VMware vCloud® Architecture Toolkit™
for Service Providers

Architecting a Messaging Strategy with Microsoft Exchange 2013 and the VMware Cloud Provider™ Program

Version 2.9

January 2018

Martin Hosken

VMware, Inc.
3401 Hillview Ave
Palo Alto, CA 94304
www.vmware.com

Contents

List of Tables

List of Figures

Overview

Today's dynamic organizations require a 100 percent reliable messaging infrastructure that lowers communication costs, increases productivity, simplifies administration, and decreases IT overhead. Environments that require continuous uptime and fast recovery from disaster scenarios are critical to meeting business requirements and service-level agreements (SLAs). The "being prepared" approach to providing application high availability is aimed not only at meeting customer SLAs, but also at reducing risk of revenue loss and maintaining compliance.

Planning for disasters and minimizing their impact is critical for enterprises and government agencies, and VMware Cloud Providers™ can assist in providing data center resources and services to leverage application-layer high availability solutions from Microsoft and other independent software vendors. Designing and deploying a messaging solution, such as Microsoft Exchange Server 2013, in conjunction with a highly available hybrid cloud solution from a VMware Cloud Provider can help businesses mitigate these risks and still maintain control over deployment and operational costs.

Many different approaches to architecting a VMware based hybrid cloud exist, depending on the specific use case requirements and technologies employed. However, the end goal is always the same. That is, a hybrid cloud is a cloud computing environment in which an organization provides and manages some resources in-house and has others provided externally. This vCloud Architecture Toolkit for Service Providers (vCAT-SP) solution document describes the VMware Cloud Provider Program facility solution for supporting a geographically dispersed Microsoft Exchange Server 2013 environment protected by Database Availability Group (DAG) technology, which is the new building block for highly available and disaster recoverable messaging solutions.

Target Audience

This paper is targeted towards architects, engineers, application owners, and business leaders involved in the key decision making process, and anyone else interested in guidance on designing a hybrid messaging infrastructure solution on VMware technologies.

While the design outlined in this document is targeted toward a medium or large customer with enterprise messaging workloads, this style of solution can be leveraged by smaller environments.

Use Case

Because businesses of all sizes are demanding improved high availability for virtualized business critical applications to meet SLAs, organizations are upgrading or migrating to Microsoft tier 1 applications that can leverage application availability across multiple data centers. One such offering is Microsoft Exchange Server 2013, which provides DAG features that can be leveraged for high availability and disaster recovery. Microsoft Exchange Server 2013 DAG technology is an integrated, flexible, and cost-efficient solution that can provide redundancy within and across data centers. In addition, Microsoft Exchange Server 2013 DAG solutions provide fast application failover for maximum availability and data protection of business-critical messaging systems. The migration to Microsoft Exchange Server 2013 can also replace older physical infrastructures, which are difficult to manage and expensive to support.

VMware Cloud Providers that take advantage of these application-layer high availability solutions to offer cloud consumers hybrid cloud-based virtualized on-premises and off-premises solutions are supporting the rapidly changing way that organizations deploy and manage their IT assets and resources. VMware Cloud Providers offer IT organizations a single flexible and scalable platform to manage and reduce system complexity and IT operational overhead. By providing this, application-based hybrid cloud service offerings give both consumers and providers a flexible solution with significant advantages in cost, efficiency, and availability. In addition, with fewer servers consuming less space, power, and cooling, IT organizations can gain significant capital and operational expense savings by moving from an IT infrastructure that provides high availability to tier 1 applications to a VMware Cloud Provider data center.

This vCAT-SP solution architecture focuses on deploying Microsoft Exchange Server 2013 and taking advantage of its DAG functionality to provide a hybrid cloud approach to application availability with a VMware Cloud Provider. It also focuses on helping IT organizations achieve business continuity objectives at a lower risk and cost.

Introduction to Microsoft Exchange Server DAGs

With Microsoft Exchange 2013, the data protection methods in Microsoft Exchange 2007 and 2010 have evolved into the latest version of the DAG, which represents the new building block for highly available and disaster recoverable Microsoft Exchange solutions.

