

Dell Red Hat OpenStack Cloud Solution Reference Architecture Guide - Version 4.0.1



Contents

| | |
|--|----|
| Trademarks..... | 4 |
| Notes, Cautions, and Warnings..... | 5 |
| Glossary..... | 6 |
| Overview..... | 8 |
| OpenStack Maturity..... | 8 |
| Hardware Options..... | 8 |
| Networking and Network Services..... | 8 |
| Taxonomy..... | 9 |
| Red Hat Enterprise Linux OpenStack Platform 7..... | 10 |
| Key Benefits..... | 11 |
| OpenStack Architecture..... | 12 |
| OpenStack Components..... | 12 |
| Server Options..... | 14 |
| PowerEdge R630 Server..... | 14 |
| PowerEdge R730xd Servers..... | 14 |
| Base Hardware Configurations..... | 15 |
| Configuration Notes..... | 16 |
| Optional Servers..... | 16 |
| PowerEdge R430..... | 17 |
| PowerEdge R730..... | 17 |
| Storage Options..... | 18 |
| Red Hat Ceph Storage..... | 18 |
| Optional Dell Storage..... | 19 |
| Dell Storage Arrays..... | 19 |
| Cinder Multi-Backend Support..... | 19 |
| Operational Notes..... | 20 |
| Backup/Recovery..... | 20 |
| Service Layout..... | 20 |
| Ephemeral Storage..... | 21 |
| Deployment..... | 21 |

| | |
|---|--------|
| Network Architecture..... | 23 |
| Infrastructure Layouts..... | 23 |
| Network Components..... | 23 |
| Server Nodes..... | 23 |
| Access Switch or Top of Rack (ToR)..... | 24 |
| Aggregation Switches..... | 25 |
| Core..... | 25 |
| Layer-2 and Layer-3 Switching..... | 25 |
| vLANs..... | 25 |
| Out of Band Management Network..... | 26 |
| Dell OpenSwitch Solution..... | 26 |
| Solution Bundle..... | 28 |
| Solution Bundle Rack Layout..... | 28 |
| Solution Bundle Network Configuration..... | 30 |
| Solution Admin Host (SAH) Networking..... | 32 |
| Optional Solution Configurations..... | 33 |
| Optional Dell Storage with the Solution Bundle..... | 33 |
| Solution Bundle Expansion..... | 37 |
| Rack 1..... | 37 |
| Rack 2..... | 37 |
| Rack 3..... | 37 |
| Larger Configurations..... | 38 |
| Update History..... | 39 |
| Initial Release..... | 39 |
| Version 1..... | 39 |
| Version 2..... | 39 |
| Version 3..... | 39 |
| Version 4..... | 40 |
| References..... | 41 |
| To Learn More..... | 41 |

Trademarks

© 2014-2015 Dell Inc. All rights reserved. Reproduction of this material in any manner whatsoever without the express written permission of Dell Inc. is prohibited. For more information, contact Dell.

Trademarks used in this text: Dell™, the DELL logo, Dell Precision™, OptiPlex™, Latitude™, PowerEdge™, PowerVault™, OpenManage™, EqualLogic™, Dell Compellent™, KACE™, FlexAddress, Dell Networking™, and Vostro™ are trademarks of Dell Inc. Intel®, Pentium®, Xeon®, Core®, and Celeron® are registered trademarks of Intel Corporation in the U.S. and other countries. AMD® is a registered trademark and AMD Opteron™, AMD Phenom™, and AMD Sempron™ are trademarks of Advanced Micro Devices, Inc. Microsoft®, Windows®, Windows Server®, MS-DOS®, and Windows Vista® are either trademarks or registered trademarks of Microsoft Corporation in the United States and/or other countries.




Red Hat®, Red Hat® Enterprise Linux®, the Shadowman logo, and JBoss are 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. Novell® is a registered trademark and SUSE™ is a trademark of Novell Inc. in the United States and other countries. Oracle® is a registered trademark of Oracle Corporation and/or its affiliates. Citrix®, Xen®, XenServer® and XenMotion® are either registered trademarks or trademarks of Citrix Systems, Inc. in the United States and/or other countries. VMware®, Virtual SMP®, vMotion®, vCenter®, and vSphere® are registered trademarks or trademarks of VMware, Inc. in the United States or other countries.

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.

Other trademarks and trade names may be used in this publication to refer to either the entities claiming the marks and names or their products. Dell Inc. disclaims any proprietary interest in trademarks and trade names other than its own.

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.

Glossary

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

Cloud Computing

See <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.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 <http://en.wikipedia.org/wiki/DevOps>

Hypervisor

Software that runs virtual machines (VMs).

LAG

Link Aggregation Group.

LOM

LAN on motherboard.

Node

One of the servers in the system.

Pod

An installation comprised of three racks, based upon server and network sizing.

SAH

The Solution Admin Host (SAH) is a physical server that supports the infrastructure roles needed for the cluster.

Storage Node

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

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 on 10G or 40G interfaces; 1G interfaces are not supported.

Overview

This Reference Architecture focuses on helping organizations to deploy an OpenStack® based cloud.

This jointly-engineered and validated architecture details the Dell™ Red Hat® Cloud Solutions with Red Hat Enterprise Linux™ OpenStack Platform and encompasses software, hardware, and integration of the solution components. The architecture provides prescriptive guidance and recommendations for Compute, Infrastructure, Storage nodes, networking design, and various system configurations.

OpenStack Maturity

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

At publication the current release of OpenStack is codenamed *Kilo*. It builds upon previous releases with 342 new features to support software development, and managing data and application infrastructure, at scale. It is developed by over 1400 individuals employed by more than 133 organizations. Please see <http://www.OpenStack.org/software/kilo>.

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

Hardware Options

To reduce time spent on hardware specification for an initial system this Reference Architecture offers specific choices for servers, storage, and networking.

The recommended hardware is general-purpose, and enables a wide range of configuration options, including optimized configurations for:

- Compute roles
- Infrastructure roles
- Storage roles

As noted throughout this Reference Architecture, Dell constantly adds capabilities to expand this offering.

Each of the Dell PowerEdge™ server configurations in this Reference Architecture is designed as a getting-started setup for OpenStack Compute, Infrastructure, and Storage. Dell recommends starting with OpenStack software using components from this configuration because the hardware and operations processes are a flexible foundation to expand upon. By design, you can expand the Reference Architecture configuration as your cloud deployment grows, so your investment is protected.

