q tagging

Network HCI	OpenStack controller	OpenStack
External Network VLAN for Tenants (Floating IP Network)	Not Connected	Connected, tagged
iDRAC physical connection to the Management/OOB VLAN	Connected, Untagged	Connected, Untagged
Internal Networks VLAN for Tenants	Connected, Tagged	Connected, Tagged
Management/OOB Network VLAN	Not Connected	Not Connected
Private API Network VLAN	Connected, Tagged	Connected, Tagged
Provisioning VLAN	Connected, Tagged	Connected, Untagged
Public API Network VLAN	Not Connected	Not Connected
Storage Clustering VLAN	Not Connected	Not Connected
Storage Network VLAN	Connected, Tagged	Connected, Tagged
Tenant Tunnel Network	Connected, Untagged	Connected, Tagged

Solution Red Hat Ceph storage configuration

The Red Hat Ceph storage cluster provides data protection through replication, block device cloning, and snapshots. By default the data is striped across the entire cluster, with three replicas of each data entity. The number of storage nodes in a single cluster can scale to hundreds of nodes and many petabytes in size.

Red Hat Ceph storage considers the physical placement (position) of storage nodes within defined fault domains (i.e., rack, row, and data center) when deciding how data is replicated. This reduces the probability that a given failure may result in the loss of more than one data replica.

The Red Hat Ceph storage cluster services include:

- **Ceph Dashboard** — Ceph web based monitoring tool hosted on a VM.
- **RADOS Gateway** — Object storage gateway.
- **Object Storage Daemon (OSD)** — Running on storage nodes, the OSD serves data to the Red Hat Ceph storage clients from disks on the storage nodes. Generally, there is one OSD process per disk drive.
- **Monitor (MON)** — Running on Controller nodes, the MON process is used by the Red Hat Ceph storage clients and internal Red Hat Ceph storage processes, to determine the composition of the

cluster and where data is located. There should be a minimum of three MON processes for the Red Hat Ceph storage cluster. The total number of MON processes should be odd.

- **Ceph Manager Daemon (ceph-mgr)** — Running on Controller nodes alongside the MON processes, it provides additional monitoring and interfaces to external monitoring and management systems

> **Note:** If MON processes on Controller nodes become a bottleneck, then additional MON processes can be added to the cluster by using dedicated machines, or by starting MON processes on storage Nodes. A custom Services engagement can be arranged; please contact your Dell EMC sales representative for assistance.

The Storage Network VLAN is described in the Red Hat Ceph storage documentation as the public network. The Storage Cluster Network VLAN is described in the Red Hat Ceph storage documentation as the cluster network.

A supported distribution by Red Hat with production level support of Ceph is used in this solution: Red Hat Ceph storage 4, which also includes the Red Hat Ceph Storage Dashboard VM. The Red Hat Ceph Storage Dashboard also includes Red Hat Ceph storage troubleshooting and servicing tools and utilities. Red Hat Ceph Storage Dashboard is installed on a virtual machine that runs on the Solution Admin Host (SAH). Note that:

- The SAH must have access to the Controller and Storage nodes through the Private API Access VLAN in order to manage Red Hat Ceph storage; and for the monitoring process on all Storage nodes to return status and performance telemetry.

- The Controller nodes must have access to the Storage nodes through the Storage Network VLAN in order for the MON processes on the Controller nodes to be able to query the Red Hat Ceph storage MON processes, for the cluster state and configuration.

- The Compute nodes must have access to the Storage nodes through the Storage Network VLAN in order for the Red Hat Ceph storage client on that node to interact with the storage nodes, OSDs, and the Red Hat Ceph storage MON processes.

- The Storage nodes must have access to the Storage Network VLAN, as previously stated, and to the Storage Cluster Network VLAN.

Solution with 25GbE/100GbE networking overview

Since the Solution is designed for a production environment, key OpenStack services are made highly available (HA) by clustering the OpenStack controller nodes. The networking is based upon 25GbE bonds for data networks, and the network switches are configured for HA. The Out of Band Management (iDRAC's) network is not HA, and is 1GbE. The 100GbE networking is used in the solution for the user/tenant traffic.

For basic hardware configuration refer to *Bill of Materials* on page 61.

Solution 25GbE/100GbE rack layout

The Solution includes three (3) storage nodes, configured in a Red Hat Ceph storage cluster, which is tied into Cinder, Glance, and Nova.

See *Table 21: Solution Admin Host (SAH) Dell EMC PowerEdge R640* on page 65 *Table 22: Controller node Dell EMC PowerEdge R640* on page 65 *Table 23: Compute node Dell EMC PowerEdge R640* on page 66, and *Table 24: Storage node Dell EMC PowerEdge R740xd* on page 66 for hardware configurations. The Solution includes:

- Node 1: Solution Admin Host with the Red Hat OpenStack Director and the Red Hat Ceph Storage Dashboard installed.
- Nodes 2 - 4: Dell EMC PowerEdge R640 OpenStack Controllers
- Nodes 5 - 7 Dell EMC PowerEdge R640 Nova Compute Nodes
- Nodes 8 - 10: Dell EMC PowerEdge R740xd Storage Nodes

- Network Switches: Two (2) Dell EMC Networking S5232F-ON, and one (1) Dell EMC Networking S3048-ON

Note: The following rack is not to scale but shows the node types and usage.

Dell Networking S3048-ON

Dell Networking S5232F-ON

Dell Networking S5232F-ON

PowerEdge R640 – SAH
PowerEdge R640 – Controller 0
PowerEdge R640 – Controller 1
PowerEdge R640 – Controller 2
PowerEdge R640 – Compute 0
PowerEdge R640 – Compute 1
PowerEdge R640 – Compute 2

PowerEdge R740xd - Storage 0

PowerEdge R740xd - Storage 1

PowerEdge R740xd - Storage 2

Dell EMC SC Series – (optional)

Dell EMC Unity Series – (optional)