A DAG is made of up to 16 mailbox servers that host a set of replicated databases and provide automatic database-level recovery from failures, issues, or outages that affect individual mailbox servers or databases. Microsoft recommends minimizing the number of deployed DAGs to simplify administration. However, with certain design factors, multiple DAGs might be required. For instance:

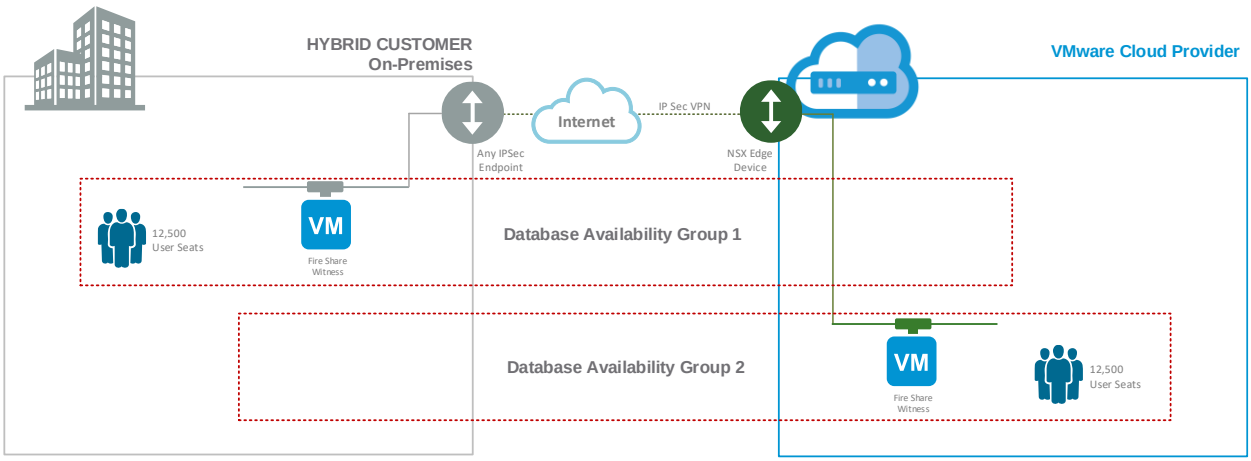
- If your design requires deployment of more than 16 mailbox servers.
- If you have active mailbox users in multiple sites (active-active site configuration).
- If you require separate DAG-level administrative boundaries for operational reasons.
- If you have mailbox servers in separate Active Directory domains (DAGs are domain bound).

In this solution architecture, due to the large number of mobile and remote workers being employed by the business, the IT organization is deploying an active-active site configuration, with active users connecting globally from multiple remote offices and through a range of mobile devices.

For this reason, the Microsoft Exchange architecture requires at least two DAGs, with each DAG spanning both sites. A file share witness is located in both the on-premises data center and in the VMware Cloud Provider Program facility to prevent database copies from becoming active in both sites, in the event of a network failure between the two facilities.

For instance, if there is a network outage between the two data center facilities, the primary site for that DAG will get two votes (the DAG member and the file share witness) as opposed to the single vote by the DAG member at the secondary data center. The majority vote retains quorum and the mailbox databases remain mounted.

Figure 1. File Share Witness Majority Vote Architecture



The Microsoft Exchange DAG feature is built on a non-shared disk architecture, with each server having its own copy of the database. This copy can be deployed on either VMFS or RDMs with log replay used to replicate data from the active to the passive nodes.

DAGs are built on top of Windows Server Failover Clustering (WSFC) technology, which provides a failover policy and quorum management. Although WSFC is required by DAGs, unlike traditional Microsoft Exchange failover cluster instances, there is no requirement to use shared disks. While VMware does not support VMware vSphere® vMotion® or VMware vSphere Distributed Resource Scheduler™ on clustered Microsoft Exchange Server virtual machines with a shared disk architecture, such as failover

cluster instances, this restriction does not apply to DAGs that are built on a non-shared disk architecture. Therefore, using VMware vSphere High Availability, vSphere vMotion, and DRS with DAGs is fully supported by VMware.

This means that with vSphere vMotion, a VMware ESXi™ host can be powered down for planned maintenance at any time without interruption to client requests. In the event of an unplanned hardware failure, vSphere HA can quickly reboot a Microsoft Exchange Server virtual machine, which can then rejoin the DAG session.

In this planned maintenance scenario, vSphere vMotion can be employed to proactively live migrate an availability group replica to a different host to allow hardware maintenance without requiring a DAG failover event. With vSphere vMotion, there is no disruption of Microsoft Exchange mailbox server services during the migration and no interruption to the client's email sync connections or any in-flight message transportation. By coupling vSphere vMotion with DAG technologies, you can eliminate the need to fail over the Microsoft cluster and reduce service interruptions for operational hardware maintenance or renewal.

