Dell EMC Ready Bundle for Red Hat OpenStack Platform

Dell EMC PowerEdge R-Series Architecture Guide

Version 10.0



Dell EMC Converged Platforms and Solutions

Contents

List of Figures....iv

List of Tables....v

Trademarks....6
Glossary....7
Notes, Cautions, and Warnings....12

Chapter 1: Overview....13
OpenStack *Newton*....14
Hardware Options....14
Networking and Network Services....14
Taxonomy....15
Red Hat OpenStack Platform 10....17
Key Benefits....17

Chapter 2: OpenStack Architecture....18
OpenStack Components....19

Chapter 3: Server Options....21
Dell EMC PowerEdge R630 Server....22
Dell EMC PowerEdge R730xd Servers....22
Base Hardware Configurations....22
Configuration Notes....24
Optional Compute Servers....24
Dell EMC PowerEdge R430....24
Dell EMC PowerEdge R730....25
Dell EMC PowerEdge R730xd....25

Chapter 4: Storage Options....26
Storage Options Overview....27
Local Storage....27
Red Hat Ceph Storage....28
Optional Dell Storage....28
Dell Storage Arrays....28

Chapter 5: Network Architecture....30
Network Architecture Overview....31
Infrastructure Layouts....31
Network Components....31
Server Nodes....31

Access Switch or Top of Rack (ToR)..... 32
Aggregation Switches.....32
Core..... 33
Layer-2 and Layer-3 Switching..... 33
VLANs.....33
Out of Band Management Network..... 34
Dell EMC OpenSwitch Solution..... 34

Chapter 6: Operational Notes.....35
Backup/Recovery.....36
High Avaibility.....36
Service Layout.....36
Deployment Overview..... 37

Chapter 7: Solution Bundle..... 39
Solution Bundle Overview.....40
Solution Bundle Rack Layout.....40
Solution Bundle Network Configuration..... 42
Solution Admin Host (SAH) Networking..... 45
Optional Solution Configurations.....46
Optional Dell Storage with the Solution Bundle.....46
Solution Bundle Expansion..... 48
Rack 1..... 48
Rack 2..... 49
Rack 3..... 49
Larger Configurations.....49

Appendix A: Update History..... 50
Initial Release.....51
Version 1..... 51
Version 2..... 51
Version 3..... 51
Version 4..... 52
Version 5..... 52
Version 6..... 52

Appendix B: References.....54
To Learn More..... 55

List of Figures

Figure 1: OpenStack Taxonomy..... 16

Figure 2: Dell EMC PowerEdge R730xd Servers - 2.5" and 3.5" Chassis Options.....22

Figure 3: Solution Bundle and Red Hat Ceph Storage Cluster..... 41

Figure 4: Cluster Network Logical Architecture with Optional Dell Storage PS Series.....43

Figure 5: Solution Admin Host Internal Network Fabric..... 46

Figure 6: Solution with Optional Dell Storage.....47

List of Tables

Table 1: Deployed Core Services....15
Table 2: Deployed Non-Core Services....15
Table 3: Optional Services....15
Table 4: OpenStack Core Components....19
Table 5: Non-Core Deployed Components....20
Table 6: Controller Node Hardware Configurations – Dell EMC PowerEdge R630....22
Table 7: Compute Node Hardware Configurations – Dell EMC PowerEdge R630....23
Table 8: Solution Admin Host Hardware Configurations – Dell EMC PowerEdge R630....23
Table 9: Storage Node Hardware Configurations – Dell EMC PowerEdge R730xd....23
Table 10: ToR Switches....31
Table 11: Channel Bonding Modes Supported....32
Table 12: Overcloud: Node Type to Services....36
Table 13: OpenStack Node Type to Network 802.1q Tagging....43
Table 14: Storage Node Type to Network 802.1q Tagging....44

Trademarks

Copyright © 2014-2017 Dell Inc. or its subsidiaries. All rights reserved.

Microsoft® and Windows® are registered trademarks of Microsoft Corporation in the United States and/or other countries.

Red Hat®, Red Hat Enterprise Linux®, and Ceph are trademarks or registered trademarks of Red Hat, Inc., registered in the U.S. and other countries. Linux® is the registered trademark of Linus Torvalds in the U.S. and other countries. Oracle® and Java® are registered trademarks of Oracle Corporation and/or its affiliates.

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.

The Midokura® name and logo, as well as the MidoNet® name and logo, are registered trademarks of Midokura SARL.

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.

Bundle

A customer-orderable solution that consists of:

- All server, network, and storage hardware needed to install and operate the solution as outlined
- All necessary solution software licenses needed to install and operate the solution as outlined

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

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.

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.

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 → 40Gb; 1G interfaces are not supported.

VM

Virtual Machine - a simulation of a computer system.

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:

- *OpenStack Newton*
- *Hardware Options*
- *Networking and Network Services*
- *Taxonomy*
- *Red Hat OpenStack Platform 10*

An OpenStack® based cloud is now a common need by many orgainizations and Dell EMC 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
  - Compute nodes
  - Infrastructure nodes
  - Storage nodes
- Network design
- Software layout
- Offers suggestion for other system configurations

OpenStack *Newton*

This Architecture Guide is based on the 14th OpenStack release codename *Newton*. This release now offers greater scalability, resiliency and user experience. Newton was developed by over 2,581 developers employed by more than 309 organizations. Please see *http://www.openstack.org/software/newton*.

Dell EMC and Red Hat designed this Architecture Guide to make it easy for Dell EMC Ready Bundle for Red Hat OpenStack Platform customers to build their own operational readiness cluster and design their initial offerings, using the current releases. Dell EMC and Red Hat provide the support and services customers need to stand up production-ready OpenStack clusters.

The code base for Red Hat OpenStack Platform is evolving at a very rapid pace. Please see *https://access.redhat.com/site/support/policy/updates/OpenStack/platform* for more information.

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
- Infrastructure nodes
- Storage nodes

Dell EMC recommends starting with OpenStack software using components from this Architecture Guide 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, 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 RHOSP 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 OpenVswitch with the VLAN option.

Networking includes:

- Core and layered networking capabilities
- 10GbE networking
- NIC bonding
- Redundant trunking top-of-rack (ToR) switches into core routers

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

See *Network Architecture* on page 30 for guidelines. Detailed designs are available through Dell EMC consulting services.

Taxonomy

The Dell EMC Ready Bundle for Red Hat OpenStack Platform is built using the following core OpenStack components, as delivered in the Red Hat OpenStack Platform. See *Table 1: Deployed Core Services* on page 15 and *Table 2: Deployed Non-Core Services* on page 15.

Table 1: Deployed Core Services

Component	Code Name
Block Storage	Cinder with Red Hat Ceph Storage and Dell Storage PS Series or SC Series
Compute	Nova
Identity	Keystone
Image Service	Glance
Networking	Neutron

Table 2: Deployed Non-Core Services

Component	Code Name
Bare Metal Provisioning	Ironic and TripleO
Dashboard	Horizon
Orchestration	Heat
Telemetry	Ceilometer
Validation Testing	Tempest

Caution: Before using Tempest, review the Tempest documentation at *http://docs.openstack.org/developer/tempest/*.

There are several optional[1] OpenStack components that are available but not part of the base solution. See *Table 3: Optional Services* on page 15.

Table 3: Optional Services

Database	Trove
Data Processing	Sahara
Database	Trove
DNS as a Service	Designate
File Share Service	Manila
Key Management	Barbican

The taxonomy presented in *Figure 1: OpenStack Taxonomy* on page 16 reflects infrastructure components, and OpenStack-specific components, that are under active development by the community, Dell EMC, and Red Hat. The taxonomy reflects that there are two sides for cloud users:

[1] Available through a custom Services engagement.