Dell EMC PowerMax Series –
(optional)
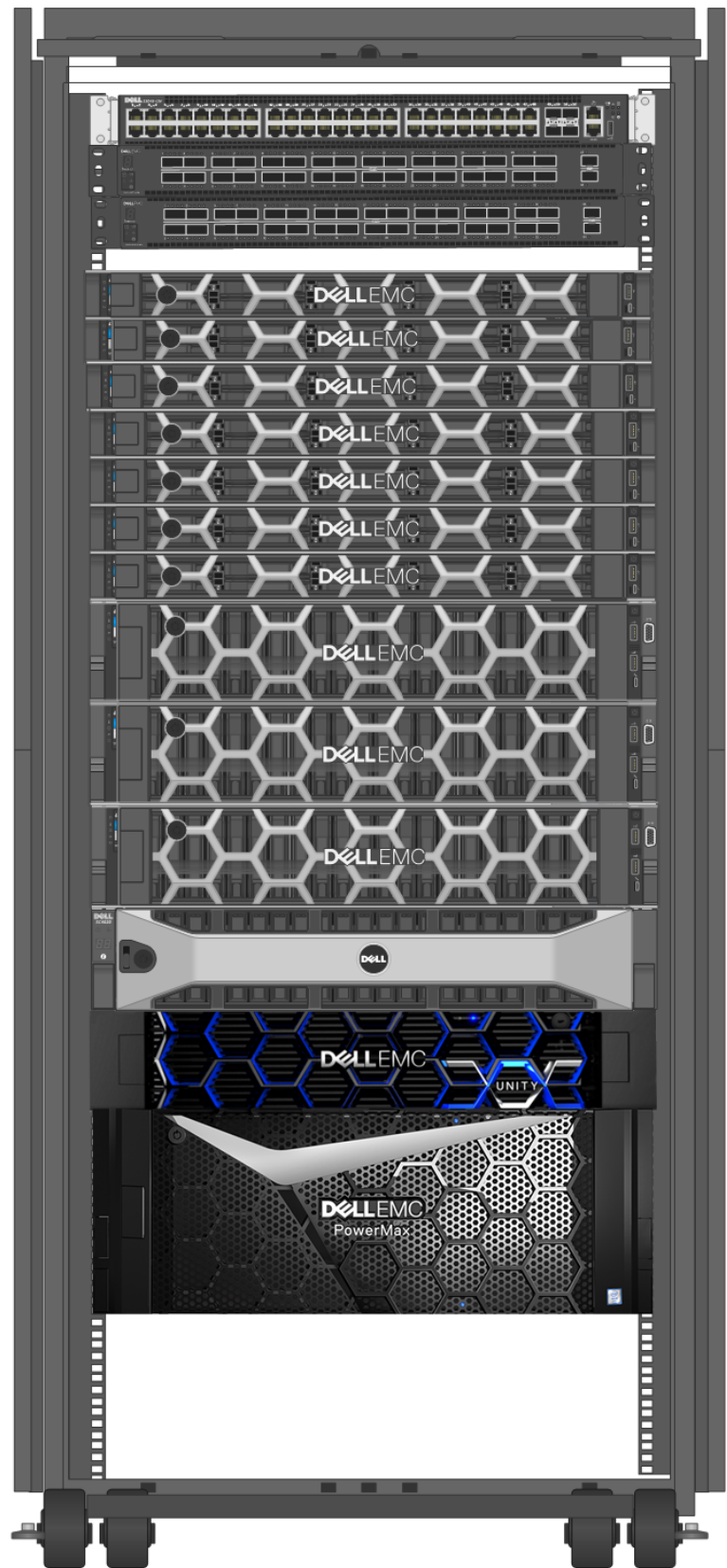
Figure 5: Solution with 25GbE/100GbE, Red Hat Ceph storage cluster, optional Dell EMC Unity storage, optional Dell EMC PowerMax storage and optional SC series storage

> ✎ **Note:** The following rack is not to scale but shows the node types and usage.

Dell Networking S3048-ON
Dell Networking S5232F-ON
Dell Networking S5232F-ON
PowerEdge XE2420 – DCN
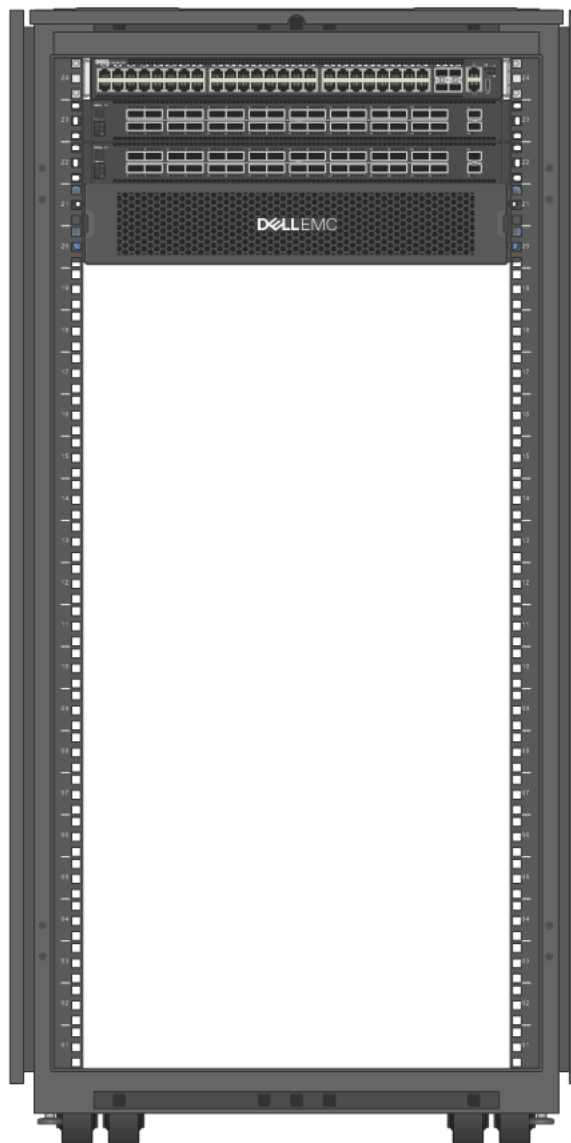
Figure 6: Solution with 25GbE/100GbE and Dell EMC PowerEdge XE2420

Solution 25GbE/100GbE network configuration

The network for the Dell EMC Ready Architecture for Red Hat OpenStack Platform has been designed to support production-ready servers with a highly available network configuration.

The node type will determine how the switches are configured for delivering the different networks. *Table 10: OpenStack node type to network 802.1q tagging* on page 50 and *Table 12: Storage node type to network 802.1q tagging* on page 51 outline the networks to the node types.