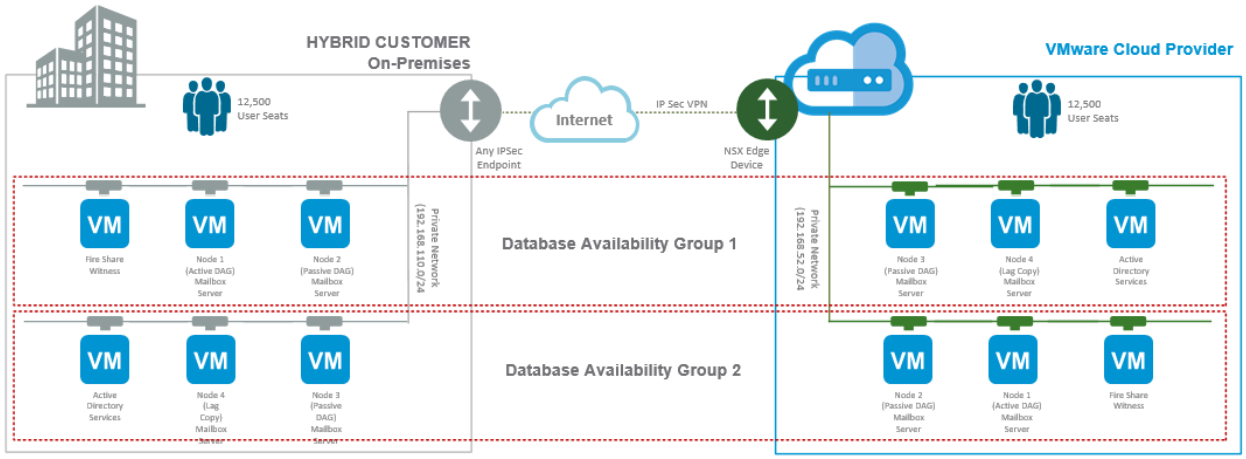
In the unplanned hardware failure scenario, the Microsoft Exchange Server environment can be vulnerable if further host failures occur during the time between the loss of a passive database and its restoration, because, depending on available bandwidth and network conditions, the resynchronization of the passive node can take a significant period of time to complete. vSphere HA helps alleviate this issue by restarting the failed passive DAG virtual machine on another available host in the VMware vSphere cluster. This facilitates a faster restore to full protection of the mailbox database and reduces the amount of time spent by the DAG in the failed state. In the event of an unplanned physical host failure, you do not need to wait for the physical host to be serviced and brought back online to restore the passive DAG copy online. Instead, vSphere HA automatically detects the host failure and immediately reboots the passive DAG virtual machine on a different available ESXi host.

With these integrated VMware mechanisms for high availability, it is possible to achieve better levels of service uptime using DAGs with vSphere than on physical hardware.

The primary objective of this document is to demonstrate how to reduce the impact of hardware and software failures by using virtualized Microsoft Exchange Server 2013 between an on-premises data center and a VMware Cloud Provider facility to architect a high availability and disaster recovery solution for tier 1 virtualized business critical mailbox databases. This maximizes continuous availability of the applications being serviced and also provides business continuity during disaster scenarios. This solution architecture also aims to:

- Demonstrate business-critical levels of high performance and availability between the IT organization's private data center and the VMware Cloud Provider's facilities.
- Provide resiliency that can meet recovery time and point objectives when faced with application, storage, network, or compute node failures.
- Demonstrate how to achieve business-continuity SLAs in partnership with a VMware Cloud Provider to lower risk and operational costs.

Figure 2. Solution Architecture Overview



Cloud Customer Requirements

This section describes the function of this sample solution architecture, and the VMware Cloud Provider Program customer profile, business, and IT organization design requirements.

5.1 Functional Requirements

To contextualize this VMware Cloud Provider Program solution architecture and its design decision points, this paper presents a use case of Rainpole.com, a fictitious organization with 25,000 employees. The employee population is split evenly between the corporate office workers and remote office and mobile workers located in geographically dispersed regions. The IT organization has access to a single on-premises data center capable of housing server and storage hardware. The primary objective of this new architecture is to mitigate this single point of failure for the organization's messaging system while maintaining control and limiting operational costs.