- Site-specific infrastructure
- Standards-based API (shown in pink) interactions

The standards-based APIs are the same between all OpenStack deployments, and let customers and vendor ecosystems operate across multiple clouds. The site-specific infrastructure combines open and proprietary software, Dell EMC hardware, and operational processes to deliver cloud resources as a service.

The implementation choices for each cloud infrastructure are highly specific to the requirements of each site. Many of these choices can be standardized and automated using the tools in this Architecture Guide. Conforming to best practices helps reduce operational risk by leveraging the accumulated experience of Dell EMC, Red Hat and the broader OpenStack community.

Red Hat OpenStack Platform Director is used to deploy the solution (Overcloud) from the Undercloud. The Undercloud is a single server that runs a subset of OpenStack services used to deploy, manage (hardware and software) and update the Overcloud servers. In the Dell EMC Ready Bundle for Red Hat OpenStack Platform the OpenStack Controllers, Computes and Red Hat Ceph Storage servers comprise the Overcloud servers.

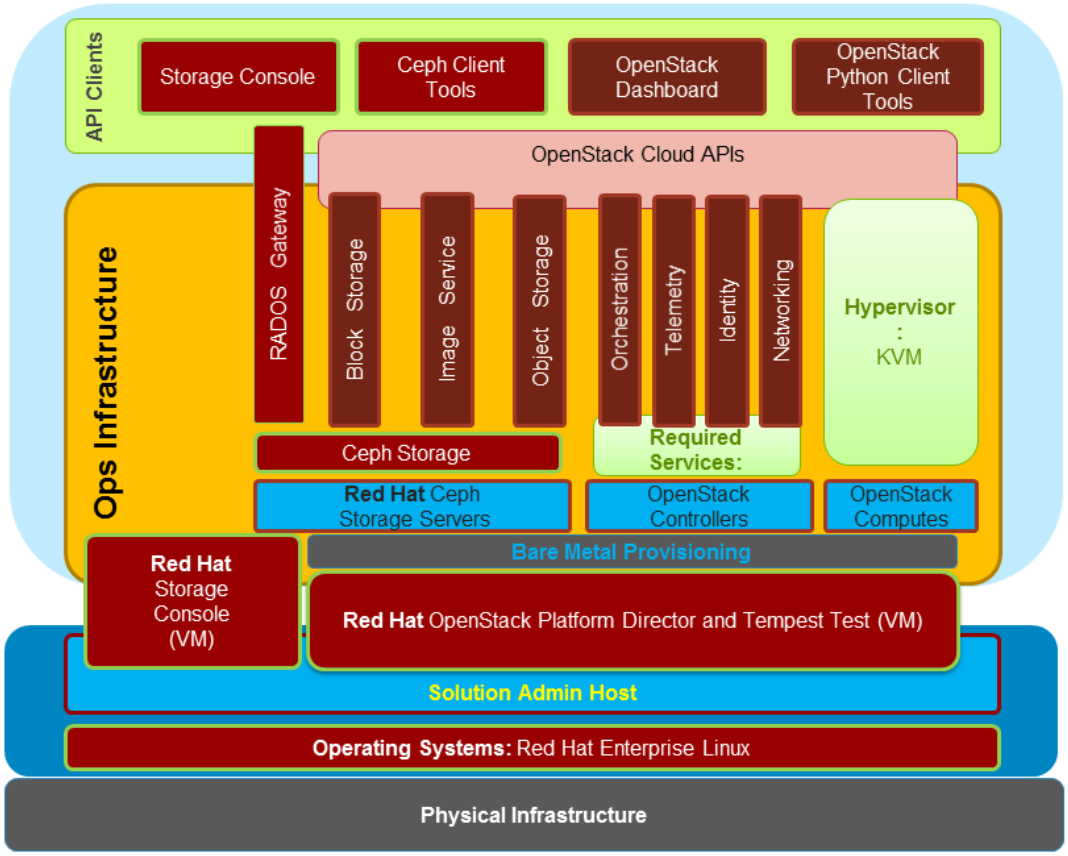


Figure 1: OpenStack Taxonomy

Red Hat OpenStack Platform 10

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.

The current Red Hat system is based on OpenStack Newton, and 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
- Web browser-based GUI for both users and administration.

The Red Hat OpenStack Platform IaaS cloud is implemented by a collection of interacting services that control its computing, storage, and networking resources. The cloud is managed using a web-based interface which allows administrators to control, provision, and automate OpenStack resources. Additionally, the OpenStack infrastructure is facilitated through an extensive API, which is also available to end users of the cloud.

Red Hat OpenStack Platform is purposely designed recognizing the unique dependencies that OpenStack has upon its underlying Linux® installation. Red Hat uniquely co-engineers and integrates Red Hat OpenStack technology with Red Hat Enterprise Linux Server 7.3, ensuring a stable, production-ready cloud platform. Version 10 boasts all of the core features and functions of the community Newton release and adds some additional innovations by Red Hat, resulting in a hardened, stable cloud platform.

Key Benefits

- 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 Red Hat Enterprise Linux; 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 supported by Red Hat for a three (3) year “production phase” lifecycle, well beyond the six-month release cycle of unsupported, community OpenStack.
- 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 Platform 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.

Chapter 2

OpenStack Architecture

Topics:

- *OpenStack Components*

While OpenStack has many configurations and capabilities, the primary components for the Red Hat OpenStack Platform 10 (Newton), as defined in *Taxonomy* on page 15.

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

OpenStack Components

The following component descriptions are from the OpenStack Foundation website. Extensive documentation for the OpenStack components is available at *http://docs.openstack.org/*.

Table 4: OpenStack Core Components

Function	Code Name	Description
Bare Metal Provisioning	Ironic	*Openstack Bare Metal Provisioning* aims to provision bare metal machines instead of virtual machines.
Block Storage	Cinder	*OpenStack Block Storage* provides persistent block level storage devices for use with OpenStack compute instances. The block storage system manages the creation, attaching, and detaching of the block devices to servers. Block storage volumes are fully integrated into OpenStack Compute and the Dashboard enabling cloud users to manage their own storage needs.
Compute	Nova	*OpenStack cloud operating system* enables enterprises and service providers to offer on-demand computing resources, by provisioning and managing large networks of virtual machines. Compute resources are accessible via APIs for developers or users and web interfaces for administrators and users.
Identity	Keystone	*Identity Service* provides a central directory of users mapped to the OpenStack services they can access. It acts as a common authentication system across the cloud operating system and can integrate with existing backend directory services.
Image Service	Glance	*OpenStack Image Service* provides discovery, registration, and delivery services for virtual disk images. The Image Service API server provides a standard REST interface for querying information about virtual disk images stored in a variety of back-end stores.
Networking	Neutron	*OpenStack Neutron* is a pluggable, scalable and API-driven system for managing networks and IP addresses. Like other aspects of the cloud operating system, it can be used by administrators and users to increase the value of existing datacenter assets.
Orchestration	Heat	*OpenStack Orchestration* is a template-driven engine that enables application developers to describe and automate the deployment of infrastructure. The flexible template language can specify compute, storage and networking configurations as well as detailed post-deployment activity to automate the full provisioning of infrastructure as well as services and applications.
Orchestration	TripleO	*OpenStack TripleO* is an OpenStack program that utilizes OpenStack's own cloud facilities to install and operate OpenStack clouds. This builds on Nova, Neutron, and Heat to automate small clouds to datacenter roll-outs.

