Accelerating High-Performance OpenStack Deployments with EMC XtremIO All-flash Storage and VMware

WHITE PAPER

Table of Contents

Executive Summary 3

Accelerating Open Source Development with EMC XtremIO All-Flash Storage 4

- Introduction to OpenStack 4
- Introduction to VMware Integrated OpenStack 4
- Introduction to EMC XtremIO 5
  - X-Brick 5
  - Scale-Out Architecture 6
  - Content-Aware Uniform Data Placement 6
  - In-Memory Metadata 6
  - Inline Always On Data Services 6
  - Optimized VM Cloning 6

VMware Integrated OpenStack and EMC XtremIO Storage Integration 7

- VAAI 7
- OpenStack Storage 8
  - Ephemeral 8
  - Block 8
  - Image 8
  - Object 8
- VMDK Driver 8

VMware Integrated OpenStack and XtremIO Reference Architecture 9

- Overview 9
- Logical View of Physical Infrastructure 9
- Logical View of Virtual Infrastructure 10
- OpenStack Component Integrations with VMware Technology 11
- VMware Integrated OpenStack Control Plane Overview 12

Performance Metrics 12

- Test 1: VMware Integrated OpenStack Control Plane Deployment 13
- Test 2: Instance Deployment 13
- Test 3: Deploy a Large Number of Instances 13
- Test 4: Image Imports 13

Summary 14

Executive Summary

OpenStack is an open source framework that provides a fast, agile environment for developing and managing the next-generation applications which enable public and private clouds. VMware vSphere® is a virtualization platform that reduces costs, increases agility, and enhances the productivity achievable with compute, network, and storage investments. In 2014, VMware announced VMware Integrated OpenStack, which allows VMware administrators to bring the speed and agility of a public cloud programming experience to their organization's software developers.

In addition to merging industry-leading data center virtualization with an open, agile cloud framework, VMware Integrated OpenStack provides deep integrations between VMware vSphere and EMC® XtremIO all-flash storage that simplify and accelerate an extremely high performance OpenStack deployment. This paper explores how XtremIO all-flash storage uses a scale out architecture, in-memory metadata, and agile integrated Copy Data Management to enable unprecedented agility, efficiency, and performance gains in VMware Integrated OpenStack environments.

Accelerating Open Source Development with XtremIO All-Flash Storage

VMware recognizes that open source APIs can provide developers in enterprise IT organizations with an experience similar to public clouds, and has helped lead the creation of several open frameworks. In cases where customers have seen strong potential in an existing framework, VMware has enabled the use of those frameworks through products such as vSphere Big Data Extensions™ (Hadoop) and Pivotal CF (Cloud Foundry). In 2014, VMware added OpenStack to this list with the announcement of the VMware Integrated OpenStack distribution of the open source OpenStack code.

While an open source framework provides many advantages, OpenStack's utilization of available resources is inherently inefficient because it must create and deploy an image copy to run every virtual machine (VM) in the infrastructure. The large number of base image copy operations adds significant network overhead, impacts application performance, and creates storage sprawl.

VMware Integrated OpenStack overcomes this performance impact through tight integration with EMC XtremIO all-flash storage technology. The virtual copy function in XtremIO is an in-memory metadata operation, so no copying actually occurs in the data plane. Offloading the copy operations to XtremIO arrays also eliminates the image management traffic between host and storage, and the related performance overhead. That advantage continues as the deployment is scaled. The XtremIO scale-out architecture does not require performance optimization to accommodate growth. The result is an exceptionally high-performance environment that virtually eliminates the performance reductions associated with disk I/O, and reduces the time required for deployment from months to days, or even hours. A proven reference architecture simplifies implementation and creates a robust infrastructure with the speed to concurrently handle high performance applications such as online transaction processing (OLTP), big data analytics, and virtual desktop infrastructures (VDIs).

To understand how the integrations between VMware Integrated OpenStack and XtremIO accelerate application performance, it's important to understand the three primary infrastructure components: OpenStack, VMware Integrated OpenStack, and XtremIO.

Introduction to OpenStack

OpenStack is an open source framework for creating an infrastructure-as-a-service (IaaS) cloud. It provides vendor-neutral, API access to cloud services, and is extremely well-suited to applications that are specifically designed for cloud computing environments. OpenStack users can control data center processing, storage, and networking resources through an API, a web-based dashboard, or command-line tools.