The company has determined that the average number of emails sent and received per day for each user is approximately 100, with an average email size of 75 KB. Each user will be assigned a 2 GB mailbox. The organization requires a deleted item retention window of 14 days to give users ample time to recover unintentionally deleted emails.

5.1.1 High Availability and Site Resiliency

By employing the services of a VMware Cloud Provider, the Microsoft Exchange infrastructure can be distributed across two or more sites. The SLA currently in place determines the degree of high availability and the placement of the Microsoft Exchange organization infrastructure.

In this solution architecture, the organization has one data center and has engaged with a VMware Cloud Provider to provide a second data center facility because the organization has determined site resiliency is required. Therefore, the Microsoft Exchange 2013 architecture is based on a multisite deployment with site resiliency. The organization has decided that a 24-hour recovery point objective is sufficient.

In addition to site resiliency, the organization wants the ability to power down a server for maintenance without affecting the user population's messaging capabilities and without switching to a server at its passive copy data center.

5.1.2 Backup and Recovery

Microsoft Exchange Server 2013 includes several features that provide native data protection that, when implemented correctly, can eliminate the need for traditional backups. Traditional backups are typically only used for long-term data storage and compliance. Disaster recovery, recovery of accidentally deleted items, and point-in-time database recovery are addressed by native features in Microsoft Exchange 2013, such as high availability database copies in a database availability group, recoverable items folders, archiving, multiple-mailbox search, message retention, and lagged database copies.

However, in the event of a server failure, recovery is necessary, and rebuilding a failed database can take hours or days when using Microsoft Exchange 2013 native data protection features. Having a backup can reduce the time required to bring a failed database back online. The downsides to using backups are the administrative overhead and the additional storage capacity required for the backup files. In addition to the disk capacity required to house the backup files, each server must have access to a restore LUN to reduce the impact on active databases while restoring database files.

For these reasons, in this solution architecture, Rainpole.com's IT organization has decided to forgo traditional backups in favor of using a Microsoft Exchange 2013 native data protection strategy.

5.1.3 Number of Database Copies

Before determining the number of database copies required by the organization, it is important to understand the two types of database copies.

- High availability database copy – Has a log replay time of zero seconds. When a change is made in the active database copy, changes are immediately replicated to passive database copies.
- Lagged database copy – Has a preconfigured delay built into the log replay time. When a change is implemented in the active database copy, the logs are copied over to the server hosting the lagged database copy, but are not immediately implemented. This provides point-in-time protection which can be used to recover from logical corruption of a database (logical corruption occurs when data has been added, deleted, or manipulated in a way the user or administrator did not expect). Lagged database copies allow up to 14 days of lagged data.

If vSphere did not provide workload mobility features, another design factor that would require consideration is database copies for hardware servicing. If only one high availability mailbox database copy is at the primary site, the operations team is required to switch over to a database copy hosted at a secondary data center when the host server needs to be powered off for servicing. Traditionally, to prevent this, you would be required to maintain a passive database at the same geographic location as the active database to support hardware maintenance. However, with DAGs, this is mitigated because vSphere vMotion and DRS are fully supported by VMware.

Despite this, a further architectural consideration is that Microsoft recommends having a minimum of three high availability database copies before removing traditional forms of backup. As Rainpole.com has chosen to forgo traditional forms of backup, they require at least three copies of each database to be implemented as part of the design. In addition, Rainpole.com's IT organization has chosen to add a fourth, lagged database copy, to protect against logical corruption.

5.2 Functional Requirements Implementation Details

This section discusses how the functional requirements are implemented as part of the solution architecture.

5.2.1 High Availability and Site Resiliency

As mentioned in Section 5.1, Functional Requirements, individual servers must be designed to be redundant and have the ability to fail over if a host fails or hardware servicing is required. In addition, in the event of a site disaster at the on-premises data center, the VMware Cloud Provider Program data center must be able to service the entire user population.

The IT organization has determined that database failover between mailbox servers within the same data center must be automatic with no loss of data. However, administrative actions must be required in the case of a site failover.

Application Architectural Overview

Before providing the vSphere architecture for this solution, it is important to fully understand the application architecture and the factors that have been employed to achieve it. This is a major consideration in designing a hybrid cloud solution on vSphere that can maximize availability and performance and minimize operational overhead. To help achieve this understanding, this section describes the architectural design of the Microsoft Exchange 2013 application.