Networking and Network Services

Network configuration is based upon using the Neutron-based options supported by the OSP code base, and not relying 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 teaming
- Redundant trunking top-of-rack (ToR) switches into core routers

This enables the solution to operate in a full production environment.

See [Network Architecture](#) on page 23 for guidelines. Detailed designs are available through Dell consulting services.

Taxonomy

This solution contains the following core OpenStack components as delivered in the Red Hat Enterprise Linux OpenStack Platform. See [Table 1: Base Components](#) on page 9.

Table 1: Base Components

| Component | Code Name |
|--------------------|--|
| Block Storage | Cinder with Ceph and Dell Storage PS Series or SC Series |
| Compute | Nova |
| Dashboard | Horizon |
| Identity | Keystone |
| Image Service | Glance |
| Validation Testing | Tempest |
| Networking | Neutron |
| Orchestration | Heat |
| Telemetry | Ceilometer |

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

There are several optional¹ OpenStack components that are available but not part of the base solution. See [Table 2: Optional Components](#) on page 9.

Table 2: Optional Components

| Component | Code Name |
|-------------------------|-----------|
| Bare Metal Provisioning | Ironic |
| Database | Trove |
| Data Processing | Sahara |
| DNS as a Service | Designate |
| File Share Service | Manila |
| Key Management | Barbican |

¹ Available through a custom Services engagement, for customer evaluation only.

The taxonomy presented in [Figure 1: Dell Red Hat OpenStack Cloud Solution Base Taxonomy](#) on page 10 reflects infrastructure components and OpenStack-specific components, that are under active development by the community, Dell, and Red Hat. The taxonomy reflects that there are two sides for cloud users:

- 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 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 Reference Architecture. Conforming to best practices helps reduce operational risk by leveraging the accumulated experience of Dell, Red Hat and the broader OpenStack community.

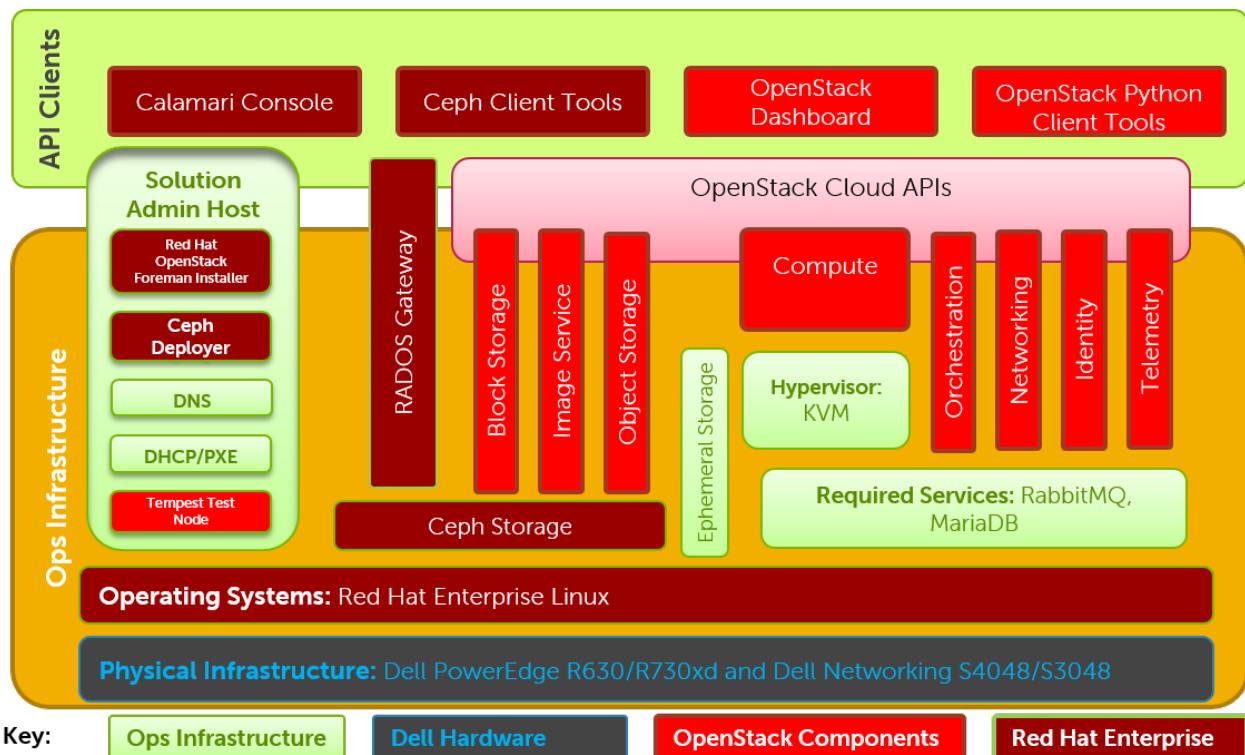


Figure 1: Dell Red Hat OpenStack Cloud Solution Base Taxonomy

Red Hat Enterprise Linux OpenStack Platform 7

Red Hat Enterprise Linux OpenStack Platform is a production-ready OpenStack solution that provides an integrated foundation to create, deploy, and scale a secure and reliable public or private OpenStack cloud. It delivers a managed cloud platform built on Red Hat Enterprise Linux, co-engineered and integrated with Red Hat's OpenStack technologies, giving you the agility to scale and quickly meet customer demands without compromising on availability, security, or performance.

Red Hat Enterprise Linux OpenStack Platform is purposely designed with the recognition of the unique dependencies OpenStack has on the underlying Linux it's installed on. Red Hat uniquely co-engineers and integrates Red Hat OpenStack technology with Red Hat Enterprise Linux Server 7, ensuring a stable, production-ready cloud platform. Version 7 boasts all of the core features and functions of the

community Kilo release and adds some additional innovations by Red Hat, resulting in a hardened, stable cloud platform.

Key Benefits

- Co-engineered and Integrated: OpenStack depends on Linux for performance, security, hardware enablement, networking, storage, and other primary services. Red Hat Enterprise Linux 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 the underlying operating platform.
- Deploy with confidence, as Red Hat Enterprise Linux OpenStack Platform provides a hardened and stable branch release of OpenStack and Linux, which is supported by Red Hat for a three (3) year “production phase” life cycle, well beyond the six-month release cycle of unsupported community OpenStack. Security fixes, bug fixes, performance enhancements, and some features can be back-ported from future releases without disrupting production environments.
- 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 certified ISV certifications, so that you can 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 Enterprise Linux OpenStack Platform deployments, ensuring the same level of broad support and compatibility customers enjoy today in the Red Hat Enterprise Linux ecosystem.
- Bring security to the cloud. Rely on 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.²

² Available through a custom Services engagement, for customer evaluation only.