While OpenStack provides the abstraction layer for consuming infrastructure resources, it requires a virtualization infrastructure layer with compute, network, and management functionality beneath it. A plug-in model enables a choice of these technologies from a variety of vendors.

The two main stakeholders for OpenStack clouds are the application developers and the cloud infrastructure team. Application developers want to consume infrastructure resources as quickly and flexibly as possible through OpenStack APIs. The cloud infrastructure team is responsible for operating and maintaining the OpenStack cloud and infrastructure resources. Both need infrastructure operation and management tools which monitor and maintain the OpenStack level as well as the network infrastructure to simplify the management and orchestration of a dynamic environment.

Introduction to VMware Integrated OpenStack

VMware Integrated OpenStack is a DefCore-compliant OpenStack distribution that dramatically simplifies OpenStack deployment. It allows IT operations staff to produce an operational cloud with a wide range of enterprise-grade data center services in days instead of months.

VMware Integrated OpenStack enables accelerated OpenStack deployment through integrated drivers for VMware technologies and OpenStack source code that is optimized and hardened to run on VMware products. IT developers, therefore, gain the benefits of a public cloud experience through simple, vendor-neutral OpenStack APIs to a private VMware Software Defined Data Center (SDDC). Users also gain access to an ecosystem of development resources, and the freedom to run workloads across a heterogeneous infrastructure.

VMware Integrated OpenStack simplifies OpenStack operations by leveraging tested VMware technologies such as vRealize® Operations Manager™ for monitoring and troubleshooting, and vRealize LogInsight™ for diagnostics across OpenStack service logs. Users with VMware experience can install and manage an OpenStack cloud from existing VMware vSphere or VMware vCenter Server™ interfaces without extensive OpenStack expertise.

As a fully validated architecture that is tested and supported by VMware, VMware Integrated OpenStack offers the fastest and most reliable route to creating a production-grade OpenStack environment. It is available free to vSphere Enterprise Plus customers, vCloud Suite customers, and vSphere with Operations Management™ Enterprise Plus customers. A single support contract covers the entire infrastructure, including the OpenStack open source code.

VMware Integrated OpenStack joins a host of VMware and OpenStack technology integrations. As a gold member of the OpenStack foundation since 2012, VMware has also integrated OpenStack code into VMware vSphere and VMware NSX®.

VMware Integrated OpenStack includes the following core OpenStack projects:

- NOVA (Compute)
- Neutron (Network)
- Cinder (Block Storage)
- Glance (Image Storage)
- Swift (Object Storage)
- Keystone (Identity)
- Horizon (the Web Portal)
- Heat (Orchestration)
- Celiometer (Telemetry)

Introduction to XtremIO

EMC XtremIO is an all-flash enterprise storage solution that is optimized for virtual environments and purpose-designed to exploit the capabilities of flash media. Its scale-out architecture with integrated Copy Data Management supports always-on, always-inline intelligent data services that deliver industry leading data reduction, improve storage efficiencies, and boost workload and service agility.

One hundred percent in-memory, granular metadata enables class-leading performance that is uniquely consistent and predictable under realistic and varied workload conditions. Even with always-on inline data services such as deduplication and compression, XtremIO delivers consistent performance and fast, efficient VM and service provisioning for all workloads at any scale.

X-Brick

An X-Brick is the cluster building block of XtremIO. Each X-Brick contains two active-active controllers and a drive enclosure that holds 25 eMLC flash drives. Between the two controllers, the X-Brick provides ample processing resources (32 or more CPU cores), abundant RAM (512GB or more), and four host ports (two iSCSI and two Fibre Channel).

X-Bricks are available in a 5TB starter X-Brick, 10TB X-Brick, 20TB X-Brick, and 40TB X-Brick form factors.