For the CSP Profile with non OVS Offload logical network:
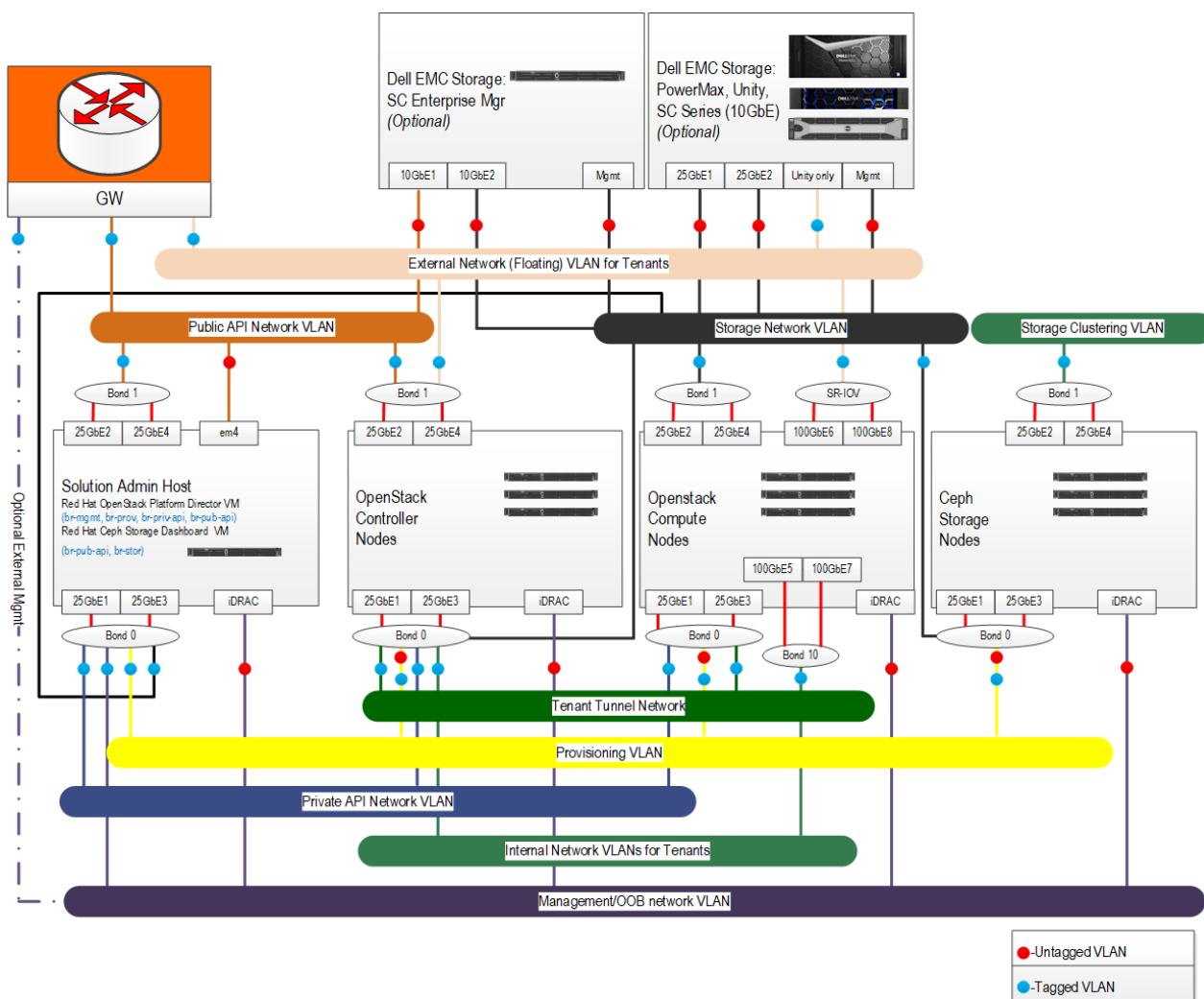
Figure 7: 25GbE/100GbE cluster network logical architecture for CSP profile

For the CSP Profile with OVS Offload logical network:
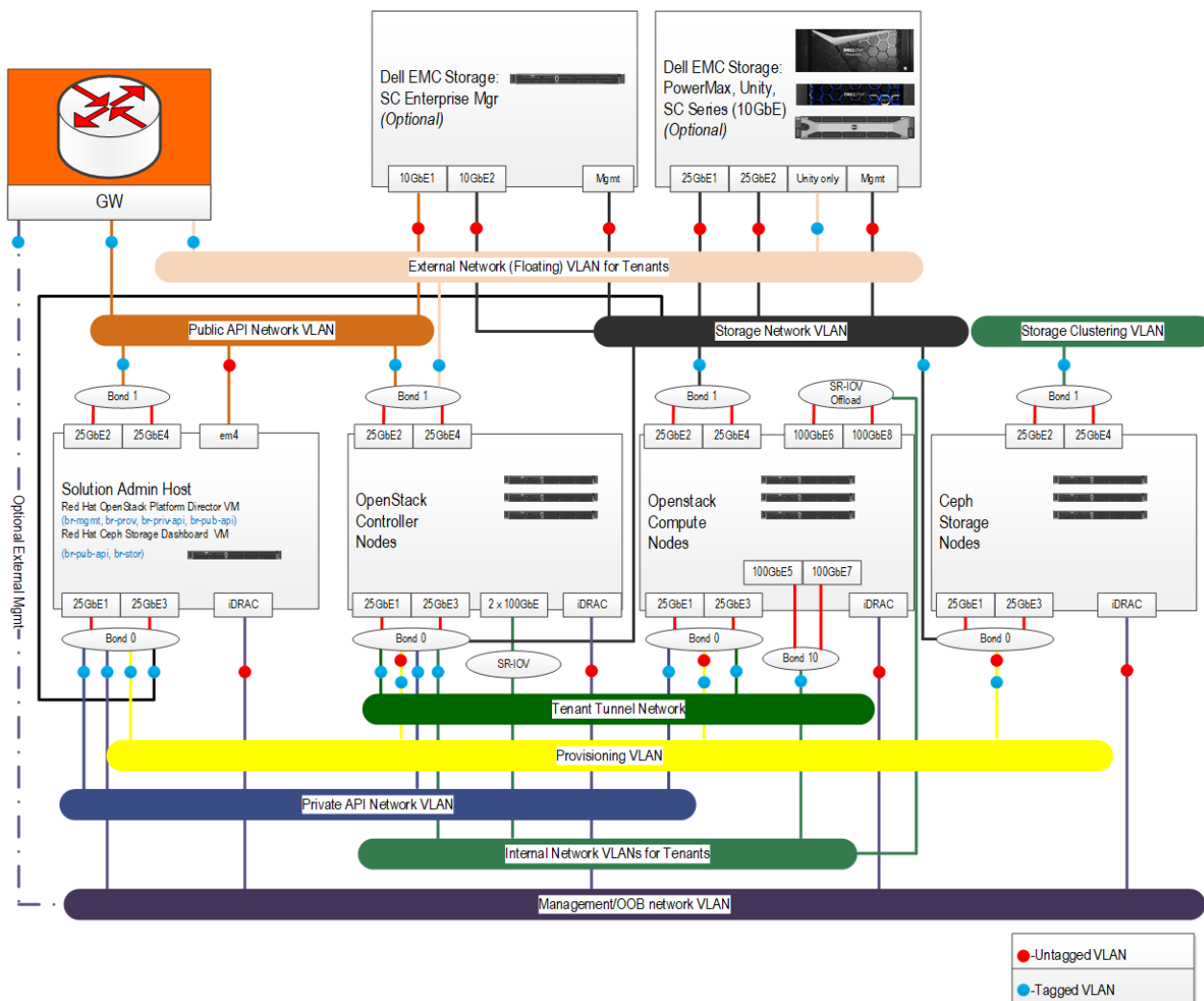
Figure 8: 25GbE/100GbE cluster network logical architecture for CSP OVS Offload profile