Table 5: Non-Core Deployed Components

Function	Code Name	Description
Dashboard/ Portal	Horizon	*OpenStack Dashboard* provides administrators and users a graphical interface to access, provision and automate cloud-based resources. The extensible design makes it easy to plug in and expose third party products and services.
Database	Trove	*Trove* is Database as a Service for OpenStack.
Data Processing	Sahara	*Sahara* provides users with a simple means to provision data processing frameworks, such as Apache Hadoop.
DNS as a Servers	Designate	*Designate* is a a multi-tenant DNS as a service for OpenStack.
File Share as a Service	Manila	*Manila* implements a method to provide coordinated asccess to shared or distibuted file system.
Key Management	Barbican	*Barbican* is the OpenStack Key Manager Service, providiing secure storage, provisioning and management of secrete data.
Telemetry	Ceilometer	*OpenStack Telemetry* aggregates usage and performance data across the services deployed in an OpenStack cloud. This powerful capability provides visibility and insight into the usage of the cloud across dozens of data points and allows cloud operators to view metrics globally or by individual deployed resources.
Validation Testing	Tempest	*Tempest* is a set of integration tests to be run against a live OpenStack cluster. Tempest includes batteries of tests for OpenStack API validation, various scenarios, and other specific tests useful in validating an OpenStack deployment.

Chapter 3

Server Options

Topics:

- *Dell EMC PowerEdge R630 Server*
- *Dell EMC PowerEdge R730xd Servers*
- *Base Hardware Configurations*
- *Configuration Notes*
- *Optional Compute Servers*

The base validated Solution supports the Dell EMC PowerEdge R630 and Dell EMC PowerEdge R730xd Server lines. See *Optional Compute Servers* on page 24 for other options.

Note: Detailed parts lists are included in the Dell EMC Ready Bundle for Red Hat OpenStack Platform Hardware Deployment Guide .

Dell EMC PowerEdge R630 Server

The Dell EMC PowerEdge R630 server is a hyper-dense, two-socket, 1U rack server.

With computing capability previously only seen in 2U servers, the ultra-dense Dell EMC PowerEdge R630 two-socket 1U rack server delivers an impressive solution for cloud solutions, virtualization environments, large business applications, or transactional databases.

The Dell EMC PowerEdge R630 server is versatile and highly configurable for a variety of solutions, supporting the latest Intel® Xeon® processor E5-2600 v4 product family, 24 DIMMs of high-performance DDR4 memory and a broad range of local storage options.

Dell EMC PowerEdge R730xd Servers

The Dell EMC PowerEdge R730xd is an exceptionally flexible and scalable two-socket 2U rack server, that delivers high performance processing and a broad range of workload-optimized local storage possibilities, including hybrid tiering.

Designed with an incredible range of configurability, the Dell EMC PowerEdge R730xd meets the needs of many different storage workloads with the latest Intel® Xeon® processor E5-2600 v4 product family, 24 DIMMs of high-performance DDR4 memory, and a broad range of local storage options.

Both 2.5" and 3.5" chassis options are available and have been validated.



Figure 2: Dell EMC PowerEdge R730xd Servers - 2.5" and 3.5" Chassis Options

Base Hardware Configurations

Table 6: Controller Node Hardware Configurations – Dell EMC PowerEdge R630

Machine Function	Solution Bundle Controller Nodes
Platform	Dell EMC PowerEdge R630
CPU	2 x E5-2650v4 (12-core)
RAM (Minimum)	128 GB

Machine Function	Solution Bundle Controller Nodes
LOM	2 x 1Gb, 2 x Intel X520 10Gb SFP+
Add-in Network	1 x Intel X520 DP 10Gb DA/SFP+
Disk	8 x 600GB 10K SAS 12Gbps
Storage Controller	PERC H730
RAID	RAID 10

Table 7: Compute Node Hardware Configurations – Dell EMC PowerEdge R630

Machine Function	Solution Bundle Compute Nodes
Platform	Dell EMC PowerEdge R630
CPU	2 x E5-2650v4 (12-core)
RAM (Minimum)	128 GB
LOM	2 x 1Gb, 2 x Intel X520 10Gb SFP+
Add-in Network	1 x Intel X520 DP 10Gb DA/SFP+
Disk	8 x 600GB 10k SAS 12 Gbps
Storage Controller	PERC H730
RAID	RAID 10

Table 8: Solution Admin Host Hardware Configurations – Dell EMC PowerEdge R630

Machine Function	Solution Bundle Infrastructure Nodes
Platform	Dell EMC PowerEdge R630
CPU	2 x E5-2650v4 (12-core)
RAM (Minimum)	64 GB
LOM	2 x 1Gb, 2 x Intel X520 10Gb SFP+
Add-in Network	1 x Intel X520 DP 10Gb DA/SFP+
Disk	8 x 600GB 10k SAS 12 Gbps
Storage Controller	PERC H730
RAID	RAID 10

Table 9: Storage Node Hardware Configurations – Dell EMC PowerEdge R730xd

Machine Function	Solution Bundle Storage Nodes
Platforms	Dell EMC PowerEdge R730xd
CPU	2 x E5-2650v4 (12-core)
RAM (Minimum)	64 GB
LOM	1 x 1Gb, 2 x Intel X520 10Gb
Add-in Network	2 x Intel X520 DP 10Gb DA/SFP+

Machine Function	Solution Bundle Storage Nodes
Disk	Flex Bay: 2 X 300GB 15K 2.5-inch (OS) Front Drives: 3 X 400GB SSD 12 x 2TB or 4TB NL SAS 7.2K 3.5-inch
Storage Controller	PERC H730
RAID	RAID 1 (operating system) Pass through SSD Pass through each data disk

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

Configuration Notes

The Dell EMC Ready Bundle for Red Hat OpenStack Platform Hardware Deployment Guide contains the full bill of materials (BOM) listing for the Dell EMC PowerEdge R630 and Dell EMC PowerEdge R730xd server configurations.

The Dell EMC PowerEdge R630 and Dell EMC PowerEdge R730xd configurations are used with 10GbE networking. To ensure that the network is HA ready, an additional network card is required in each node. Refer to the Dell EMC Ready Bundle for Red Hat OpenStack Platform Hardware Deployment Guide, which outlines the supported cards and includes them as part of the solution.

Caution: You must ensure that the firmware on all servers is up to date. Otherwise, unexpected results may occur. See *Tested BIOS and Firmware* in the Dell EMC Ready Bundle for Red Hat OpenStack Platform Hardware Deployment Guide.

Optional Compute Servers

Two additional servers have been validated for specific roles in the solution:

- *Dell EMC PowerEdge R430* on page 24
- *Dell EMC PowerEdge R730* on page 25
- *Dell EMC PowerEdge R730xd* on page 25

Dell EMC PowerEdge R430

The Dell EMC PowerEdge R430 is an option for the OpenStack Computes. It is a powerful, compact computing system that can support up to:

- 384 GB RAM
- 10 x 2.5 hot-plug drives
- 2PCIe 3.0 slots

This server is used in environments where power and cooling constraints need to be met. The smaller form factor (24" depth), and low power usage, ensure that the server can be utilized in your infrastructure with minimal impact.

Please contact your Dell EMC sales representative for options and configurations.

Dell EMC PowerEdge R730

The Dell EMC PowerEdge R730 is an option for the OpenStack Compute nodes. With a 2U form factor, the Dell EMC PowerEdge R730 can support:

- Up to 7 PCIe cards
- HDD options include:
  - Up to 9TB in 2.5" hard drives, or
  - Up to 48TB in 3.5" hard drives

This server can be used where a large ephemeral storage pool is desired, or if there is a need for additional PCIe cards.

Please contact your Dell EMC sales representative for options and configurations.