OpenStack Architecture

While OpenStack has many configurations and capabilities, we focus on the primary components for Red Hat Enterprise Linux® OpenStack Platform 7 (Kilo), as Dell has defined in [Taxonomy](#) on page 9 above.



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 3: OpenStack Components

| Function | Code Name | Description |
|------------------------|-----------|--|
| 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 services 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. |
| 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. |
| 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. |
| Integration 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, Scenarios, and other specific tests useful in validating an OpenStack deployment. |
| Networking | Neutron | OpenStack Networking 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. |

| Function | Code Name | Description |
|---------------|------------|---|
| 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. |
| 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. |
| 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. |

Server Options

The base Solution supports the PowerEdge R630 and R730xd Server lines. See [Optional Servers](#) on page 16 for other options. The following sections describe the supported server models and configurations required. Detailed part lists and rack layouts are included in the *Dell Red Hat OpenStack Cloud Solution Build of Materials Guide*.

Server topics discussed include:

- [PowerEdge R630 Server](#) on page 14
- [PowerEdge R730xd Servers](#) on page 14
- [Base Hardware Configurations](#) on page 15
- [Configuration Notes](#) on page 16
- [Optional Servers](#) on page 16

PowerEdge R630 Server

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

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

The PowerEdge R630 server is versatile and highly configurable for a variety of solutions, supporting the latest Intel® Xeon® processor E5-2600 v3 product family, 24 DIMMs of high-performance DDR4 memory and a broad range of local storage options.

PowerEdge R730xd Servers

The 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 PowerEdge R730xd meets the needs of many different storage workloads with the latest Intel® Xeon® processor E5-2600 v3 product family, 24 DIMMs of high-performance DDR4 memory, and a broad range of local storage options.





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

Base Hardware Configurations

Table 4: Controller Node Hardware Configurations – PowerEdge R630

| Machine Function | Solution Bundle Controller Nodes |
|--------------------|----------------------------------|
| Platform | PowerEdge R630 |
| CPU | 2 x E5-2650v3 (10-core) |
| RAM (Minimum) | 128 GB |
| LOM | 2 x 1Gb, 2 x Intel X520 10Gb |
| Add-in Network | 1 x Intel X520 DP 10Gb DA/SFP+ |
| Disk | 4 x 500GB 7.2k NLSAS |
| Storage Controller | PERC H730 |
| RAID | RAID 10 |

Table 5: Compute Node Hardware Configurations – PowerEdge R630

| Machine Function | Solution Bundle Compute Nodes |
|--------------------|--------------------------------|
| Platform | PowerEdge R630 |
| CPU | 2 x E5-2650v3 (10-core) |
| RAM (Minimum) | 128 GB |
| LOM | 2 x 1Gb, 2 x Intel X520 10Gb |
| Add-in Network | 1 x Intel X520 DP 10Gb DA/SFP+ |
| Disk | 6 x 600GB 10k SAS |
| Storage Controller | PERC H730 |
| RAID | RAID 10 |

Table 6: Infrastructure Node Hardware Configurations – PowerEdge R630

| Machine Function | Solution Bundle Infrastructure Nodes |
|------------------|--------------------------------------|
| Platform | PowerEdge R630 |
| CPU | 2 x E5-2630v3 (10-core) |
| RAM (Minimum) | 32 GB |
| LOM | 2 x 1Gb, 2 x Intel X520 10Gb |

| Machine Function | Solution Bundle Infrastructure Nodes |
|--------------------|--------------------------------------|
| Add-in Network | 1 x Intel X520 DP 10Gb DA/SFP+ |
| Disk | 4 x 500GB 7.2 NLSAS |
| Storage Controller | PERC H730 |
| RAID | RAID 10 |

Table 7: Storage Node Hardware Configurations – PowerEdge R730xd

| Machine Function | Solution Bundle Storage Nodes |
|--------------------|--|
| Platforms | PowerEdge R730xd |
| CPU | 2 x E5-2650v3 (10-core) |
| RAM (Minimum) | 48 GB |
| LOM | 1 x 1Gb, 2 x Intel X520 10Gb |
| Add-in Network | 2 x Intel X520 DP 10Gb DA/SFP+ |
| Disk | Flex Bay: 2 X 300GB 10K 2.5-inch (OS) Front Drives: 3 X 200GB SSD 13 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 account representative before changing the recommended hardware configurations.

Configuration Notes

The *Dell Red Hat OpenStack Cloud Solution Bill of Materials Guide* contains the full bill of materials (BOM) listing for the PowerEdge R630 and R730Xd server configurations.

The R630 and 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 Red Hat OpenStack Cloud Solution Bill of Materials Guide* which outlines the supported cards and includes them as part of the solution.

Optional Servers

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

- [PowerEdge R430](#) on page 17
- [PowerEdge R730](#) on page 17

PowerEdge R430

The 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 means that the server can be utilized in your infrastructure with minimal impact.

Please contact your Dell sales representative for options and configurations.

PowerEdge R730

The PowerEdge R730 is an option for the OpenStack Computes. You can either add them to an existing cluster or use them as the base Computes.

With a 2U form factor, the 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 sales representative for options and configurations.

Storage Options

OpenStack has several storage services, including:

- Cinder
- Glance
- Swift³

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 our customer's needs.

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

As with other OpenStack services, there are client and server components for each of these. 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. This solution modifies the default configuration for these services.

This solution include three alternate implementations of Cinder that enable the cluster to fit many needs:

- [Red Hat Ceph Storage](#) on page 18
- [Optional Dell Storage](#) on page 19
- [Cinder Multi-Backend Support](#) on page 19

This solution provides Ceph as an alternate implementation for Glance and Block and Object stores.

Red Hat Ceph Storage

Red Hat Ceph Storage is a scale-out, shared-nothing, distributed, software-defined storage system. The Solution bundle includes Red Hat Ceph Storage. Red Hat Ceph Storage is used as the storage backend for Nova, Cinder and Glance. Red Hat Ceph Storage is used as block storage. Storage nodes run the Red Hat Ceph Storage software. Compute and Controller nodes run the Red Hat Ceph Storage block client.

The Red Hat Ceph Storage option provides object storage for OpenStack VM and for clients external to OpenStack. The object storage interface is an implementation of:

- The S3 RESTful API
- The basic data access model of the Swift RESTful API

³ Available through a custom Services engagement, for customer evaluation only.