For the HCI profile with 25GbE/100GbE cluster network:
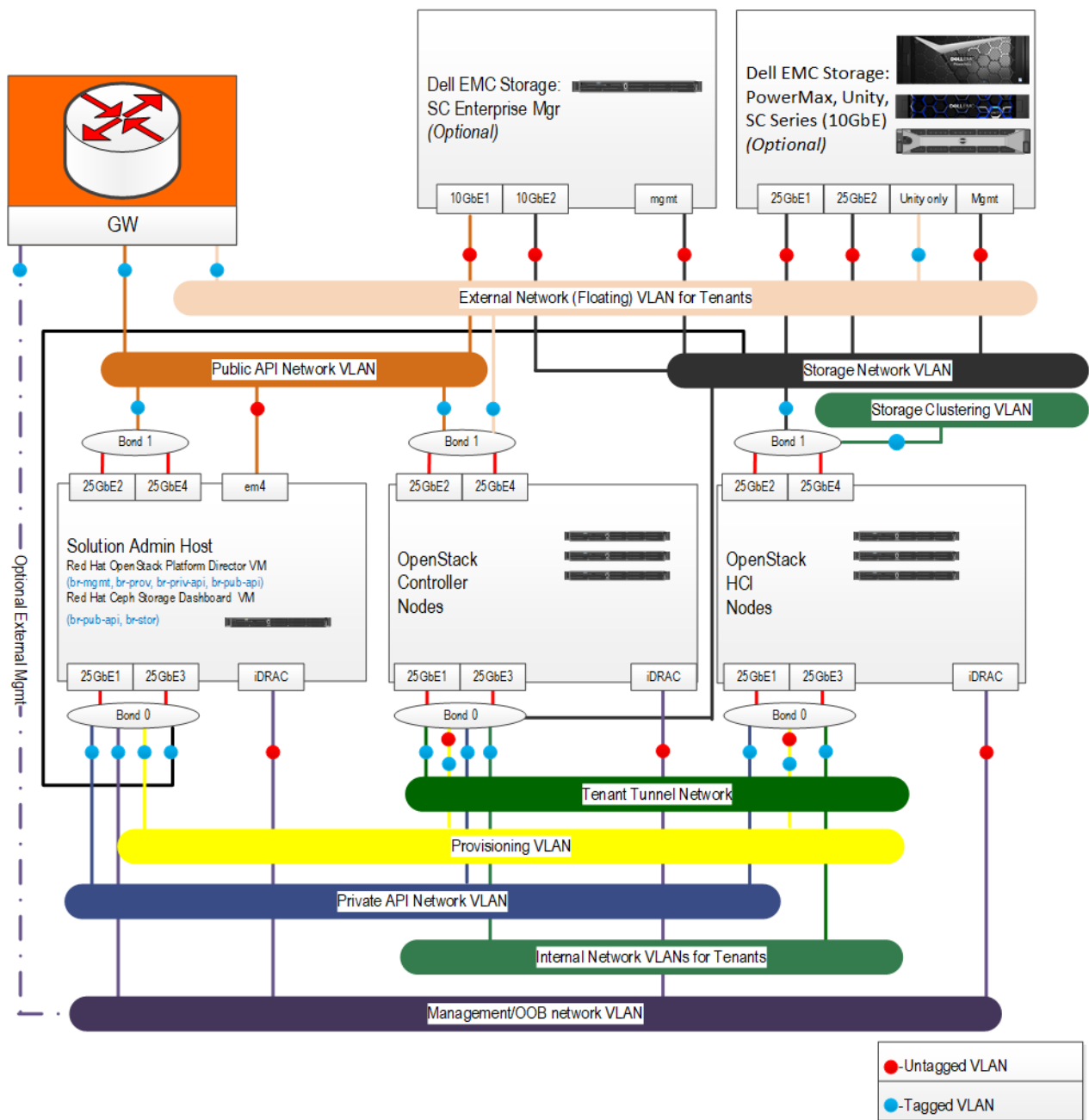
Figure 9: 25GbE/100GbE cluster network logical architecture for HCI profile