Dell EMC PowerEdge R730xd

The Dell EMC PowerEdge R730xd is an option for the OpenStack Compute nodes. With a 2U form factor, the Dell EMC PowerEdge R730xd can support:

- Up to 7 PCIe cards
- HDD options include:
  - Up to 36 TB in 2.5" hard drives, or
  - Up to 96 TB in 3.5" hard drives



Note: All hard disk drives must be the same make, model, and size for this configurations, also Flex Bays should not be populated.

This server can be used where a large ephemeral storage pool is desired, or if there is a need for additional PCIe cards.

Please contact your Dell EMC sales representative for options and configurations.

Chapter 4

Storage Options

Topics:

- *Storage Options Overview*
- *Local Storage*
- *Red Hat Ceph Storage*
- *Optional Dell Storage*

OpenStack has several storage services, including:

- Cinder
- Glance
- Swift [2]

Together these services provide virtual machines (VMs) with block, image, and object storage. In turn, the services employ block 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] Availabe through a Ceph cluster with the Swift API enabled or through a custom 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.

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 Bundle 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, or Dell Networking PS Series or SC Series 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 *RBD*. During the planning process, decisions can be made to place ephemeral on local or share, it is recommended that a shared backend is used that will allow live migration.

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

The Dell EMC Ready Bundle 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 27
- *Red Hat Ceph Storage* on page 28
- *Optional Dell Storage* on page 28

In addition, the Dell EMC Ready Bundle for Red Hat OpenStack Platform provides Red Hat Ceph Storage as an alternate implementation for Glance, Block, and Object stores.

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 7: Compute Node Hardware Configurations – Dell EMC PowerEdge R630* on page 23) using the eight (8) 600 GB disks in a RAID 10, there will be approximately 2 TB of storage available.

Red Hat Ceph Storage

The Dell EMC Ready Bundle 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 stores data by running multiple Object Storage Daemons (OSD) on each Storage node. 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 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 R730xd system with 16 drive bays available for Red Hat Ceph Storage, 3 bays could be used for SSD journal drives and 12 bays for HDD data drives, leaving 1 bay open for a spare drive. This is based upon the current industry guideline of a 4:1 HDD to SSD journal ratio.

Optional Dell Storage

Two supported Dell Storage options have been validated for the solution:

- Dell Storage PS Series
- Dell Storage SC Series

Dell Storage Arrays

Dell Storage PS Series and SC Series storage arrays are designed to provide simplified deployment and administration of consolidated storage environments. Dell Storage PS Series and SC Series storage systems are self-optimized, utilizing embedded load-balancing technologies that react to workload demands.

The core capabilities of Dell Storage PS Series and SC Series storage products include comprehensive software components and host integration, which simplify administrative tasks and assist with storage management. Application-layer integration with OpenStack enables Cinder to:

- Provision and manage volumes on Dell Storage PS Series and SC Series storage
- Utilize SAN-based snapshots for protection capability

Additionally, the Dell Storage SC Series:

- Provides automatic data tiering, based upon usage.
- Requires an additional server to be added to the cluster to support the Dell Storage Enterprise Manager. These solutions may require a service motion in order to implement them.

Note: iSCSI connections are only supported with the Dell Storage PS Series or SC Series. Fibre Channel connections are not supported in this Architecture Guide.

Chapter 5

Network Architecture

Topics:

- *Network Architecture Overview*
- *Infrastructure Layouts*
- *Network Components*

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

Network Architecture Overview

The Dell EMC Ready Bundle for Red Hat OpenStack Platform with Red Hat OpenStack Platform uses either of the recommended Ethernet switches, presented in *Table 10: ToR Switches* on page 31, as the top-of-rack connectivity to all OpenStack-related nodes.

Note: If already in your environment, you can substitute the alternate switches for the recommended switches.

Table 10: ToR Switches

Recommended Switches	Alternate Switches
Dell Networking S3048-ON 1/10-Gigabit	Dell Networking S55 1/10-Gigabit
Dell Networking S4048-ON 10-Gigabit or S6000-ON 40-Gigabit	Dell Networking S4810 10-Gigabit

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, top-of-rack (ToR) 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. Configurations for 1GbE and 10GbE are included in this Architecture Guide. The following component blocks make up this network:

- *Server Nodes* on page 31
- *Access Switch or Top of Rack (ToR)* on page 32
- *Aggregation Switches* on page 32
- *Core* on page 33
- *Layer-2 and Layer-3 Switching* on page 33
- *VLANs* on page 33
- *Out of Band Management Network* on page 34
- *Dell EMC OpenSwitch Solution* on page 34

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 channel bonding across the servers and switches.

There are several types (or modes) of channel 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 Dell Networking S4048-ONs configured with a VLTi across them. The configuration settings are explained in greater detail in the Dell EMC Ready Bundle for Red Hat OpenStack Platform Deployment Guide.

Table 11: Channel Bonding Modes Supported

	Channel 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)

A single port is an option when bonding is not required. However, it is neither used nor validated in the Dell EMC Ready Bundle 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 viable options.

Access Switch or Top of Rack (ToR)

The servers connect to ToR switches, of which there are typically two per rack. Dell EMC's recommended switches are:

- **1GbE Connectivity** - Dell Networking S3048-ON
- **10GbE Connectivity** - Dell Networking S4048-ON

The 10GbE configuration utilizes Dell Networking S4048-ON or S6000-ON switches as the ToR switches. Dell EMC recommends that this pair of switches run Virtual Link Trunking (VLT) for HA, which enables the servers to terminate their LAG interfaces into two different switches instead of one. This configuration enables active-active bandwidth utilization, and provides redundancy within the rack if one switch fails or requires maintenance.

The uplink to the aggregation pair is 80Gb, using a LAG from each ToR switch. This is achieved by using two 40Gb interfaces in a LAG connecting to the aggregation pair. Therefore, a collective bandwidth of 160Gb is available from each rack. Each rack is managed as a separate entity, from a switching perspective, and ToR switches connect only to the aggregation switches.

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

Aggregation Switches

For a deployment of from one to three racks of 10G servers, Dell EMC recommends the Dell Networking S4048-ON or S6000-ON as the aggregation switch. They are both 10GbE and 40GbE-capable.

The 40GbE interfaces on the switches can be converted into four 10GbE interfaces, thereby converting the switches into multiple 10GbE-capable ports. ToR switches connect to aggregate switches via uplinks of 10GbE interfaces from the ToR to the Dell Networking S3048-ON.

Dell EMC's recommended architecture uses Virtual Link Trunking (VLT) between the two ToR switches in each rack, and then aggregation to a core switch. This enables a multi-chassis LAG from the ToR switches in each rack. The stacks in each rack can divide their links between this pair of switches to achieve powerful active-active forwarding, while using full bandwidth capability, with no requirement for spanning tree.

For the 10G server deployment, Dell EMC's recommendation depends upon:

- The scale at which the rack layouts are planned
- Required future scaling

When designing a large deployment, Dell EMC recommends the Dell Networking S6000-ON for aggregation for smaller scale and the Dell Networking Z9100 for larger deployments. The Dell Networking Z9100 is a 32-port, 40G high-capacity switch that can aggregate up to 15 racks of high-density Dell EMC PowerEdge R630 and Dell EMC PowerEdge R730xd servers. The rack-to-rack bandwidth needed in OpenStack is most suitably handled by a 40G-capable, non-blocking switch. The Dell Networking Z9100 can provide a cumulative bandwidth of 1.5TB of throughput at line-rate traffic from every port.

Core

The aggregation layer could itself be the network core in many cases, but otherwise it would connect to a larger core. Discussion of this topic is beyond the scope of this document.