The Object Gateway processes run on the Controller nodes. Object IO load is distributed across Object gateway processes.

Optional Dell Storage

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

- [Dell Storage Arrays](#) on page 19

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
- Dell Storage SC Series provides automatic data tiering based on usage
- iSCSI connections are only supported with either the Dell Storage PS Series or SC Series. Fibre Channel connections are not supported.

The Dell Storage SC Series will require 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.

Cinder Multi-Backend Support

This solution can take advantage of Cinder's Multi-Backend and Multi-Instance support, using different storage types to meet the requirements that the application requires. You can meet different use cases based upon:

- Performance
- Large data transfers
- Temporary storage
- Or other criteria

For example, you can create a Ceph cluster with a performance group and a second storage option of a Dell Storage PS Series group to support large data transfers. Then, using Cinder multi-backend, configure the virtual machine volume requirements as required.

Operational Notes

This section provides a basic overview of the following system aspects:

- [Backup/Recovery](#) on page 20
- [Service Layout](#) on page 20
- [Ephemeral Storage](#) on page 21
- [Deployment](#) on page 21

Backup/Recovery

Backup and recovery have not been addressed in this configuration. The Red Hat Enterprise Linux OpenStack Platform Installer Virtual Server, while not needed for normal operations of the services, is not redundant or backed up.

Service Layout

During the deployment each service configured by the solution needs to be on a particular hardware type.

For each server platform, two types of nodes have been designed:

- Infrastructure
- Storage

Red Hat Enterprise Linux OpenStack Platform Installer is designed for flexibility, enabling you to try different configurations in order to find the optimal service placement for your workload. [Table 8: Node Type to Services](#) on page 20 presents the recommended layout of each service.

The Red Hat Enterprise Linux OpenStack Platform Installer and the Red Hat Ceph Storage Admin are deployed to the Solution Admin Host as manually-configured VMs. This enables each tool to control its respective resources.

Table 8: Node Type to Services

| Hardware Type | Service | Node to Deploy |
|----------------|--------------------------|-----------------------|
| Infrastructure | Ceilometer | OpenStack Controllers |
| Infrastructure | Cinder-scheduler | OpenStack Controllers |
| Infrastructure | Cinder-volume | OpenStack Controllers |
| Infrastructure | Database-server | 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 |

| Hardware Type | Service | Node to Deploy |
|--------------------------|---|--|
| 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 Enterprise Linux OpenStack Platform Installer | Solution Admin Host (KVM) |
| Infrastructure | Tempest Test Node | Solution Admin Host (KVM) |
| Infrastructure | Red Hat Ceph Storage Admin (Calamari) | Solution Admin Host (KVM) |
| Infrastructure | Red Hat Ceph Storage Monitor | OpenStack Controllers |
| Storage | Red Hat Ceph Storage (Block) ⁴ | 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 |

Ephemeral Storage

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

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 *rdn*. During the planning process, decisions must be made as to:

- Where the ephemeral drives are stored
- The storage location (storage must be sized accordingly)

Boot from volume/snapshot will use one of the back ends for Cinder.

Deployment

Deployment consists of three phases:

1. Hardware Setup:

- Rack and stack
- Cabling
- Server BIOS configuration
- Server RAID configuration
- Switch configuration

2. Software Setup:

⁴ Available through a custom Services engagement, for customer evaluation only.

- Deploy Solution Admin Host for provisioning services:
 - Deploy Red Hat Ceph Storage Admin Node to the Solution Admin Host
 - Deploy Red Hat Enterprise Linux OpenStack Platform Installer Virtual Server
 - Deploy Tempest Test Node to the Solution Admin Host (optional)
- Provision all nodes with operating systems
- Validate all nodes networking
- Provision services to nodes based on their node types
- Post-deployment adjustments, including but not limited to:
 - Enabling fencing
 - Enable local storage for ephemeral
 - Transition shared ephemeral storage to local storage from shared storage

3. Environment Tests

- Tempest can be used to validate the deployment. At minimum the following 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

Network Architecture

The Dell Red Hat OpenStack Cloud Solution with Red Hat Enterprise Linux OpenStack Platform uses either of the recommended Ethernet switches, presented in [Table 9: TOR Switches](#) on page 23, 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 9: TOR Switches

| Recommended Switches | Alternate Switches |
|------------------------------------|----------------------------------|
| Dell Networking S3048 1/10-Gigabit | Dell Networking S55 1/10-Gigabit |
| Dell Networking S4048 10-Gigabit | Dell Networking S4810 10-Gigabit |

This Reference Architecture is used to support consistency in rapid deployments through the minimal differences in the network configuration. Please contact your sales representative to find out if there other viable options available.

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 data network consists of the server NICs, the 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 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 Reference Architecture. The following component blocks make up this network:

- [Server Nodes](#) on page 23
- [Access Switch or Top of Rack \(ToR\)](#) on page 24
- [Aggregation Switches](#) on page 25
- [Core](#) on page 25
- [Layer-2 and Layer-3 Switching](#) on page 25
- [vLANs](#) on page 25
- [Out of Band Management Network](#) on page 26
- [Dell OpenSwitch Solution](#) on page 26

Server Nodes

Server connections to the network switches are setup to be highly available.

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 three are recommended for the Solution. The OpenStack Controller, Compute nodes, Red Hat Ceph Storage nodes, and Solution Admin Host can use any of the three channel bonding modes:

- `active-backup` (mode = 1)
- `balance-xor` (mode = 2)
- `802.3ad` or `LACP` (mode = 4)



Note: Other modes, such as `balance-rr` (mode=0), `broadcast` (mode=3), `balance-tlb` (mode=5), and `balance-alb` (mode=6), are not supported.

For all nodes, the endpoints are terminated to switch ports, that have been configured for the particular channel bonding mode, across two Dell Networking S4048s that are built up with a VLTi across them. The configuration settings are explained in greater detail in the *Dell Red Hat OpenStack Cloud Solution Deployment Guide*.

Table 10: Channel Bonding Modes Supported

| | Channel Bonding Type | | |
|----------------------------|-------------------------------------|-----------------------------------|------------------------------------|
| Node Type | <code>active-backup</code> (Mode 1) | <code>balance-xor</code> (mode 2) | <code>802.3ad</code> (LACP mode 4) |
| Solution Admin Host | Yes | Yes | Yes (solution default) |
| OpenStack Controller Nodes | Yes | Yes | Yes (solution default) |
| OpenStack Compute Nodes | Yes | Yes | Yes (solution default) |
| Red Hat Ceph Storage Nodes | Yes | Yes | Yes (solution default) |

A single port is an option when Bonding is required. However, it is not used in the Solution. 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 sales representative to find out if there other options available.