For the Edge profile with 25GbE cluster network:
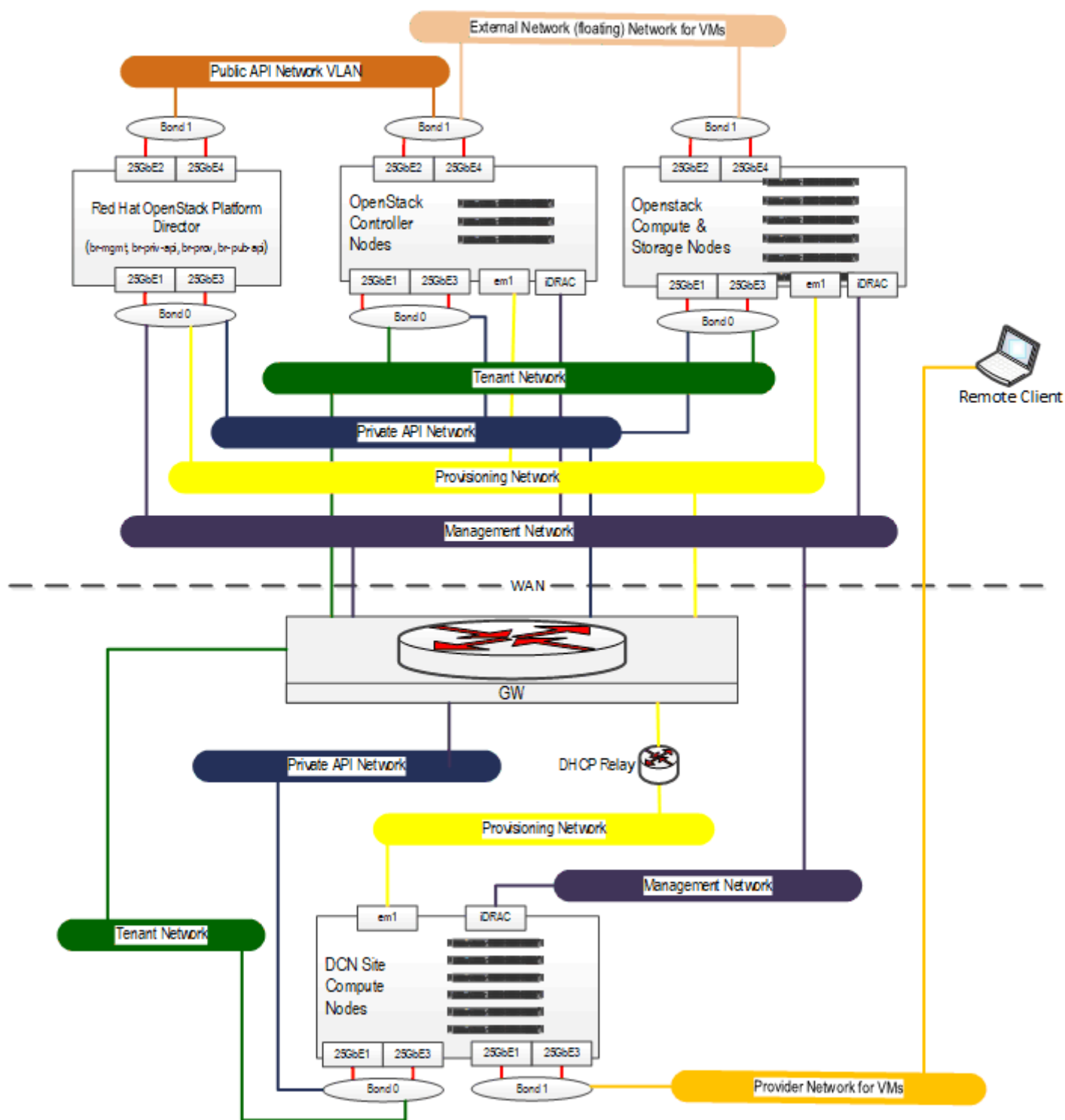
Figure 10: 25GbE cluster network logical architecture for Edge profile

Appendix
A

Bill of Materials

Topics:

*   *Bill of Materials for Dell EMC Ready Architecture for Red Hat OpenStack Platform DCN*
*   *Bill of Materials for Dell EMC PowerEdge R-Series solution - Hyper-Converged Infrastructure*
*   *Bill of Materials for Dell EMC PowerEdge R-Series solution - Mellanox NIC's*
*   *Bill of Materials for Dell EMC PowerEdge R-Series solution - Intel NIC's*
*   *Nodes overview*
*   *Subscriptions and Network Switches in the Solution*

This guide provides Bill of Material information necessary to purchase the proper hardware to deploy the Dell EMC Ready Architecture for Red Hat OpenStack Platform.

Note: For cable, racks, power please contact your Dell EMC support representative.

Bill of Materials for Dell EMC Ready Architecture for Red Hat OpenStack Platform DCN

Up to 20 Dell EMC PowerEdge R-Series compute servers are supported per DCN site in any combination of the following:

- Dell EMC PowerEdge XE2420
- Dell EMC PowerEdge R640
- Dell EMC PowerEdge R740
- Dell EMC PowerEdge R740xd

Edge compute node configuration with 25GbE networking

Table 14: Compute node Dell EMC PowerEdge XE2420

Platform	Dell EMC PowerEdge XE2420
CPU	Intel® Xeon® Intel(R) Xeon(R) Gold 6238 CPU @ 2.10GHz
RAM (minimum)	192GB DDR-4 2933 MHz
NIC Mezzanine	1 x Intel® XXV710 DP OCP 25GbE DA/SFP+
Add-in network	3 x Intel® XXV710 DP 25GbE DA/SFP+ Adapter
Disk	4 x 900GB 10k SAS 12Gbps
Storage controller	PERC H740P controller
RAID	RAID 10

Table 15: Compute node Dell EMC PowerEdge R640

Platform	Dell EMC PowerEdge R640
CPU	Intel® Xeon® Gold 6230 CPU @ 2.10GHz
RAM (minimum)	192GB DDR-4 2933 MHz
LOM	1 x Intel® X710 10GbE SFP+
Add-in network	2 x Intel® XXV710 DP 25GbE DA/SFP+ Adapter
Disk	8 x 600GB 10k SAS 12Gbps
Storage controller	PERC H740 Mini controller
RAID	RAID 10

Table 16: Compute node Dell EMC PowerEdge R740

Platform	Dell EMC PowerEdge R740
CPU	Intel® Xeon® Gold 6230 CPU @ 2.10GHz
RAM (minimum)	192GB DDR-4 2933 MHz
LOM	1 x Intel® X710 10GbE SFP+
Add-in network	2 x Intel® XXV710 DP 25GbE DA/SFP+ Adapter

Platform	Dell EMC PowerEdge R740
Disk	8 x 600GB 10k SAS 12Gbps
Storage controller	PERC H740 Mini controller
RAID	RAID 10

Table 17: Compute node Dell EMC PowerEdge R740xd

Platform	Dell EMC PowerEdge R740xd
CPU	Intel® Xeon® Gold 6230 CPU @ 2.10GHz
RAM (minimum)	192GB DDR-4 2933 MHz
LOM	1 x Intel® X710 10GbE SFP+
Add-in network	2 x Intel® XXV710 DP 25GbE DA/SFP+ Adapter
Disk*	8-24 x 600GB 10k SAS 12Gbps
Storage controller	PERC H740 Mini controller
RAID	RAID 10

Note: * When choosing Dell EMC PowerEdge R740xd as compute nodes, all nodes must have the same number of disks.