Scale-Out Architecture

Traditional storage architectures employ a two-controller, active-passive scale-up design in which the controller becomes a performance-limiting bottleneck, especially in the case of high-performance flash media. XtremIO eliminates this bottleneck with a scale-out architecture designed for easy scalability and consistency. There is no single point of failure as the cluster expands, and adding X-Brick cluster building blocks to existing XtremIO clusters provides linear capacity and performance scalability with consistently sub-millisecond response times. Each cluster supports up to 8 X-Bricks, with up to 16 active-active controllers delivering over one million, mixed 8Kb IOPs. Inherent load balancing ensures that all applications continuously benefit from the entire performance potential of the cluster. This scale-out architecture is critical to providing comprehensive inline data services with no performance impact.

Content-Aware Uniform Data Placement

XtremIO uses a content-aware, fingerprint-based data placement algorithm to distribute data evenly across all available storage controller and drive resources. Spreading all volumes and data across available X-Bricks delivers superior, consistent, and predictable random I/O read and write performance, due to the ability to maximize parallel processing efficiencies.

In-Memory Metadata

XtremIO maintains 100 percent of its metadata in the controller RAM; there is no destaging of metadata on flash drives. In-memory metadata lets XtremIO perform metadata-heavy operations such as instant, high-performance writable XtremIO Virtual Copies (XVC). XtremIO employs a globally distributed, granular, in-memory metadata management system which scales in conjunction with the addition of X-Bricks due to the scale-out architecture employed. A high-speed Infiniband RDMA fabric connects all the controllers in a cluster to facilitate inter-controller communication and inter-X-Brick data movement.

Inline Always On Data Services

XtremIO data services are delivered inline, all the time. There is no post-processing of data, which minimizes the write operations and extends the life of flash media while delivering consistent performance. The data services include thin provisioning, global deduplication, inline compression, the XtremIO Data Protection (XDP) flash-specific data protection method, Data at Rest Encryption (D@RE), and fully writable and full performance in-memory XtremIO Virtual Copies (XVC).

Volumes created on XtremIO are thin-provisioned by default. As you write data to the volume, it is globally deduplicated and compressed prior to being written to media, thereby extending the lifetime of the flash media. The XtremIO Data Protection (XDP) flash-specific data protection method is then employed before writing any data. XDP replaces traditional RAID methods and helps lower capacity overhead, minimize write operations, and provide outstanding performance—better than any RAID level while ensuring outstanding fault-tolerance. The data is then encrypted on the drives themselves. Once you have data on your volume, you can use XVC to efficiently create and manage copies of the volumes, datastores, or VMs.

XVC enables high-performance, capacity-efficient copy operations. The Virtual Copies are made near-instantaneously, perform at the same level as the parent volume, are writeable, and do not require additional storage space because they are inherently duplicate. XVCs are created by in-memory metadata operations only, at RAM speed. No disc IOs are generated while making an XVC, ensuring consistent performance for workloads sharing the storage infrastructure.

Optimized VM Cloning

When a VM administrator issues a command to clone a VM, or a number of them, a VMware Storage APIs for Array Integration (VAAI) XCOPY command is issued to the underlying storage. XCOPY moves the copy operation to the array, eliminating the high data traffic between the host and storage tiers during copy operations. XtremIO uses an in-memory metadata architecture that allows XCOPY to take place completely in memory. Since clones are inherently duplicate, this is a completely metadata operation. As a result, creating VM copies does not consume any disc I/Os, enabling near-instant copies while reducing capacity requirements and eliminating any performance impact. As a result, VM copy operations can be carried out without any infrastructure impact.

VMware Integrated OpenStack and EMC XtremIO Storage Integration

Integrations between VMware vSphere (a key component of VMware Integrated OpenStack) and XtremIO enable enhanced performance that is unmatched by open or proprietary hypervisor-based OpenStack solutions. The following three integration areas contribute to significant performance gains in VMware Integrated OpenStack deployments:

- VAAI
- OpenStack Storage
- VMDK Driver

VAAI

VMware vSphere Storage APIs – Array Integration (VAAI), also referred to as hardware acceleration or hardware offload APIs, are a set of APIs to enable communication between VMware vSphere ESXi™ hosts and storage devices. The APIs define a set of "storage primitives" that enable the ESXi host to offload certain storage operations to the array, which reduces resource overhead on the ESXi hosts and can significantly improve performance for storage-intensive operations such as storage cloning and zeroing.