6.1 Mailbox Servers and Database Distribution

Given the functional requirements, Rainpole.com's IT organization has determined the number of mailbox servers and the mailbox database distribution. Microsoft advises customers that each copy of a database be hosted on a separate mailbox server. Therefore, Rainpole.com needs eight mailbox servers across two DAGs to support the three highly available database copies, and one lagged copy per DAG (one copy per server).

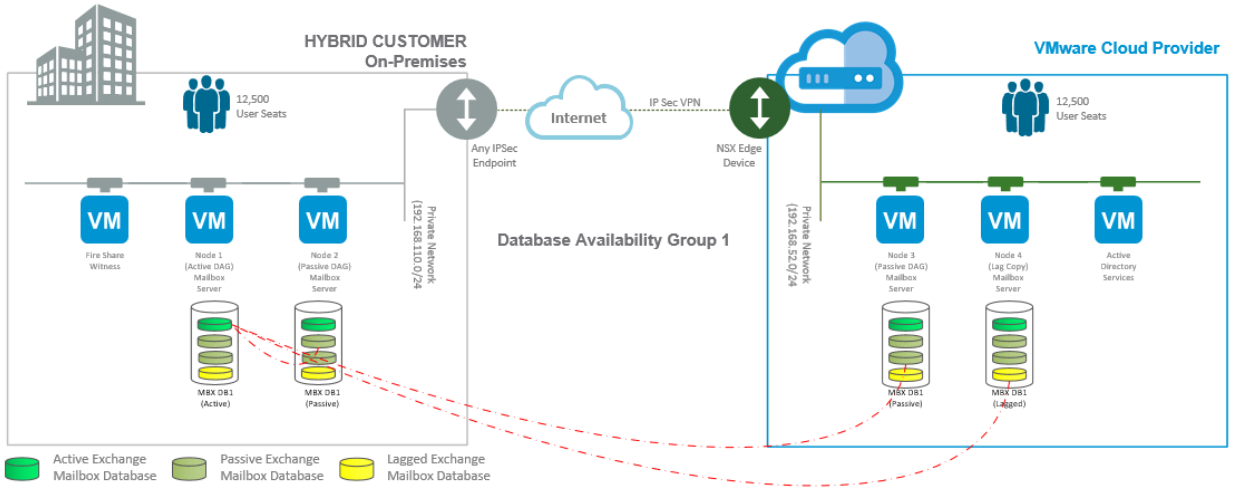
The following table illustrates the mailbox database distribution among the required mailbox servers for one of the two DAGs. The second DAG is the mirror image of the first DAG, with mailbox database 1 (DB 1) having copies one and two at the on-premises data center, and the third and the lagged copy at the VMware Cloud Provider data center facility. Other mailbox databases are to be distributed in a similar manner, balancing the workload evenly across all mailbox servers.

Table 1. Database Distribution Among Servers DAG 1

On-Premises DC	Active Server	Server 1 DB Copy	Server 2 DB Copy		VMware Cloud Provider Program DC		Active Server	Server 3 DB Copy	Server 4 DB Copy
DB 1	Server 1	1	2		DB 1		Server 3	3	Lag
DB 2	Server 2	2	1		DB 2		Server 4	Lag	3
DB 3	Server 1	1	2		DB 3		Server 3	3	Lag
DB 4	Server 2	2	1		DB 4		Server 4	Lag	3
DB 5	Server 1	1	2		DB 5		Server 3	3	Lag
DB 6	Server 2	2	1		DB 6		Server 4	Lag	3
…					...				
DB 18	Server 2	2	1		DB 18		Server 4	Lag	3

1	*High Availability Database Copy 1 (Active)*
2	*High Availability Database Copy 2 (Passive)*
3	*High Availability Database Copy 3 (Passive)*
4	*Lagged Database Copy*

Figure 3. Database Availability Group Distribution Across Mailbox Servers



The Microsoft Exchange 2013 messaging system design enables the organization to withstand up to two mailbox server failures without loss of data. For instance, if mailbox server 2 fails, the passive copies (number 2) for each database hosted by mailbox server 2 will activate on mailbox server 1. If mailbox server 1 then also fails, the third database copy hosted at the VMware Cloud Provider facility can be activated by the operations team.

With two mailbox servers at each site hosting active mailboxes (12,500 users per site), the entire population of 25,000 users is divided equally among the four servers (two servers per DAG) resulting in 6,250 users per server during normal operational conditions, with no failed servers. With a single mailbox server failure, a secondary mailbox server is required to handle the workload generated by 12,500 users, and therefore needs to be sized accordingly. This application architecture for Microsoft Exchange 2013 provides highly available database copies and a high level of data protection and redundancy.