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 aggregation layer, which is why VLT is used on the aggregation switches. The Red Hat OpenStack Platform Director requires a layer-2 domain in order to provision servers.

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

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

There are many tools for OOB management 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 network 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 Platform Director Node, 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, and the Horizon GUI. Depending upon the network configuration these networks may be either shared or routed, as needed. The Red Hat OpenStack Platform Director Node 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 Storage PS Series or SC Series 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**.

Out of Band Management Network

The management network for all the servers and switches is aggregated into a Dell Networking S3048-ON switch, located in each rack of up to 3 racks, or a pod. It uplinks on a 10G link to the S4048-ON switches.

The Out of Band (OOB) Management network is used for several functions:

- The highly available software uses it to reboot and partition servers.
- When an uplink to a router is added and the iDRACs configured to use it as a gateway, there are tools for monitoring the servers and gather metrics on them. 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.

- 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 170 Gigabit 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 6

Operational Notes

Topics:

- *Backup/Recovery*
- *High Availability*
- *Service Layout*
- *Deployment Overview*

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

Backup/Recovery

Backup and recovery have not been addressed in this configuration. Since the Red Hat OpenStack Platform Director Virtual Server is not needed for normal operations of the services, it is neither redundant nor backed up.

High Avaibility

In order for the solution to be ready for production, different systems need to be failure-tolerant. The Architecture Guide design utililizes 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 hosting the control plane services.
- Control plane services made highly available utilizing *ha-proxy*, *cora sync 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 Avaibility

  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 Storage PS Series or SC Series arrays are available.

Note: The Solution Admin Host, and the servers hosted on it (Red Hat Storage Console VM and Red Hat OpenStack Platform Director Node), 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 Bundle for Red Hat OpenStack Platform needs to reside upon a particular hardware type. For each server platform, two types of nodes have been designed:

- R630 for Computes, Controllers, Solution Admin Hosts, or Infrastructure hardware type
- R730xd for Storage nodes or Storage hardware type

Red Hat OpenStack Platform Director is designed for flexibility, enabling you to try different configurations in order to find the optimal service placement for your workload. *Table 12: Overcloud: Node Type to Services* on page 36 presents the recommended layout of each service.

The Red Hat OpenStack Platform Director and the Red Hat Ceph Storage Admin are deployed to the Solution Admin Host as individual VMs. This enables each tool to control its respective resources.

Table 12: Overcloud: Node Type to Services

Hardware Type	Service	Node to Deploy
Infrastructure	Ceilometer	OpenStack Controllers

Hardware Type	Service	Node to Deploy
Infrastructure	Cinder-scheduler	OpenStack Controllers
Infrastructure	Cinder-volume	OpenStack Controllers
Infrastructure	Database-server	OpenStack Controllers
Infastructure	Glance-Image	OpenStack Controllers
Infrastructure	HA-Proxy (Load Balancer)	OpenStack Controllers
Infrastructure	Heat	OpenStack Controllers
Infrastructure	Keystone-server	OpenStack Controllers
Infrastructure	Neutron-server	OpenStack Controllers
Infrastructure	Nova-Controller	OpenStack Controllers
Infrastructure	Nova dashboard-server	OpenStack Controllers
Infrastructure	Nova-multi-compute	Three or more Compute Nodes
Infrastructure	Pacemaker	OpenStack Controllers
Infrastructure	RabbitMQ-server (Messaging)	OpenStack Controllers
Infrastructure	Red Hat Ceph Storage RADOS Gateway	OpenStack Controllers
Infrastructure	Red Hat OpenStack Platform Director	Solution Admin Host (KVM)
Infrastructure	Red Hat Storage Console Admin	Solution Admin Host (KVM)
Infrastructure	Red Hat Ceph Storage Monitor	OpenStack Controllers
Storage	Red Hat Ceph Storage (Block)[3]	Three or more Storage Servers
Optional Services		
Storage	Dell Storage PS Series Array	Dell Storage PS Series Arrays
Storage	Dell Storage SC Series Array	Dell Storage SC Series Arrays
Storage	Dell Storage Enterprise Manager	Dell 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:

[3] Available through a custom Services engagement.

    - Deploy Red Hat Storage Console VM VM to the Solution Admin Host
    - Deploy Red Hat OpenStack Platform Director Virtual Server VM to the Solution Admin Host
  - Discover nodes
  - Import discovered nodes into Red Hat OpenStack Platform Director
  - Configure Overcloud files
  - Provision Overcloud
  - Validate all nodes' networking
  - Post-deployment adjustments, 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

Chapter 7

Solution Bundle

Topics:

- *Solution Bundle Overview*
- *Solution Bundle Rack Layout*
- *Solution Bundle Network Configuration*
- *Solution Admin Host (SAH) Networking*
- *Optional Solution Configurations*
- *Solution Bundle Expansion*
- *Larger Configurations*

This core architecture provides prescriptive guidance and recommendations, jointly engineered by Dell EMC and Red Hat, for deploying Dell EMC Ready Bundle for Red Hat OpenStack Platform version 10 with Dell EMC infrastructure.

Solution Bundle Overview

Our 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 Bundle consists of the components described in *Figure 3: Solution Bundle and Red Hat Ceph Storage Cluster* on page 41, and represents the base upon which all optional components and expansion of the Dell EMC Ready Bundle for Red Hat OpenStack Platform are built.

Using the recommended R630 and R730xd servers, and given a virtual machine with two (2) cores; 4GB of memory; and a 40GB local ephemeral drive, you can expect to run around 90 virtual machines with a 1.5 oversubscription of CPU cores. At 120 virtual machines, you will have:

- 2-to-1 CPU core oversubscription
- Local ephemeral storage undersubscribed
- Memory just starting to be oversubscribed

Since the Solution Bundle 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 10Gbe bonds for data networks, and the network switches are configured for HA. The Out of Band Management network is not HA, and is 1GbE.

Note: The Solution Admin Host that hosts the Red Hat OpenStack Platform Director Node and the Red Hat Storage Console VM is not made highly available, and must be appropriately managed for backup and recovery.

Please review and discuss the specifics with your Dell EMC sales representative.

Solution Bundle Rack Layout

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

See *Table 6: Controller Node Hardware Configurations – Dell EMC PowerEdge R630* on page 22 and *Table 9: Storage Node Hardware Configurations – Dell EMC PowerEdge R730xd* on page 23 for hardware configurations. The Solution Bundle includes:

- Node 1: Dell EMC PowerEdge R630 Solution Admin Host with the Red Hat OpenStack Platform Director Installed
- Nodes 2 - 4: Dell EMC PowerEdge R630 OpenStack Controllers
- Nodes 5 - 7 Dell EMC PowerEdge R630 Nova Compute Nodes
- Nodes 8 - 10: Dell EMC PowerEdge R730xd Storage Nodes
- Network Switches: Two (2) Dell Networking S4048-ON or S6000-ON, and one (1) Dell Networking S3048-ON

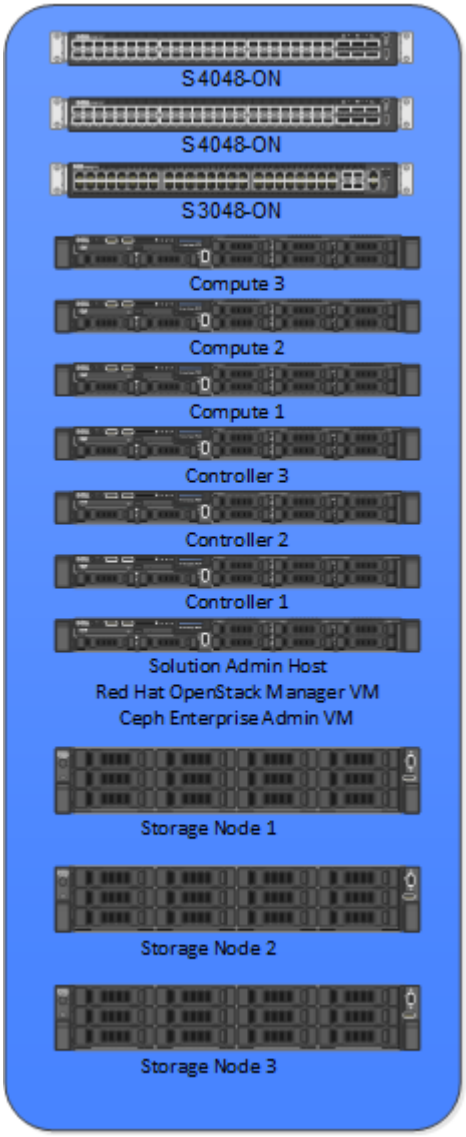


Figure 3: Solution Bundle and Red Hat Ceph Storage Cluster

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:

- **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.