In the VMware Integrated OpenStack environment, XtremIO uses VAAI to relieve VMware vSphere of the task of performing copy operations. Offloading copy operations to the array in place of VMware vSphere enables much faster processing because the copy data need not be transferred between the VMware vSphere host and the array, and because XtremIO performs copy operations much faster than any hypervisor due to its in-memory metadata management architecture.

Building your environment on storage which supports VAAI therefore provides a critical performance advantage, dramatically increasing the performance and efficiency of VMware Integrated OpenStack. The combination of VAAI support, a wholly in-memory metadata architecture, and always-on inline data services let XtremIO optimize many common VM operations more efficiently than other flash or non-flash VAAI-compliant storage options.

XtremIO supports all VAAI primitives:

- **Full Copy/XCOPY** delivers hardware-accelerated data copying by performing all duplication and migration within the physical array. Customers can achieve considerably faster job completion times for VMware Storage vMotion and VM creation and deployment (from templates and virtual machine cloning) operations. Full Copy is also referenced as the XCOPY SCSI command.
- **Zero Blocks/Write Same** delivers hardware-accelerated zero initialization, greatly reducing common I/O tasks such as those used in creating new virtual machines. This feature is especially beneficial when performing routine application-level Block Zeroing.
- **Atomic Test & Set (ATS)/Hardware-Assisted Locking** delivers improved locking controls in the Virtual Machine File System (VMFS), allowing far more virtual machines per datastore and reducing simultaneous block virtual machine boot times. Hardware-assisted locking improves the performance of common tasks such as virtual machine migration, powering many virtual machines on or off, and creating a virtual machine from a template.
- **Block Delete** allows space to be reclaimed using the SCSI UNMAP feature, which enables the reclamation of blocks of thin-provisioned LUNs by informing the array that specific blocks are obsolete (in ESXi 5.x and later hosts).

OpenStack Storage

OpenStack users must become familiar with four types of storage: Ephemeral, Block, Image, and Object. VMware Integrated OpenStack ties the first three to the XtremIO array, which dramatically improves OpenStack storage performance. Object storage requires integration through a third party, and is therefore not linked to XtremIO through VMware Integrated OpenStack.

Ephemeral

The root disk, which is where the system partition of the operating system resides, is ephemeral storage. It is the initial disk given to an OpenStack VM, which is known as an "instance" in OpenStack terminology. This storage type exists only as long as the VM on which it runs exists—hence the term "ephemeral". When the VM is removed, the disk is deleted as well. Ephemeral storage is used with the Nova OpenStack service and the VMware vCenter Driver.

Block

Block storage is persistent because it is allocated separately from the instance. The disk remains even if the instance to which it is attached is destroyed. This storage type is commonly used in database deployment. Developers allocate a root disk, a separate disk for the database data, and a third disk for log information. This storage type uses the VMware VMDK Driver; additional disks are added with the OpenStack Cinder service.

Image

An OpenStack Image is similar to a VM template. The developer installs an operating system on a VM, and exports that VM as an image to provision additional VMs. The image provides a point-in-time copy that can be installed over and over again, speeding application deployment. Image storage performance is critical because a network might have hundreds or even thousands of users who need to create VMs on a regular basis. Inadequate storage performance will create an image copy bottleneck that reduces user productivity. Image storage is managed with the OpenStack Glance service. Like the Block storage type, Image storage uses the VMware VMDK Driver.

Object

Object storage stores files as addressable objects with individual URLs. This approach is especially useful for web applications and data archival, because it eliminates the need to know where resources are located in the file system. Stored pictures for a web application, for example, can be referenced by their URL. Object storage is managed with the OpenStack Swift service.

VMDK Driver

The VMware VMDK driver leverages the storage abstractions (datastores) available in the VMware vSphere architecture to simplify storage management in OpenStack. VMware has worked closely with its storage partner ecosystem to establish standard mechanisms for accessing different types of storage, and the protocols that the VMDK driver supports are NFS, iSCSI, Fibre Channel, and VMware Virtual SAN™ .The VMware VMDK driver allows OpenStack administrators to use VMware vSphere datastores to provide block storage to the Cinder service and image storage to the Glance service.