Note: Be sure to consult your Dell EMC account representative before changing the recommended hardware configurations.

Bill of Materials for Dell EMC PowerEdge R-Series solution - Hyper-Converged Infrastructure

This chapter provide bill of materials information necessary to purchase the proper hardware to deploy the Dell EMC Ready Architecture for Red Hat OpenStack Platform for Hyper-Converged Infrastructure on Red Hat OpenStack Platform.

The base Dell EMC PowerEdge R-Series HCI solution is comprised of:

- 1 Dell EMC PowerEdge R640 Solution Admin Host (SAH) node
- 3 Dell EMC PowerEdge R640 controller nodes
- 3 Dell EMC PowerEdge R740xd compute/storage nodes

Base configuration - Intel® NIC's

Table 18: Solution Admin Host (SAH) Dell EMC PowerEdge R640

Machine function	SAH node
Platform	Dell EMC PowerEdge R640 (1 qty)
CPU	Intel®Xeon® Gold 6230 CPU @ 2.10GHz
RAM (Minimum)	192GB DDR-4 2933 MHz
LOM	N/A
Add-in Network	2 x Intel® XXV710 DP 25GbE DA/SFP+ Adapter
Disk	8 x 600GB 10k SAS 12Gbps

Machine function	SAH node
Storage Controller	PERC H740 Mini controller
RAID	RAID 10

Table 19: Controller node Dell EMC PowerEdge R640

Machine function	Controller nodes
Platform	Dell EMC PowerEdge R640 (3 qty)
CPU	Intel(R) Xeon(R) Gold 6230 CPU @ 2.10GHz
RAM (Minimum)	192GB DDR-4 2933 MHz
LOM	N/A
Add-in Network	2 x Intel® XXV710 DP 25GbE DA/SFP+ Adapter
Disk	8 x 600GB 10k SAS 12Gbps
Storage Controller	PERC H740 Mini controller
RAID	RAID 10

Table 20: Hyper-Converged Infrastructure node Dell EMC PowerEdge R740xd

Machine function	Storage nodes		
Platform	Dell EMC PowerEdge R740xd (3 qty)		
CPU	Intel®Xeon® Gold 6230 CPU @ 2.10GHz		
RAM (Minimum)	192GB DDR-4 2933 MHz		
LOM	N/A		
Add-in Network	2 x Intel® XXV710 DP 25GbE DA/SFP+ Adapter		
Disk	Option 1: OSD and Journal Drives separate Front Drives: 12 x 2.4TB 10K SAS Mid Bay drive: 4 x 2.4 TB 10K SAS Flex Bay drive: 4 x 400GB SAS SSD	Option 2: OSD and Journal colocated Front drives: 24 x 1.8TB 10K SAS Mid Bay drives: 4 x 1.8.TB 10K SAS Flex Bay drives: 4 x 1.8TB 10K SAS	Option 3: NVMe Front Drives: 12 x 1.2TB Intel NVMe 2 x 1.8 TB 10K SAS
Storage Controller	HBA 330 BOSS Cntrl + 2 M.2 240G,R1,FH	HBA 330 BOSS Cntrl + 2 M.2 240G,R1,FH	PERC H740 Mini controller

Machine function	Storage nodes
RAID	Operating System on BOSS in RAID 1
	Configuration
	Pass through each data disk

> **Note:** NFV features support is experimental in a HCI profile. All other NFV features documented in this Architecture Guide are supported. Be sure to consult your Dell EMC account representative before changing the recommended hardware configurations.

Bill of Materials for Dell EMC PowerEdge R-Series solution - Mellanox NIC's

The base Dell EMC PowerEdge R-Series solution is comprised of:

- 1 Dell EMC PowerEdge R640 Solution Admin Host (SAH) node
- 3 Dell EMC PowerEdge R640 controller nodes
- 3 Dell EMC PowerEdge R640 compute nodes
- 3 Dell EMC PowerEdge R740xd storage nodes

Base configuration - Mellanox NIC's

Table 21: Solution Admin Host (SAH) Dell EMC PowerEdge R640

Machine function	SAH node
Platform	Dell EMC PowerEdge R640 (1 qty)
CPU	Intel®Xeon® Gold 6230 CPU @ 2.10GHz
RAM (Minimum)	192GB DDR-4 2933 MHz
LOM	1 x Mellanox 25GbE 2P ConnectX4LX SFP RNDC
Add-in Network	1 x Mellanox 25GbE 2P ConnectX4LX SFP Adapter
Disk	8 x 600GB 10k SAS 12Gbps
Storage Controller	PERC H740 Mini controller
RAID	RAID 10

Table 22: Controller node Dell EMC PowerEdge R640

Machine function	Controller nodes
Platform	Dell EMC PowerEdge R640 (3 qty)
CPU	Intel®Xeon® Gold 6230 CPU @ 2.10GHz
RAM (Minimum)	192GB DDR-4 2933 MHz
LOM	1 x Mellanox 25GbE 2P ConnectX4LX SFP RNDC

Machine function	Controller nodes
Add-in Network	1 x Mellanox 25GbE 2P ConnectX4LX SFP Adapter 2 x Mellanox 100GbE 2P ConnectX5 SFP Adapter (Used for CSP Profile)
Disk	8 x 600GB 10k SAS 12Gbps
Storage Controller	PERC H740 Mini controller
RAID	RAID 10

Table 23: Compute node Dell EMC PowerEdge R640

Machine function	Compute nodes	
Platform	Dell EMC PowerEdge R640 (3 qty)	
CPU	Intel®Xeon® Gold 6230 CPU @ 2.10GHz	
RAM (Minimum)	192GB DDR-4 2933 MHz	
LOM	1 x Mellanox 25GbE 2P ConnectX4LX SFP RNDC	
Add-in Network	Option 1: for xSP 1 x Mellanox 25GbE 2P ConnectX4LX SFP Adapter	Option 2: for CSP 1 x Mellanox 25GbE 2P ConnectX4LX SFP 2 x Mellanox 100GbE 2P ConnectX5 SFP Adapter
Disk	8 x 600GB 10k SAS 12Gbps	
Storage Controller	PERC H740 Mini controller	
RAID	RAID 10	