Access Switch or Top of Rack (ToR)

The servers connect to ToR switches. Typically there are two in each rack. The switches recommended by Dell are:

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

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

Check with your Dell Sales Representative for other options.

Aggregation Switches

For a deployment from one to three racks of 10G servers, Dell recommends the Dell Networking S4048 as the aggregation switch. It is both 10GbE and 40GbE capable.

The 40GbE interfaces on the S4048 could be converted into four 10GbE interfaces, thereby converting this switch into 64 10GbE-capable ports. ToR switches connect to aggregate switches via uplinks of 10GbE interfaces from the ToR Dell Networking S4048 to the Dell Networking S3048.

Dell's recommended architecture uses Virtual Link Trunking (VLT) between the two Dell Networking S4048 switches in each rack and then aggregation to a core switch. This feature 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. Running 40GbE Ethernet switches, like the Dell Networking Z9100, in aggregation can achieve a scale of up to hundreds of 1G deployed nodes.

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

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

When designing a large deployment, Dell recommends the Dell Networking S4048 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. It can aggregate up to 15 racks of high-density PowerEdge R630 and R730xd servers. The rack-to-rack bandwidth needed in OpenStack would be 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. Details on this topic are 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 Reference Architecture uses layer-2 as the reference up to the aggregation layer. That is why VLT is used on the aggregation switches. The Red Hat Enterprise Linux OpenStack Platform Installer 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 Enterprise Linux OpenStack Platform Installer 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, by simply adding the gateway to the network and updating the iDRAC these tools can be used.

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 RESTful API and Graphical User Interface access

vLANs

This Reference Architecture implements at a minimum eight (8) separate Layer 2 vLANs:

- **Management/Out of Band (OOB) Network** - iDRAC connections can be routed to an external network. All OpenStack HA Controllers need direct access to this for IPMI operations.

- **Internal Networks vLAN for Tenants** - Sets up the backend network for Nova and the VMs to use.
- **Public API Network vLAN** - Sets up the network connection to a router external to the cluster. The network is used for the front-end network for routable traffic to individual VMs, access to the OpenStack API, RADOW Gateway, and the Horizon GUI. Depending upon the network configuration these networks may be shared, or routed as needed. Access to the External API Network by the Tempest Test Node is required.
- **External Network vLAN for Tenants** - Set up a network that will support the floating IP's and default external gateway for tenants and virtual machines. This connection is through a router external to the cluster.
- **Provisioning Network vLAN** - Connects a NIC from all nodes into the fabric, used for setup and provisioning of the OpenStack servers.
- **Private API Network vLAN** - Used for communication between OpenStack Controllers, Tempest Test Node, and Compute nodes for RESTful API and cluster communications.
- **Storage Network vLAN** - Used by all the nodes for the data plane reads/writes to communicate to OpenStack Storage, setup, and provisioning of the Ceph storage cluster, and when included, the Dell Storage PS Series or SC Series arrays.
- **Storage Clustering Network vLAN** - Used by all Storage nodes for replication and data checks (Ceph clustering).

Out of Band Management Network

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

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

- The highly available servers, the controllers, use it for a heartbeat to ensure the servers are all up and functioning
- 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. These are out of scope of this solution.

Dell OpenSwitch Solution

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

You are expected to ensure that the switches conform to these requirements, and that they are configured according to this Reference Architecture's guidelines. The following list of requirements will enable other brands of switches to properly operate with Dell's required tools and configurations:

- Support for IEEE 802.1Q vLAN traffic and port tagging
- Support using one untagged and multiple tagged vLANs on the same port
- 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 Link Aggregation Groups (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

Solution Bundle

This core architecture provides prescriptive guidance and recommendations jointly engineered by Dell and Red Hat for deploying Dell Red Hat OpenStack Cloud Solution 7 with Dell infrastructure.

Our aims 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 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 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 Ceph Cluster](#) on page 29, and represents the base from which all optional components and expansion of the Solution are built upon.

Using the recommended R630 and R730xd servers, 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 two-to-one CPU core oversubscription, still have local ephemeral storage under subscribed, and memory just starting to be oversubscribed.

As 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 on 10GbE bonds for data networks and the network switches are setup for HA. The Out of Band Management network is not HA and is 1GbE. Please review and discuss with your Sales Team the specifics.

Solution Bundle Rack Layout

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

See [Table 4: Controller Node Hardware Configurations – PowerEdge R630](#) on page 15 and [Table 7: Storage Node Hardware Configurations – PowerEdge R730xd](#) on page 16 for hardware configurations. The Solution Bundle includes:

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

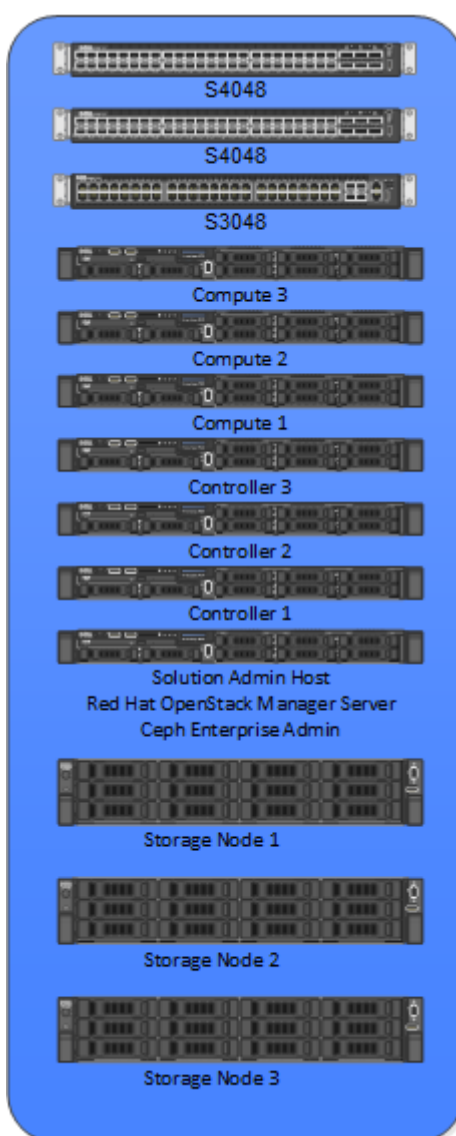


Figure 3: Solution Bundle and Ceph Cluster

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 results in the loss of more than one data replica.