The VMDK driver is the only storage driver that is specified in the Nova, Cinder, and Glance configuration files instead of leveraging individual array drivers for each device in the environment. The OpenStack administrator then indicates the names of the VMware vSphere datastores for OpenStack to use in regular expression format in the configuration files.

VMware Integrated OpenStack builds on the benefits of the VMware VMDK Driver by providing a simple graphical user interface for managing the VMware vSphere datastores that provide storage to Cinder and Glance. This storage management GUI is part of the VMware Integrated OpenStack plugin for the VMware vSphere Web Client. IT administrators do not need to manually edit configuration files at the command line, or learn a completely different interface to perform OpenStack storage administration operations.

The VMware VMDK driver supports the standard storage-related operations that an OpenStack user would expect to be available:

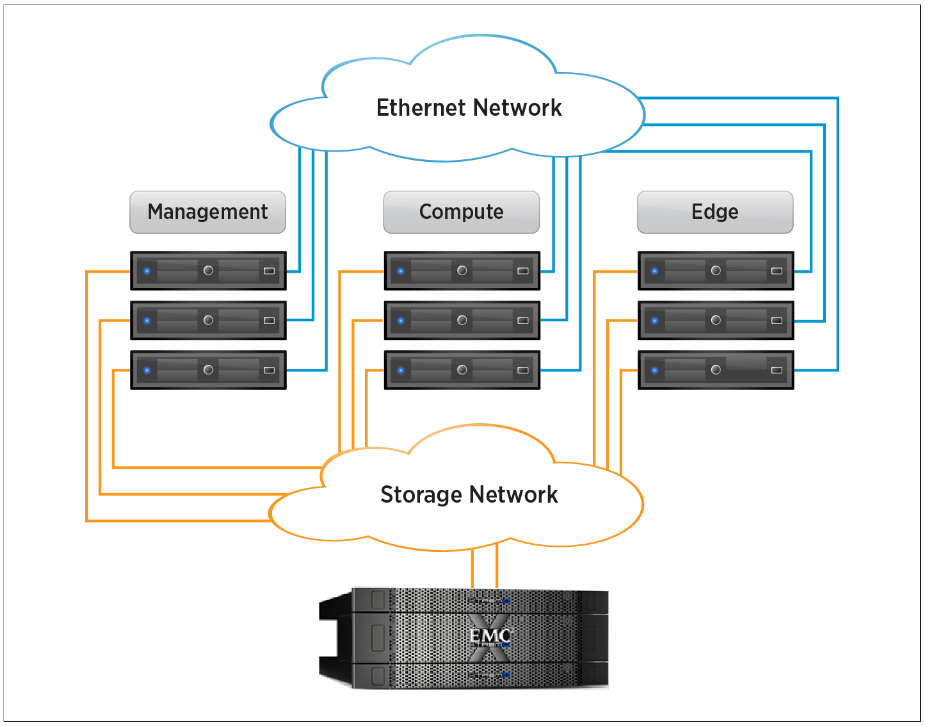
- Create volume
  - Supported VMDK types – thin, thick, eagerZeroedThick
  - From scratch
  - From glance image
  - From an available source volume – full clone and linked clone
  - From a snapshot – full clone and linked clone
- Attach volume to an instance
- Detach volume from an instance
- Create a volume snapshot
- Delete a volume snapshot
- Upload an available volume as a glance image
- Delete a volume

VMware Integrated OpenStack and XtremIO Reference Architecture

Overview

This section provides a summary of the infrastructure used during the performance testing for VMware Integrated OpenStack deployed on a VMware vSphere infrastructure and EMC XtremIO all-flash array.