Table 24: Storage node Dell EMC PowerEdge R740xd

Machine function	Storage nodes
Platform	Dell EMC PowerEdge R740xd (3 qty)
CPU	Intel(R) Xeon(R) Gold 6230 CPU @ 2.10GHz
RAM (Minimum)	192GB DDR-4 2933 MHz
LOM	1 x Mellanox 25GbE 2P ConnectX4LX SFP RNDC
Add-in Network	1 x Mellanox 25GbE 2P ConnectX4LX SFP Adapter

Machine function	Storage nodes		
Disk	Option 1: OSD and Journal Drives separate Front Drives: 12 x 2.4TB 10K SAS Mid Bay drive: 4 x 2.4 TB 10K SAS Flex Bay drive: 4 x 400GB SAS SSD	Option 2: OSD and Journal colocated Front drives: 24 x 1.8TB 10K SAS Mid Bay drives: 4 x 1.8.TB 10K SAS Flex Bay drives: 4 x 1.8TB 10K SAS	Option 3: NVMe Front Drives: 12 x 1.2TB Intel NVMe 2 x 1.8 TB 10K SAS
Storage Controller	HBA 330 BOSS Cntrl + 2 M.2 240G,R1,FH	HBA 330 BOSS Cntrl + 2 M.2 240G,R1,FH	PERC H740 Mini controller
RAID	Operating System on BOSS in RAID 1 Configuration Pass through each data disk		

Note: Be sure to consult your Dell EMC account representative before changing the recommended hardware configurations.

Bill of Materials for Dell EMC PowerEdge R-Series solution - Intel® NIC's

The base Dell EMC PowerEdge R-Series solution is comprised of:

- 1 Dell EMC PowerEdge R640 Solution Admin Host (SAH) node
- 3 Dell EMC PowerEdge R640 controller nodes
- 3 Dell EMC PowerEdge R740xd compute nodes
- 3 Dell EMC PowerEdge R740xd storage nodes

Base configuration - Intel® NIC's

Table 25: Solution Admin Host (SAH) Dell EMC PowerEdge R640

Machine function	SAH node
Platform	Dell EMC PowerEdge R640 (1 qty)
CPU	Intel®Xeon® Gold 6230 CPU @ 2.10GHz
RAM (Minimum)	192GB DDR-4 2933 MHz
LOM	N/A
Add-in Network	2 x Intel® XXV710 DP 25GbE DA/SFP+ Adapter
Disk	8 x 600GB 10k SAS 12Gbps

Machine function	SAH node
Storage Controller	PERC H740 Mini controller
RAID	RAID 10

Table 26: Controller node Dell EMC PowerEdge R640

Machine function	Controller nodes
Platform	Dell EMC PowerEdge R640 (3 qty)
CPU	Intel(R) Xeon(R) Gold 6230 CPU @ 2.10GHz
RAM (Minimum)	192GB DDR-4 2933 MHz
LOM	N/A
Add-in Network	2 x Intel® XXV710 DP 25GbE DA/SFP+ Adapter
Disk	8 x 600GB 10k SAS 12Gbps
Storage Controller	PERC H740 Mini controller
RAID	RAID 10

Table 27: Compute node Dell EMC PowerEdge R740

Machine function	Compute nodes	
Platform	Dell EMC PowerEdge R740 (3 qty)	
CPU	Intel®Xeon® Gold 6230 CPU @ 2.10GHz	
RAM (Minimum)	192GB DDR-4 2933 MHz	
LOM	N/A	
Add-in Network	Option 1: for xSP 2 x Intel® XXV710 DP 25GbE DA/SFP+ Adapter	Option 2: for CSP 4 x Intel® XXV710 DP 25GbE DA/SFP+ Adapter
Disk	8 x 600GB 10k SAS 12Gbps	
Storage Controller	PERC H740 Mini controller	
RAID	RAID 10	

Table 28: Storage node Dell EMC PowerEdge R740xd

Machine function	Storage nodes
Platform	Dell EMC PowerEdge R740xd (3 qty)
CPU	Intel®Xeon® Gold 6230 CPU @ 2.10GHz
RAM (Minimum)	192GB DDR-4 2933 MHz
LOM	N/A
Add-in Network	2 x Intel® XXV710 DP 25GbE DA/SFP+ Adapter

Machine function	Storage nodes		
Disk	Option 1: OSD and Journal Drives separate Front Drives: 12 x 2.4TB 10K SAS Mid Bay drive: 4 x 2.4 TB 10K SAS Flex Bay drive: 4 x 400GB SAS SSD	Option 2: OSD and Journal colocated Front drives: 24 x 1.8TB 10K SAS Mid Bay drives: 4 x 1.8.TB 10K SAS Flex Bay drives: 4 x 1.8TB 10K SAS	Option 3: NVMe Front Drives: 12 x 1.2TB Intel NVMe 2 x 1.8 TB 10K SAS
Storage Controller	HBA 330 BOSS Cntrl + 2 M.2 240G,R1,FH	HBA 330 BOSS Cntrl + 2 M.2 240G,R1,FH	PERC H740 Mini controller
RAID	Operating System on BOSS in RAID 1 Configuration Pass through each data disk		

Note: When using Intel® NIC's, OpenvSwitch (OVS) hardware offloading is not supported. All other NFV features documented in this Architecture Guide are supported. Be sure to consult your Dell EMC account representative before changing the recommended hardware configurations.

Nodes overview

The minimum hardware needed is:

- 1 Solution Admin Host (SAH)
- 3 Controller nodes
- 3 Compute nodes
- 3 Storage servers

Please consult with your Dell EMC sales representative to ensure proper preparation and submission of your hardware and software orders.

Subscriptions and Network Switches in the Solution

A Dell EMC sales representative will determine the correct software subscriptions needed for the Dell EMC Ready Architecture for Red Hat OpenStack Platform and Dell EMC Networking OS10 subscriptions.

Required subscriptions:

- Red Hat OpenStack Platform
- Red Hat Ceph storage
- Dell EMC Networking OS10
- Red Hat Satellite - Optional

> **Note:** Please contact your Dell EMC sales representative.