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 special distribution of Ceph is used in this solution: Red Hat Ceph Storage 2.0, which also includes the Red Hat Storage Console VM. The Red Hat Storage Console VM also includes Red Hat Ceph Storage troubleshooting and servicing tools and utilities. Red Hat Ceph Storage 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 Bundle Network Configuration

The network for the Dell EMC Ready Bundle for Red Hat OpenStack Platform has been designed to support production-ready servers with a highly available network configuration. See *Figure 4: Cluster Network Logical Architecture with Optional Dell Storage PS Series* on page 43.

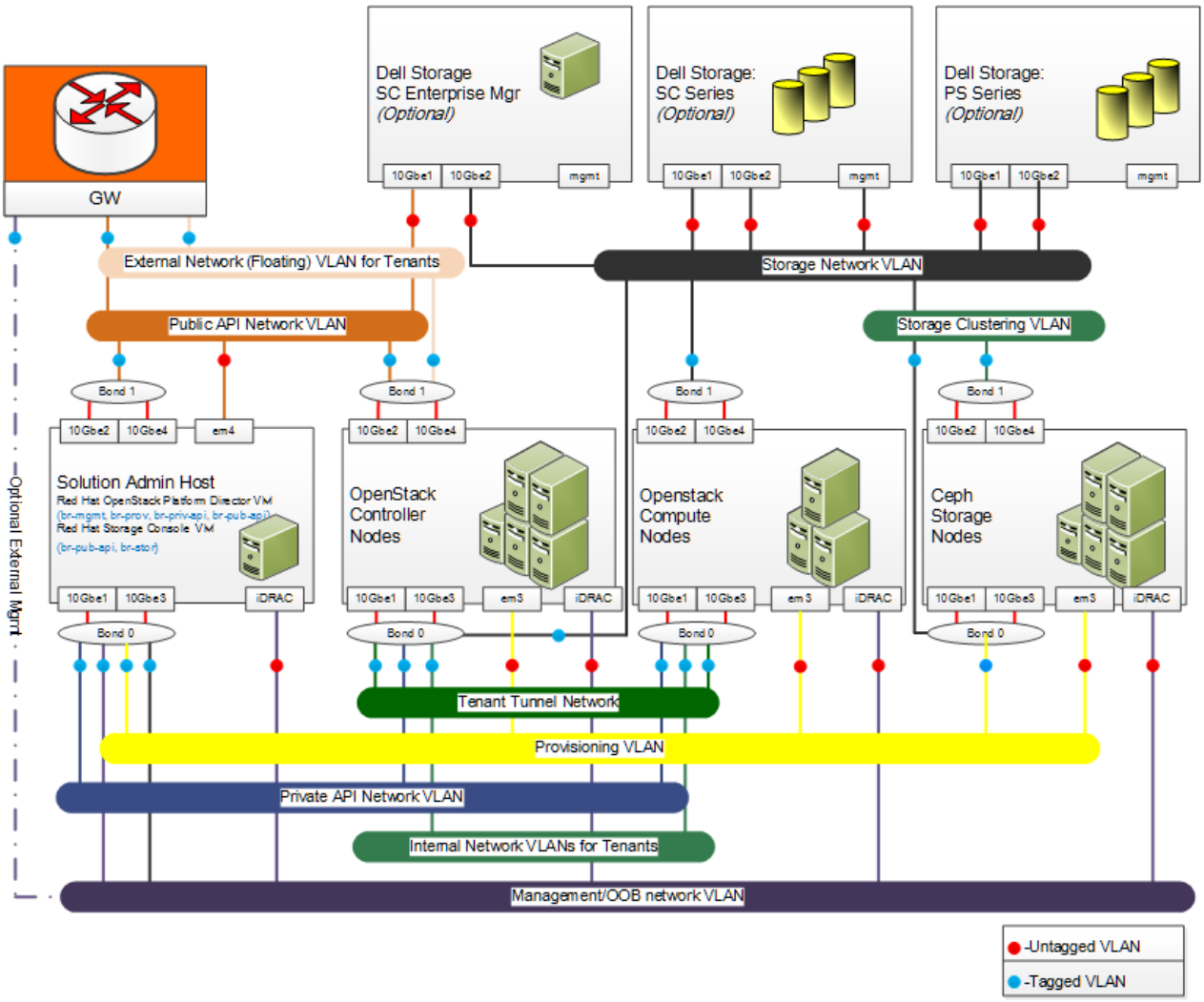


Figure 4: Cluster Network Logical Architecture with Optional Dell Storage PS Series

The node type will determine how the switches are configured for delivering the different networks. *Table 13: OpenStack Node Type to Network 802.1q Tagging* on page 43 and *Table 14: Storage Node Type to Network 802.1q Tagging* on page 44 outline the networks to the node types. The Management/OOB network is used by the Director Node and the Cluster Software to manage the OpenStack Controllers. All iDRACs are plugged into this network without tagging.

Table 13: OpenStack Node Type to Network 802.1q Tagging

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

Network	Solution Admin Host	OpenStack Controller	OpenStack Compute	Red Hat Ceph Storage
Management/OOB Network VLAN	Connected, Tagged	Not Connected	Not Connected	Not Connected
Private API Network VLAN	Connected, Tagged	Connected, Tagged	Connected, Tagged	Not Connected
Provisioning VLAN	Connected, Tagged	Connected, Untagged	Connected, Untagged	Connected, Untagged
Public API Network VLAN	Connected, Tagged[4]	Connected, Tagged	Not Connected	Not Connected
Storage Clustering VLAN	Not Connected	Not Connected	Not Connected	Connected, Tagged
Storage Network VLAN	Connected, Tagged	Connected, Tagged	Connected, Tagged	Connected, Tagged
Tenant Tunnel Network	Not Connected	Connected, Tagged	Connected, Tagged	Not Connected

Table 14: Storage Node Type to Network 802.1q Tagging

Network	PS Series Array	SC Series Enterprise Manager	SC Series Array
External Network for Tenants VLAN (Floating IP Network)	Not Connected	Not Connected	Not Connected
iDRAC physical connection to the Management/OOB VLAN	Not Connected	Not Connected	Not Connected
Internal Networks VLAN for Tenants	Not Connected	Not Connected	Not Connected
Management/OOB Network VLAN	Not Connected	Not Connected	Not Connected
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

[4] The 1Gbe port is used for provisioning the SAH during its installation only and can be disconnected afterwards.