6.1.1 Client Access Servers and Edge Transport Servers

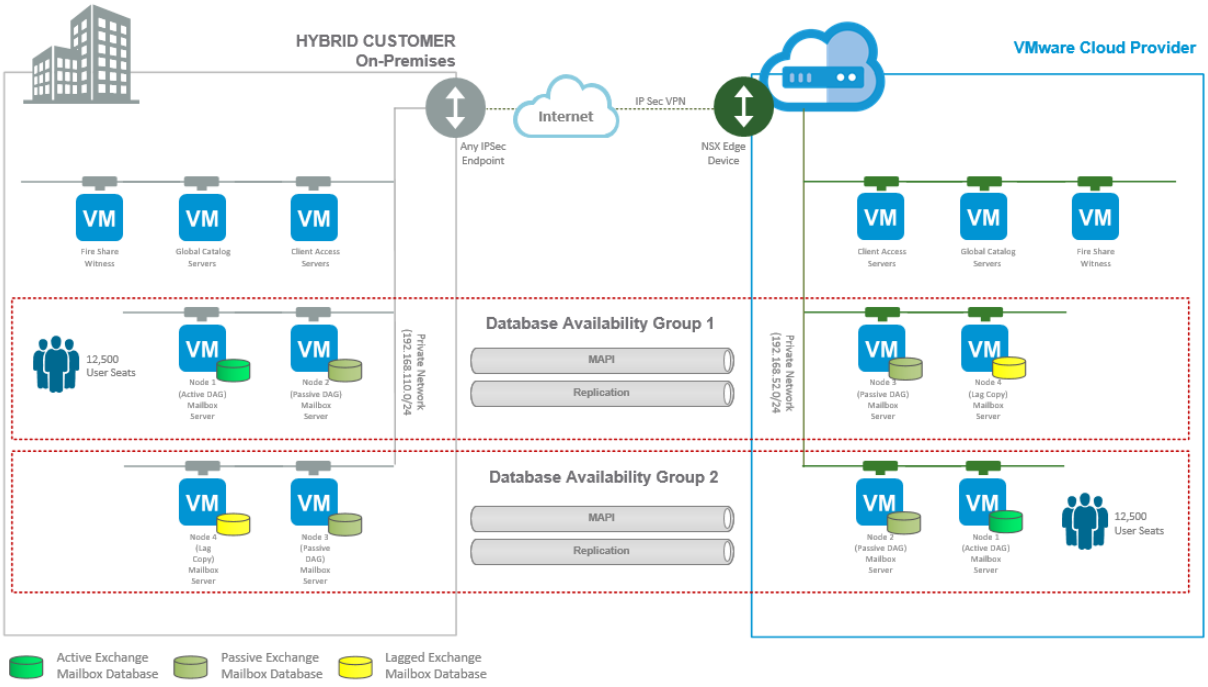
The Client Access Server (CAS) and Edge Transport Server (ETS) roles are deployed in a separate server instance from the mailbox servers.

The CASs are deployed in a 1:1 ratio with mailbox servers. Therefore, because Rainpole.com requires four mailbox servers per DAG for a total of eight mailbox servers, an additional eight servers are installed with the CAS role, distributed evenly across both data centers to provide the required resources to handle the workload generated by 25,000 users with full redundancy. Similarly, to provide the required resources for message hygiene throughput with full site redundancy, three ETSs are deployed on both the on-premises and VMware Cloud Provider Program data center facilities.