There are two services in the Red Hat Ceph Storage cluster:

- **Object Storage Daemon (OSD)** - Running on Storage nodes, the OSD serves data to the Ceph 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; contact your Dell 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 1.3, which also includes the Red Hat Ceph Storage Admin Node (Calamari). The Red Hat Ceph Storage Admin Node 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 this Solution has been designed to support production-ready servers with a highly available network configuration.

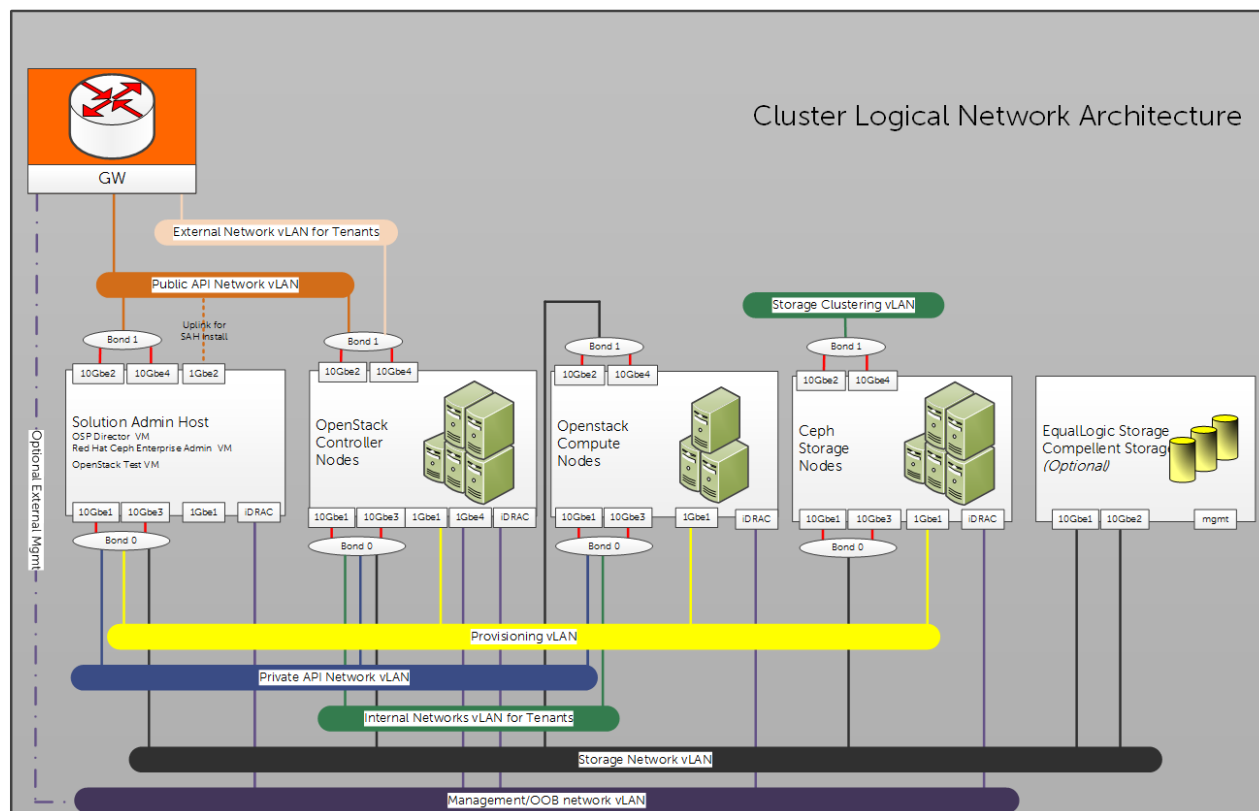


Figure 4: Cluster Network Logical Architecture with Optional Dell Storage PS Series

The node type will determine how the switches are configured in delivering the different networks. [Table 11: OpenStack Node Type to Network 802.1q Tagging](#) on page 31 and [Table 12: Storage Node Type to Network 8.2.1q Tagging](#) on page 31 outline the networks to the node types. The Management/OOB network is used by the Cluster Software to manage the OpenStack Controllers; therefore, they are the only ones that need direct connections. All iDRACs are plugged into this network without using tagging.

Table 11: OpenStack Node Type to Network 802.1q Tagging

| Network | Solution Admin Host | OpenStack Controller | OpenStack Compute | Red Hat Ceph Storage |
|--|----------------------------------|----------------------|---------------------|----------------------|
| Provisioning vLAN | Connected, Tagged | Connected, Untagged | Connected, Untagged | Connected, Untagged |
| Public API Network vLAN | Connected, Untagged ⁵ | Connected, Tagged | Not Connected | Not Connected |
| External Network for Tenant vLAN | Not Connected | Connected, tagged | Not Connected | Not Connected |
| Private API Network vLAN | Connected, Tagged | Connected, Tagged | Connected, Tagged | Not Connected |
| Internal Networks for Tenants vLAN | Not Connected | Connected, Tagged | Connected, Tagged | Not Connected |
| Storage Network vLAN | Connected, Tagged | Connected, Tagged | Connected, Untagged | Connected, Untagged |
| Storage Clustering vLAN | Not Connected | Not Connected | Not Connected | Connected, Untagged |
| Management/OOB Network vLAN | Not Connected | Connected, Untagged | Not Connected | Not Connected |
| iDRAC physical connection to the Management/OOB vLAN | Connected, Untagged | Connected, Untagged | Connected, Untagged | Connected, Untagged |

Table 12: Storage Node Type to Network 8.2.1q Tagging

| Network | PS Series Array | SC Series Enterprise Manager | SC Series Array |
|------------------------------------|---------------------|------------------------------|---------------------|
| Provisioning vLAN | Not Connected | Not Connected | Not Connected |
| Public API Network vLAN | Not Connected | Connected, Untagged | Not Connected |
| External Network for Tenants vLAN | Not Connected | Not Connected | Not Connected |
| Private API Network vLAN | Not Connected | Not Connected | Not Connected |
| Internal Networks for Tenants vLAN | Not Connected | Not Connected | Not Connected |
| Storage Network vLAN | Connected, Untagged | Connected, Untagged | Connected, Untagged |

⁵ The 1Gbe port is used for provisioning the SAH during its installation only and can be disconnected afterwards.

| Network | PS Series Array | SC Series Enterprise Manager | SC Series Array |
|--|-----------------|------------------------------|-----------------|
| Storage Clustering vLAN | Not Connected | Not Connected | Not Connected |
| Management/OOB Network vLAN | Not Connected | Not Connected | Not Connected |
| iDRAC physical connection to the Management/OOB vLAN | Not Connected | Not Connected | Not Connected |

Solution Admin Host (SAH) Networking

The Solution Admin Host has internal bridged networks for the Virtual Machines.