Solution Admin Host (SAH) Networking

The Solution Admin Host has internal bridged networks for the Virtual Machines. It is physically connected to the following networks:

- **Management Network** — used by the Red Hat OpenStack Platform Director Node for iDRAC control of all Overcloud nodes.
- **Private API Network** — Used by the Red Hat OpenStack Platform Director Node to run Tempest tests against the OpenStack private AP
- **Provisioning Network** — Used by the Red Hat OpenStack Platform Director Node 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 Platform Director Node
    - HTTP/HTTPS access to the Red Hat Storage Console VM
    - Optional - SSH Access to the Red Hat OpenStack Platform Director Node and Red Hat Storage Console VM
  - Outbound Access
    - HTTP/HTTPS access for Red Hat Ceph Storage, RHEL, and RHOSP updates
    - Used by the Red Hat OpenStack Platform Director Node to run Tempest tests using the OpenStack public API
- **Storage Network** — Used by the Red Hat Storage Console VM to monitor and manage the Red Hat Ceph Storage Cluster

Figure 5: Solution Admin Host Internal Network Fabric on page 46 displays how the networks are bridged inside the Solution Admin Host. Since the Provisioning, Private API, and Storage networks come in on one physical interface and the Public API and External Tenant networks on the other, 802.1q tagging is configured on the SAH and corresponding switch ports. This Architecture Guide does not cover any security aspects, so the appropriate Network and Security teams should be involved before connecting any machine to the externally-accessible networks.

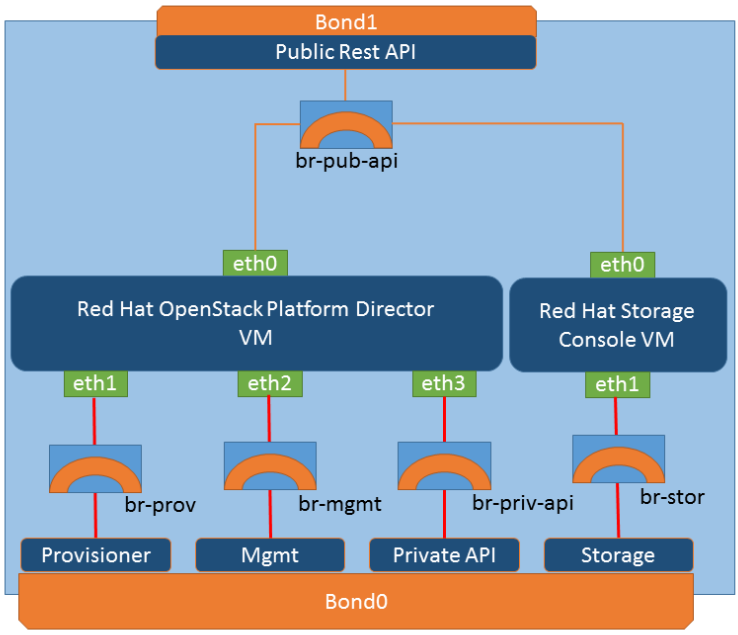


Figure 5: Solution Admin Host Internal Network Fabric

Optional Solution Configurations

Optional component changes require that the base solution be modified. For example, you can use either the Dell EMC PowerEdge R430 or Dell EMC PowerEdge R730 as OpenStack Compute nodes instead of the Dell EMC PowerEdge R630 (see *Optional Compute Servers* on page 24 for more information).

If you replace the Dell EMC PowerEdge R630 OpenStack Computes with either Dell EMC PowerEdge R430 or Dell EMC PowerEdge R730 servers there are no networking changes; just a change of the server hardware. When ordering the servers, the configuration must be similar to the Dell EMC PowerEdge R630. The servers will require:

- Enough disks to build a single RAID 10 set
- At least four (4) Intel® 10G network interface ports
- Two (2) 1Gb network interface ports
- An iDRAC Enterprise

The memory, RAID, and CPU configuration should be sized based upon expected workloads.

Optional Dell Storage with the Solution Bundle

The Solution Bundle with Dell Storage PS Series or SC Series Storage has the same characteristics; the only change is that the Storage backend software is configured to use Red Hat Ceph Storage, and Dell Storage PS Series and/or SC Series. The Storage node servers are supplemented with one or more Dell Storage PS Series arrays.

The Solution bundle shown in *Figure 6: Solution with Optional Dell Storage* on page 47 includes Dell Storage PS Series and SC Series Storage Arrays. This can be either or both, depending upon your Application and Storage needs. Your Dell EMC sales representative will work to find the proper configuration for your needs prior to ordering.

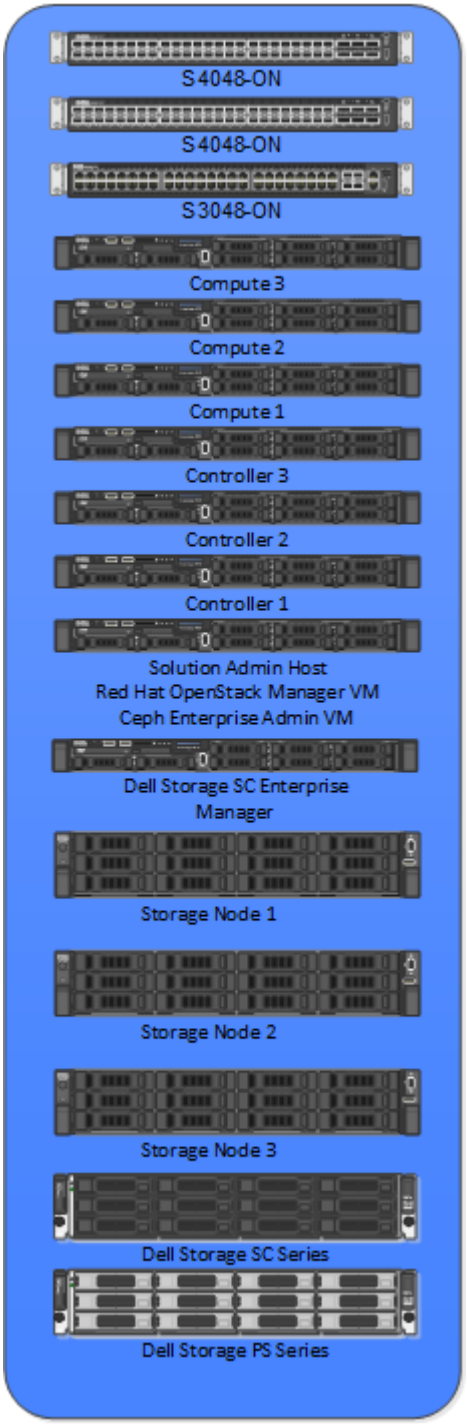


Figure 6: Solution with Optional Dell Storage

The Controller nodes will use the Storage Network VLAN to access PS Series Storage Pools, created on the Storage Group, for creation, deletion, and snapshots. Each Compute node must have access to the Storage nodes, through the Storage Network VLAN, in order for its iSCSI driver to interact with the volumes associated to Virtual Machines that it hosts.

Dell Storage PS Series Arrays are connected to the Storage Networking VLAN untagged only; all other nodes use the same layout as in *Table 13: OpenStack Node Type to Network 802.1q Tagging* on page 43 and *Figure 4: Cluster Network Logical Architecture with Optional Dell Storage PS Series* on page 43.

The Solution Bundle, when adding Dell Storage Center, has the same characteristics; the only change is the Storage backend software is configured to use both Red Hat Ceph Storage and Dell Storage Center with Dell EMC Enterprise Manager Platform. The Storage node servers are supplemented with one or more Dell Storage Centers managed by the Dell EMC Enterprise Manager. The Solution bundle shown has Dell Storage Center; this can be one or more, depending upon your Application and Storage needs. Prior to ordering, your Dell EMC sales representative will work to find the proper configuration for your needs.