6.1.2 Component Model

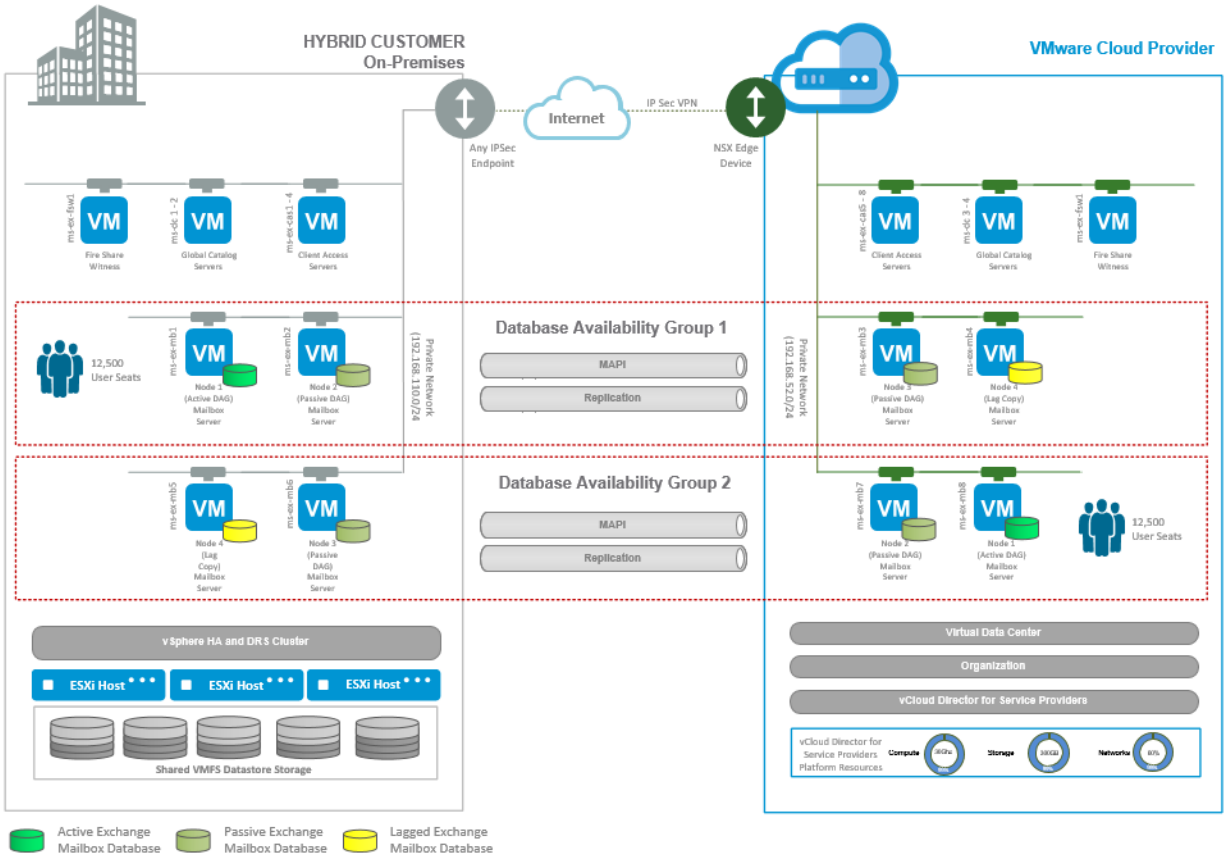
The following figure shows the virtual machine resources required to support all 25,000 users. There are four virtualized mailbox servers per DAG, as well as four corresponding CASs and three ETSs at each data center. In addition, there is one file share witness virtual machine per DAG and Active Directory Domain Controller virtual machines, which are required for the Microsoft Exchange infrastructure and also to provide Global Catalog services.

Figure 4. Microsoft Exchange Architecture



The following figure shows the physical hosts and resources required to support the Microsoft Exchange infrastructure. In this configuration, a number of dedicated physical hosts are employed at the on-premises data center to host the Microsoft Exchange virtual machines, the Active Directory Global Catalog virtual machines, the file share witness, and to provide sufficient resources for vSphere HA admission control reserved capacity. The hardware resources employed at the VMware Cloud Provider facility are provided by a dedicated VMware vCloud Director® Organizational Virtual Data Center (VDC).

Figure 5. vSphere Physical Infrastructure



In the figure, the first DAG has been configured with an active database on ms-ex-mb1 and a local passive database copy on ms-ex-mb2 at the customer's on-premises data center. A secondary passive database on ms-ex-mb5 has been configured at the VMware Cloud Provider facility for high availability and disaster recovery purposes. This architecture provides redundancy in multiple failure scenarios at both the on-premises and provider data centers.

6.1.3 DAG Replication Requirements

All DAG-enabled mailbox server virtual machines require a minimum of two network connections. One is required for MAPI traffic and the other for replication traffic (seeding). To provide timely data replication, it is important that there be less than 50 ms link latency between the on-premises data center and the VMware Cloud Provider's site. The log shipping network compression feature must also be enabled for each DAG to reduce the size of network packets. For more information about enabling log shipping network compression, refer to Microsoft Exchange product documentation.