Logical View of Physical Infrastructure

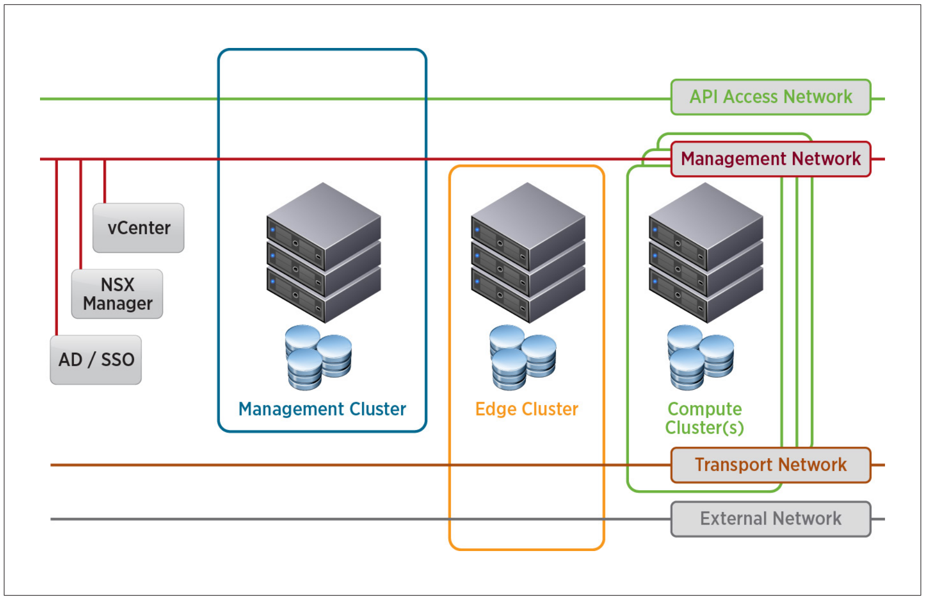


VMware Integrated OpenStack is deployed with a minimum of three VMware vSphere clusters:

- 1 Management cluster for the OpenStack control plane
- 1 Compute cluster for the OpenStack tenant workloads
- 1 Edge cluster for the VMware NSX Edge routers

The clusters leveraged 8 Gb Fibre Channel connectivity for storage traffic with the EMC XtremIO array. Management and application data traversed the 10 Gb Ethernet network.

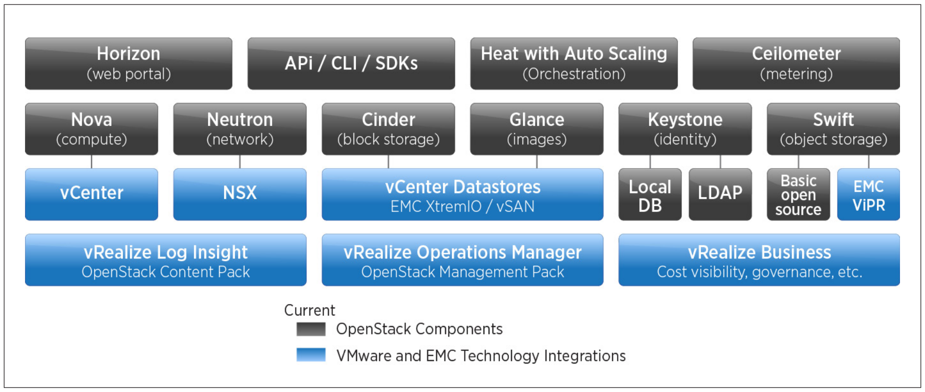
Logical View of Virtual Infrastructure



The virtual infrastructure design consisted of the aforementioned clusters, an Active Directory server, a VMware NSX Manager, and a vCenter server. There were four networks allocated as follows:

- API Access Network – the network that OpenStack cloud tenants use.
- Management Network – the network that the OpenStack control plane components use to communicate with each other.
- Transport Network – the network that VMware NSX leverages for VXLAN overlay tenant communications.
- External Network – the network that tenant workloads utilize to communicate with the outside world. Also, Floating IPs are allocated from this network.

OpenStack Component Integrations with VMware Technology



VMware Integrated OpenStack is based on the OpenStack Kilo release and includes the most commonly used projects from that release, including:

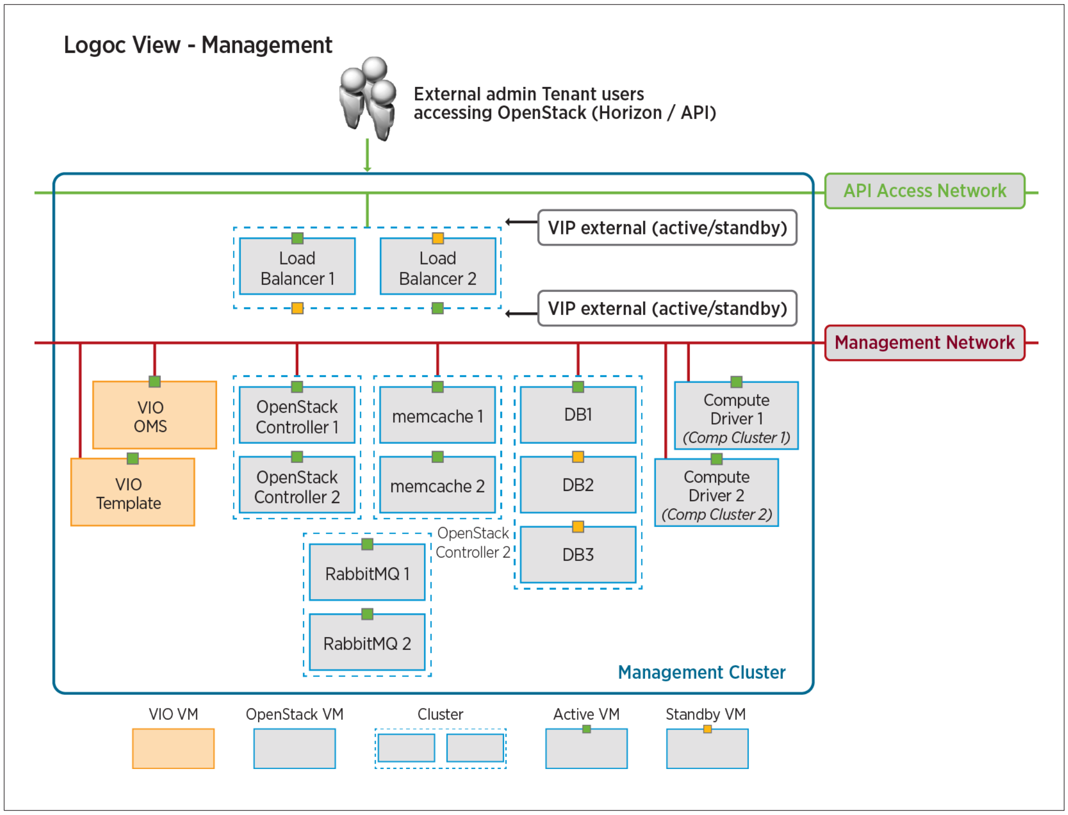
- Nova – provision tenant instances (OpenStack terminology for virtual machines).
- Neutron – provide network connectivity to tenant workloads.
- Cinder – block storage volumes (OpenStack terminology for virtual disks). EMC XtremIO storage was used for the block storage.
- Glance – provide a catalog of images (OpenStack terminology for virtual machine templates). EMC XtremIO storage was used for image storage.
- Keystone – identity management, which can leverage an LDAP source.
- Swift – object storage, which can utilize EMC ViPR.
- Ceilometer – provides metering capabilities to trigger alarms based on resource utilizations.
- Heat – the orchestration service that allows tenants to define their application infrastructure in a single YAML definition file instead of provisioning components piece-by-piece manually. Heat also supports autoscaling for dynamic expansion and retraction of application infrastructure based on Ceilometer alarm triggers.

VMware also added OpenStack intelligence to its vRealize management solutions to assist OpenStack administrators with monitoring, log analysis, and chargeback/showback capabilities.

- VMware vRealize Log Insight – provides syslog aggregation and analysis on a per-OpenStack component basis.
- VMware vRealize Operations Manager – performs health check and polls control plane services to verify that they are up and running successfully.
- VMware vRealize Business – uses real-world data to calculate the cost of operating your cloud so that administrators can chargeback internal departments for the use of the cloud resources and also perform comparisons with public cloud vendors.

These management functions can be a challenge to address in OpenStack clouds since the cloud framework itself does not include these capabilities. OpenStack administrators often have to develop their own monitoring tools, customize open source solutions, or leverage a third-party solution to effectively report on their OpenStack clouds.

VMware Integrated OpenStack Control Plane Overview



High availability is built in to the control plane. Instead of relying solely on VMware vSphere availability features, we have a pair of HAProxy load balancer VMs, a pair of controller VMs, a pair of memcached VMs, and a pair of RabbitMQ VMs. The database cluster consists of three MariaDB VMs that implement a Galera cluster, which has a quorum of three nodes. Hence three nodes instead of two.