The Dell EMC Enterprise Manager platform is used to proxy OpenStack API calls to the configured Dell Storage SC Series arrays. The Controller nodes will use the Storage Network VLAN to access the Dell EMC Enterprise Manager Node for management of volumes and snapshots.

The Compute nodes must have access to the Dell Storage SC Series through the Dell Storage Center iSCSI ports in order for its iSCSI driver to interact with the volumes associated to Virtual Machines that it hosts.

Dell EMC SC Series Arrays are connected to the Storage Networking VLAN untagged only; all other nodes will use the layout as in *Figure 4: Cluster Network Logical Architecture with Optional Dell Storage PS Series* on page 43.

Solution Bundle Expansion

The Dell EMC Ready Bundle for Red Hat OpenStack Platform bundle can be expanded by adding Compute nodes, Storage nodes, or Dell Storage (PS Series or SC Series) Arrays. You can expand up to 20 servers per rack and/or 30 rack units (Infrastructure, Compute, and Storage combined).

Expanding beyond the first rack will require the addition of aggregation network switches (see *Aggregation Switches* on page 32), additional ToR and management switches in each rack, and the appropriate power and cooling. Expansion beyond a total of three (3) racks must be designed and configured based upon your requirements. Please work with your Dell EMC sales representative to properly architect these large cluster deployments.

Note: When expanding the cluster, the Controller nodes can be expanded to more systems, but the expansion must be done in odd numbers only. For other expansion details, please speak with your Dell EMC sales representative.

Rack 1

Base Solution with Red Hat Ceph Storage, optionally Dell Storage PS Series:

- 2 S4048-ON or S6000-ON ToR Switches
- 1 S3048-ON Management Switch
- Solution Admin Host
- 3 Controller Nodes
- 3 Nova Compute Nodes
- 3 Storage Nodes
- Optional Dell Storage PS Series Storage Arrays can be added

This configuration consists of a total of 15 U's, and a total of 12 servers (allowing up to 8 more servers in Rack 1). In rack 1 you could add up to either:

- 8 R630 Nova Compute Nodes, or
- 7 R730xd Storage Nodes or
- Dell Storage PS Series or SC Series Storage Arrays

Note: You can use a combination of the three options that does not exceed a total of 20 servers or 30 rack units.

Rack 2

- 2 Z9100 or S6000-ON or S4048-ON Aggregation Switches, depending upon your load requirements
- 2 S4048-ON or S6000-ON ToR Switches
- 1 S3048-ON Management Switch

Note: To split HA across the racks, you can move one or two Controllers from Rack 1 to Rack 2.

For additional nodes, you can add up to either:

- 19 R630 Nova Compute Nodes, or
- 14 R730xd Storage Nodes or Dell Storage PS Series Storage Arrays , or
- Dell Storage PS Series or SC Series Storage Arrays

Note: You can use a combination of the three options that does not exceed a total of 20 servers or 30 rack units.

Rack 3

- 2 S4048-ON or S6000-ON ToR Switches
- 1 S3048-ON Management Switch

Note: To split HA across the racks, you can move a Controller from Rack 1 or Rack 2 to Rack 3, giving one Controller per rack.

For additional nodes, you can add up to either:

- 19 R630 Nova Compute Nodes, or
- 14 R730xd Storage Nodes, or
- Dell Storage PS Series SC Series Storage Arrays

Note: You can use a combination of the three options that does not exceed a total of 20 servers or 30 rack units.

Moving the Controllers is not documented as part of the solution.

Larger Configurations

Clusters larger than the three (3) racks of a Solution bundle must be designed, sized, and configured based on your requirements. Please work with your Dell EMC sales representative to properly architect these large deployments

Appendix A

Update History

Topics:

- *Initial Release*
- *Version 1*
- *Version 2*
- *Version 3*
- *Version 4*
- *Version 5*
- *Version 6*

This appendix lists changes to this document, per major release.

Initial Release

First Architecture Guide for the Dell EMC Ready Bundle for Red Hat OpenStack Platform with Red Hat OpenStack Platform.

Version 1

Update to support:

- Red Hat OpenStack Provisioning 5 Icehouse
- R620
- Red Hat Ceph Storage
- Cinder Multi-Backend and Multi-Instance
- Dell EMC EqualLogic (PS Series)
- HA

Version 2

Updated as follows:

- Updated to OpenStack Version: Juno
- Updated to RHOSP 6
- New network diagrams
- Support for HA-only clusters
- Support for up to three (3) racks of equipment
- Renamed *Admin Node* to *Solution Admin Host*
- Added Virtual Servers to support Provisioning nodes
- Added optional Gateways for Provisioning/Storage/Management networks

Version 3

Updated as follows:

- New network diagram
- Added support for Dell EMC PowerEdge R630 and Dell EMC PowerEdge R730xd
- Removed support for Dell EMC PowerEdge R620, Dell EMC PowerEdge R720, and Dell EMC PowerEdge R720xd
- Added support for OpenStack Neutron
- Removed support for Nova-Network
- Removed the POC from the Solution
- Updated to OpenStack Version: Kilo
- Updated to RHOSP 7
- Standardized terminology for platforms and nodes

Update 1

- New Network Diagram

    - SAH changes
        - 1Gbe to external
        - Private API to bond
- Solution Admin Diagram
    - Added Private API for Tempest Test Node
    - Added Tempest Test Node
- Solution Admin Diagram Text
    - Added Private API for Tempest Test Node
    - Added Tempest Test Node to Public API Outbound
- Server Networking
    - Added discussion about the new bond modes and what is used by solution
    - Added table outline all modes and what can be used where.
        - Indicated the “solution default” on all modes

Version 4

Updated as follows:

- OpenStack Version: Kilo
- RHOSP 7
- Updated Taxonomy
- Added support for Dell EMC PowerEdge R430 and Dell EMC PowerEdge R730
- Added Ceph Object Store support with RADOS Gateway
- Standardized terminology for platforms and nodes
- Added Dell Storage Options

Version 5

Updated as follows:

- Updated to OpenStack Version: Liberty
- Red Hat OpenStack Platform version 8
- Added Dell EMC PowerEdge R730xd 24-drive variants as optional Compute and Storage nodes

Version 6

Updated as follows:

- Updated to OpenStack Version: Mitaka
- Red Hat OpenStack Platform version 9
- Added support for minor OpenStack update
- Added support for major OpenStack upgrade
- Added support for Red Hat CloudForms version 4.2
- Added support for OpenShift Container Platform version 3.3
- Added support for CloudForms and OpenShift integration

- Added support for billing and chargeback reporting

Appendix B

References

Topics:

- *To Learn More*

Additional information can be obtained at *http://www.dell.com/en-us/work/learn/openstack-cloud* or by e-mailing *openstack@dell.com*.

If you need additional services or implementation help, please contact your Dell EMC sales representative.

To Learn More

For more information on the Dell EMC Ready Bundle for Red Hat OpenStack Platform visit *http://www.dell.com/learn/us/en/04/solutions/red-hat-openstack*.

Copyright © 2014-2017 Dell Inc. or its subsidiaries. All rights reserved. Trademarks and trade names may be used in this document to refer to either the entities claiming the marks and names or their products. Specifications are correct at date of publication but are subject to availability or change without notice at any time. Dell EMC and its affiliates cannot be responsible for errors or omissions in typography or photography. Dell EMC's Terms and Conditions of Sales and Service apply and are available on request. Dell EMC service offerings do not affect consumer's statutory rights.

Dell EMC, the DELL EMC logo, the DELL EMC badge, and PowerEdge are trademarks of Dell Inc.