The Solution Admin Host is physically connected to the following networks:

1. Public Network - used for:

a. Inbound Access -

- a.** HTTP/HTTPS access to the OpenStack Foreman Installer Node
- b.** HTTP/HTTPS access to the Red Hat Ceph Storage Admin Node
- c.** Optional - SSH Access to the OpenStack Foreman Installer Node and Red Hat Ceph Storage Admin Node

b. Outbound Access

- a.** HTTP/HTTPS access for Ceph and Red Hat Updates
- b.** Used by the Tempest Test Node to run test using the OpenStack public API

2. Provision Network - Used by the OpenStack Foreman Installer Node to service DHCP to all hosts, provision each host, and act a proxy for the updates host.

3. Private API Network - Used by the Tempest Test Node to run tests against the OpenStack private API

4. Storage Network - Used by the Red Hat Ceph Storage Admin Node to provision, monitor and manage the Ceph Cluster.

Figure 5: Solution Admin Host Internal Network Fabric on page 33 displays how the networks are bridged inside the Solution Admin Host. As the Provision, Private API, and Storage networks come in on the same physical interface, 802.1q tagging is configured on the SAH and corresponding switch ports. This Reference Architecture 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.

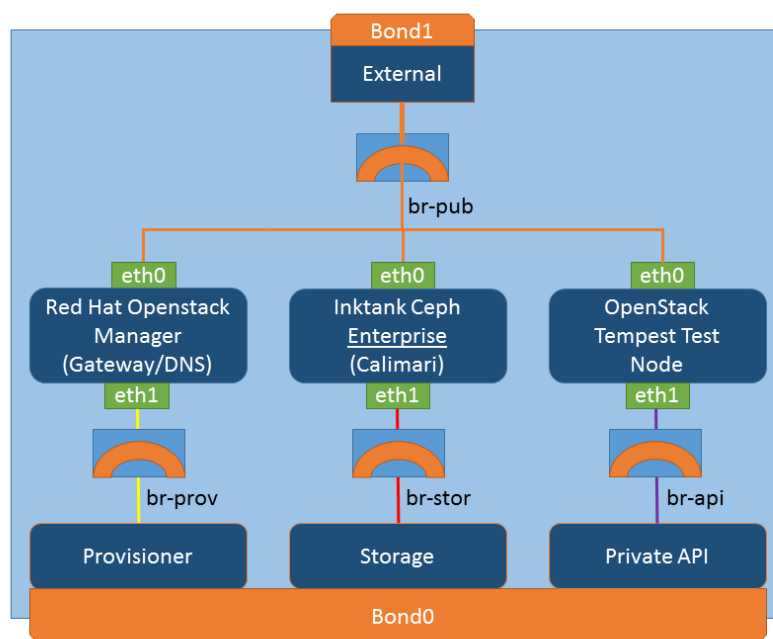


Figure 5: Solution Admin Host Internal Network Fabric

Optional Solution Configurations

Optional component changes require that the base solution be modified. For example, you can use either the PowerEdge R430 and R730 as OpenStack Compute nodes instead of the PowerEdge R630.

If you change the PowerEdge R630 OpenStack Computes to either/or PowerEdge R430 or R730 there are no networking changes; just a change of the server hardware. When ordering the servers, the configuration must be similar to the 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 the Storage backend software is setup 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 has Dell Storage PS Series and SC Series Storage Arrays; this can be one or both depending on your Application and Storage needs. Prior to ordering, your Dell sales representative will work to find the proper configuration for your needs.

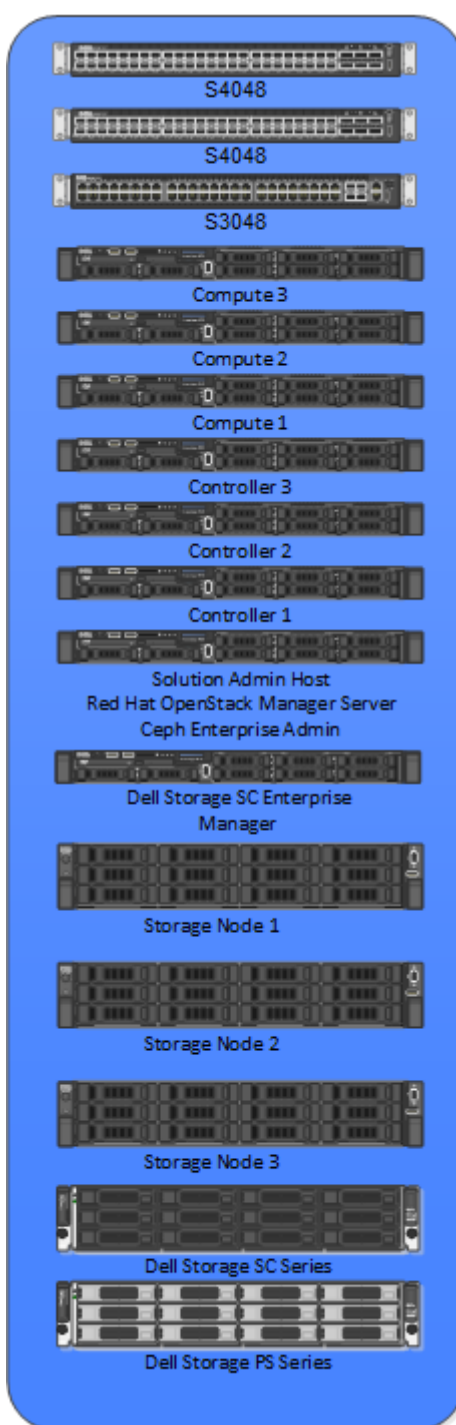


Figure 6: Solution with Optional Dell Storage

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. The Compute nodes must have access to the Storage nodes through the Storage Network vLAN in order for the iSCSI driver on that node to interact with the volumes associated to Virtual Machines hosted by that node. Dell Storage PS Series Arrays are connected to the Storage Networking vLAN untagged only, and all other nodes use the same layout as in [Table 11: OpenStack Node Type to Network 802.1q Tagging](#) on page 31 and [Figure 4: Cluster Network Logical Architecture with Optional Dell Storage PS Series](#) on page 30.