The Compute Driver VM is the only control plane component that will be solely protected by vSphere HA. There is one Compute Driver VM per Compute cluster managed by OpenStack. In production, we expect that customers will have more than one cluster in the OpenStack deployment. So, in the unlikely event of the failure of the hypervisor hosting Compute Driver1 VM, the remaining cluster is able to service user requests in the short amount of time that vSphere HA takes to recover the failed compute driver VM.

Performance Metrics

The performance improvements gained through VMware vSphere and XtremIO integrations in a VMware Integrated OpenStack environment are proven and extensive, as demonstrated by the following performance tests implemented by EMC.

- VMware Integrated OpenStack Control Plane Deployment
- Instance Deployment
- Deploy a large number of instances
- Image Imports

These tests were selected due to their storage-intensive nature. They represent a sampling of common tasks performed by OpenStack administrators and end-users.

Test 1: VMware Integrated OpenStack Control Plane Deployment

Test Description: Deploy thirteen VMs that serve as the control plane for OpenStack. The VMs provide the services that allow users to programmatically provision instances, images, networks, and other resources for their applications. The deployment process had two parts:

1. Clone a VM template 13 times to form the control plane
2. Execute Ansible playbooks to configure each instance for their proper OpenStack role

Test results:

- VM Cloning Time to Completion: 1 minute (or 3% of the time required on other storage platforms)
- Total Time to OpenStack Control Plane Deployment: 35 minutes (a 58% time savings)

Test 2: Instance Deployment

Test Description: Virtual machines are referred to as "instances" or "servers" in OpenStack terminology. Faster instance provisioning equates to quicker application infrastructure deployment and improved time-to-value for your software projects. Two different images were tested:

1. A 64 GB Ubuntu thin-provisioned VMDK image (2 GB storage used)
2. A 64 GB Ubuntu thick-provisioned VMDK image

Test results:

- Time to complete image 1 provisioning: 3 minutes
- Time to complete image 2 provisioning: 7:33 minutes (or 12% of the time required by similar operations on a non-vSphere based OpenStack implementation with other storage platforms)

It should be noted that the copy times were relatively close despite the differences in storage utilization due to VAAI and the XtremIO array's capabilities.

Test 3: Deploy a Large Number of Instances

Test Description: There are storage optimizations that allow OpenStack installations on VMware vSphere to speed instance deployments by leveraging VAAI for VMDK images. This test used the thin-provisioned image from Test 2.

Test results (time to completion):

- 50 instances: 1:47 minutes
- 100 instances: 3 minutes

Test 4: Image Imports

Test Description: OpenStack images are analogous to VMware VM templates; they consist of an operating system already deployed and awaiting customization via the instance deployment process. The Glance service's operations to import images can be time-consuming due to network bandwidth, storage protocol, disk array architecture, and other factors. This is another operation that benefits from VMware integrations with EMC XtremIO.

The two images from test 2 were used for this test:

1. A 64 GB Ubuntu thin-provisioned VMDK image (2 GB storage used)
2. A 64 GB Ubuntu thick-provisioned VMDK image

Test results (time to completion):

- Image 1: 36 seconds
- Image 2: 14:15 minutes

Summary

VMware has implemented a wide variety of integrations between VMware vSphere and third-party products that accelerate performance. EMC XtremIO improves VMware Integrated OpenStack's storage performance by providing consistent performance capabilities for workloads at any scale, in addition to ultra-fast, efficient VM and service provisioning and radically simpler storage management. VMware Integrated OpenStack with XtremIO storage achieves performance levels that are not possible with other open or proprietary hypervisors, or other all-flash, hybrid, or disk-based storage systems. As a result, VMware Integrated OpenStack with XtremIO storage enables rapid deployment and support for applications with the most demanding performance requirements.

Contributors

Trevor Roberts, Technical Marketing Manager, VMware Integrated OpenStack, VMware

Vikram Belapurkar, Solutions Marketing Manager, Virtualization and Hybrid Cloud, EMC

Michael Cooney, Technical Solution Engineer, XtremIO, EMC

VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com