Default network switch - Dell EMC Networking S3048-ON switch

Table 29: S3048-ON switch

Product	Description
S3048-ON	48 line-rate 1000BASE-T ports, 4 line-rate 10GbE SFP+ ports (1 qty)
Redundant power supplies	AC Power Supply or DC Power Supply
Fans	Fan Module I/O Panel to PSU Airflow or Fan Module PSU to I/O Panel Airflow
Validated operating systems	Dell EMC Networking OS10

Dell EMC Networking S4048-ON optional switch

Table 30: S4048-ON switch

Product	Description
S4048-ON	48x 10GbE SFP+, 6x QSFP+ (1 qty) - optional
Redundant power supplies	AC Power Supply or DC Power Supply
Fans	Fan Module I/O Panel to PSU Airflow or Fan Module PSU to I/O Panel Airflow
Validated operating systems	Dell EMC Networking OS10

Dell EMC Networking S5232F-ON switch

Table 31: S5232F-ON

Product	Description
S5232F-ON	100GbE, 40GbE, and 25GbE (2 qty)
Redundant power supplies	AC Power Supply or DC Power Supply

Product	Description
Fans	Fan Module I/O Panel to PSU Airflow or Fan Module PSU to I/O Panel Airflow
Validated operating systems	Dell EMC Networking OS10

Dell EMC Networking S5224F-ON

Table 32: S5224F-ON

Product	Description
S5224F-ON	100GbE, 40GbE, and 25GbE (2 qty)
Redundant power supplies	AC Power Supply or DC Power Supply
Fans	Fan Module I/O Panel to PSU Airflow or Fan Module PSU to I/O Panel Airflow
Validated operating systems	Dell EMC Networking OS10

Appendix
B

References

Additional information can be found at the Knowledge Base for Dell EMC Ready and Reference Architectures for OpenStack Platform. This page contains the latest copy of this guide as well as associated solution briefs and white papers. It also contains information on related solutions, including network edge, storage, and Ironic bare metal solutions. Please visit *https://www.dell.com/support/article/us/ en/19/sln310368/dell-emc-ready-architecture-for-red-hat-openstack- platform?lang=en*. Or you can e-mail *openstack@dell.com*.

For more information on Dell EMC Service Provider Solutions, visit *https://www.dellemc.com/en-us/service-providers/index.htm*

Glossary

API

Application Programing Interface is a specification that defines how software components can interact.

BMC/iDRAC Enterprise

Baseboard management controller. An on-board microcontroller that monitors the system for critical events by communicating with various sensors on the system board, and sends alerts and log events when certain parameters exceed their preset thresholds.

BOSS

The Boot Optimized Storage Solution (BOSS) enables customers to segregate operating system and data on server-internal storage. This is helpful in the Hyper-Converged Infrastructure (HCI) and Software Defined Storage (SDS) arenas, to separate operating system drives from data drives, and implement hardware RAID mirroring (RAID1) for OS drives.

CDH

Cloudera Distribution for Apache Hadoop

Cloud computing

See *http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf*

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Cluster

A set of servers dedicated to OpenStack that can be attached to multiple distribution switches.

Compute node

The hardware configuration that best supports the hypervisor server or Nova compute roles.

DevOps

Development Operations (DevOps) is an operational model for managing data centers using improved automated deployments, shortened lead times between fixes, and faster mean time to recovery. See *https://en.wikipedia.org/wiki/DevOps*.

DIMM

Dual In-line Memory Module

DNS

The domain name system (DNS) defines how Internet domain names are located, and translated into Internet Protocol (IP) addresses.

FQDD

A fully qualified device descriptor (FQDD) is a method used to describe a particular component within a system or subsystem, and is used for system management and other purposes.

FQDN

A fully qualified domain name (FQDN) is the portion of an Internet Uniform Resource Locator (URL) that fully identifies the server to which an Internet request is addressed. The FQDN includes the second-level domain name, such as "dell.com", and any other levels as required.

GUI

Graphical User Interface - A visual interface for human interaction with the software, taking inputs and generating easy to understand visual outputs.

Hypervisor

Software that runs virtual machines (VMs).

IaaS

Infrastructure as a Service.

Infrastructure node

Systems that handle the control plane and deployment functions.

ISV

Independent Software Vendor.

JBOD

Just a Bunch of Disks

LAG

Link Aggregation Group.

LOM

LAN on motherboard.

LVM

Logical Volume Management

ML2

The Modular Layer 2 plug-in is a framework that allows OpenStack to utilize different layer 2 networking technologies.

NFS

The Network File System (NFS) is a distributed filesystem that allows a computer user to access, manipulate, and store files on a remote computer, as though they resided on a local file directory.

NIC

Network Interface Card

Node

One of the servers in the cluster.

NUMA

Non-Uniform Memory Access

Overcloud

The functional cloud that is available to run guest VMs and workloads.

Pod

An installation comprised of three racks, and consisting of servers, storage, and networking.

REST

REST - Representational State Transfer (also ReST). Relies upon stateless, client-server, cacheable communications protocol to access the API.

RHOSP

Red Hat OpenStack Platform

RPC

Remote Procedure Call

SAH

The Solution Admin Host (SAH) is a physical server that supports VMs for the Undercloud machines needed for the cluster to be deployed and operated.

SDS

Software-defined storage (SDS) is an approach to computer data storage in which software is used to manage policy-based provisioning and management of data storage, independent of the underlying hardware.

SDN

Software-defined Network (SDN) is where the software will define, create, use and destroy different networks as needed.

Stamp

A stamp is the collection of all servers and network switches in the solution.

Storage Node

The hardware configuration that best supports SDS functions such as Red Hat Ceph storage.

ToR

Top-of-rack switch/router.

U

U used in the definition of the size of server, example 1U or 2U. A "U" is a unit of measure equal to 1.75 inches in height.

Undercloud

The Undercloud is the system used to control, deploy, and monitor the Overcloud - it is a single node OpenStack deployment completely under the administrators control. The Undercloud is *not* HA configured.

VLT

A Virtual Link Trunk (VLT) is the combined port channel between an attached device (ToR switch) and the VLT peer switches.

VLTi

A Virtual Link Trunk Interconnect (VLTi) is an interconnect used to synchronize states between the VLT peer switches. Both endpoints must be the same speed, i.e. 40Gb $\rightarrow$ 40Gb; 1G interfaces are not supported.

VM

Virtual Machine - a simulation of a computer system.