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

The Dell 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 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 the iSCSI driver on that node to interact with the volumes associated to Virtual Machines hosted by that node.

Dell SC Series Arrays are connected to the Storage Networking vLAN untagged only and all other nodes will use the layout as in [Figure 7: Cluster Network Logical Architecture with Optional Dell Storage SC Series](#) on page 36.

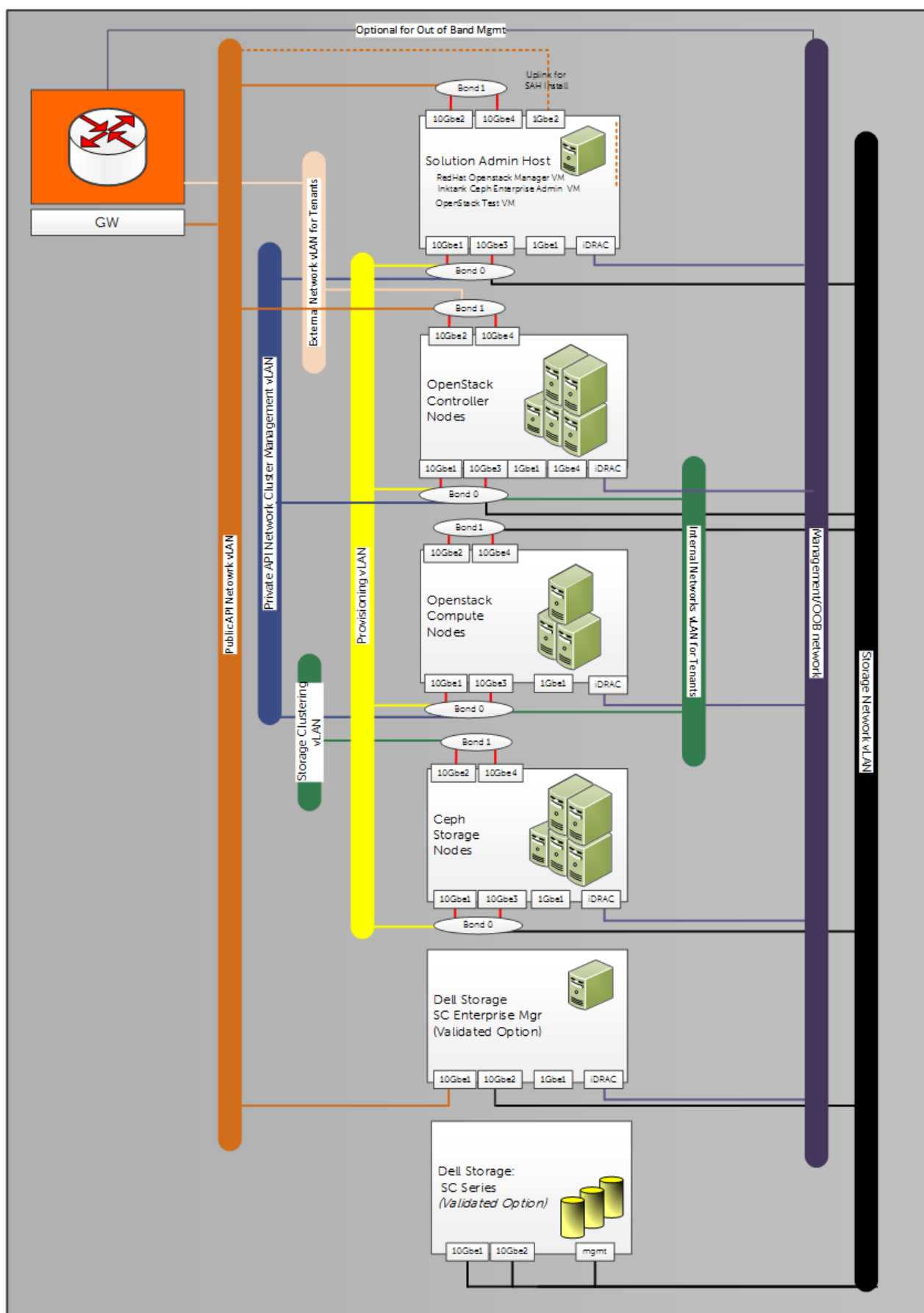



Figure 7: Cluster Network Logical Architecture with Optional Dell Storage SC Series

Solution Bundle Expansion

The Solution bundle can be expanded by adding Compute nodes, Storage nodes, or Dell Storage (PS Series or SC Series) Arrays. Using this, one could 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 as described in the networking section, additional TOR and management switches in each rack, and the appropriate power and cooling. Expansion beyond a total of three (3) racks will need to be designed and configured based on your requirements. Please work with your 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 Sales Representative.


Rack 1

Base Solution with Ceph, optionally Dell Storage PS Series:

- 2 S4048 TOR Switches
- 1 S3048 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 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 S4048 aggregation switches, depending upon your load requirements
- 2 S4048 TOR Switches
- 1 S3048 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 Arrays, or
- Dell Storage PS Series 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 TOR Switches

- 1 S3048 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 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 sales representative to properly architect these large deployments

Update History

The following changes have been made to this guide:

- [Initial Release](#) on page 39
- [Version 1](#) on page 39
- [Version 2](#) on page 39
- [Version 3](#) on page 39
- [Version 4](#) on page 40

Initial Release

First Reference Architecture for the Dell Red Hat OpenStack Cloud Solution with Red Hat Enterprise Linux™ OpenStack Platform

Version 1

Update to support:

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

Version 2

Updated as follows:

- 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 PowerEdge R630 and R730xd
- Removed support for PowerEdge R620, R720, and R720xd
- Added support for OpenStack Neutron
- Removed support for Nova-Network
- Removed the POC from the Solution

- Updated to Kilo
- Updated to RHEL OSP 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:

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

References

Additional information can be obtained at <http://www.dell.com/learn/us/en/04/solutions/red-hat-openstack> or by e-mailing openstack@dell.com.

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

To Learn More

For more information on the Dell Red Hat OpenStack Cloud Solution with Red Hat Enterprise Linux™ OpenStack Platform visit <http://www.dell.com/learn/us/en/04/solutions/red-hat-openstack>.

© 2014-2015 Dell Inc. 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 and its affiliates cannot be responsible for errors or omissions in typography or photography. Dell's Terms and Conditions of Sales and Service apply and are available on request. Dell service offerings do not affect consumer's statutory rights.

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