Designing the Solution

The sample solution architecture shown in Figure 5. vSphere Physical Infrastructure is built on Windows Server 2012 R2 Enterprise Edition with Windows Server Failover Cluster (WSFC). Each DAG consists of two mailbox server nodes in the primary on-premises data center, and two nodes at the VMware Cloud Provider facility, using a stretched, active-active failover cluster configuration. Both of the two four-node Microsoft Exchange 2013 DAGs are configured with an active database and two passive database copies. The following table describes the virtual machine's functions in this sample solution architecture:

Table 2. Architecture Virtual Machine Function

Data Center Site	Virtual Machine	Function
On-Premises	Windows Server 2012 R2 (ms-dc1)	Provides DNS, Active Directory, and Global Catalog services
On-Premises	Windows Server 2012 R2 (ms-dc2)	Provides DNS, Active Directory, and Global Catalog services
On-Premises	Windows Server 2012 R2 (ms-ex-mb1)	Provides mailbox database services (DAG 1)
On-Premises	Windows Server 2012 R2 (ms-ex-mb2)	Provides mailbox database services (DAG 1)
On-Premises	Windows Server 2012 R2 (ms-ex-mb5)	Provides mailbox database services (DAG 2)
On-Premises	Windows Server 2012 R2 (ms-ex-mb6)	Provides mailbox database services (DAG 2)
On-Premises	Windows Server 2012 R2 (ms-ex-fsw)	Provides cluster file share witness services (DAG 1)
On-Premises	Windows Server 2012 R2 (ms-ex-cas1)	Provides client access and front-end transport services
On-Premises	Windows Server 2012 R2 (ms-ex-cas2)	Provides client access and front-end transport services
On-Premises	Windows Server 2012 R2 (ms-ex-cas3)	Provides client access and front-end transport services
On-Premises	Windows Server 2012 R2 (ms-ex-cas4)	Provides client access and front-end transport services
VMware Cloud Provider Facility	Windows Server 2012 R2 (ms-ex-mb3)	Provides mailbox database services (DAG 1)
VMware Cloud Provider Facility	Windows Server 2012 R2 (ms-ex-mb4)	Provides mailbox database services (DAG 1)
VMware Cloud Provider Facility	Windows Server 2012 R2 (ms-ex-mb7)	Provides mailbox database services (DAG 2)

Data Center Site	Virtual Machine	Function
VMware Cloud Provider Facility	Windows Server 2012 R2 (ms-ex-mb8)	Provides mailbox database services (DAG 2)
VMware Cloud Provider Facility	Windows Server 2012 R2 (ms-ex-cas5)	Provides client access and front end transport services
VMware Cloud Provider Facility	Windows Server 2012 R2 (ms-ex-cas6)	Provides client access and front end transport services
VMware Cloud Provider Facility	Windows Server 2012 R2 (ms-ex-cas7)	Provides client access and front end transport services
VMware Cloud Provider Facility	Windows Server 2012 R2 (ms-ex-cas8)	Provides client access and front end transport services
VMware Cloud Provider Facility	nsx-edge-02	VMware NSX Edge™ device

7.1 Distance and Latency Considerations

The link between the customer's on-premises data center and the VMware Cloud Provider facility is a significant factor in this solution architecture. Therefore, when designing a high availability and disaster recovery solution, it is important to understand factors such as link capacity, latency, and use. Typically, network latency corresponds to distance. Round-Trip Time (RTT) can vary from 0 ms to 250 ms or higher. The following table provides distances and their corresponding link latencies. Note that the values shown here are estimates and can vary significantly depending on interconnect quality, network hops, and physical infrastructure.

Table 3. Distance and Estimated Link Latency

Approximate Distance	One-Way Latency (ms)	Round-Trip Latency (ms)
50 km	0.25	0.5
500 km	2.5	5
1,000 km	5	10
5,000 km	25	50
10,000 km	50	100
20,000 km	100	200
22,500 km	125	250

Verify that the design follows Microsoft Exchange Server guidelines for latency and that the required interconnect capacity requirements are known when evaluating connectivity options for the solution.

7.2 Server Sizing

Microsoft recommends hyper-threading be disabled on all production Microsoft Exchange servers. As with any virtualized business critical application, when sizing for a new virtualized production deployment, you do not want to oversubscribe processors. Maintain a 1:1 ratio of logical cores to virtual processors on the host. In addition:

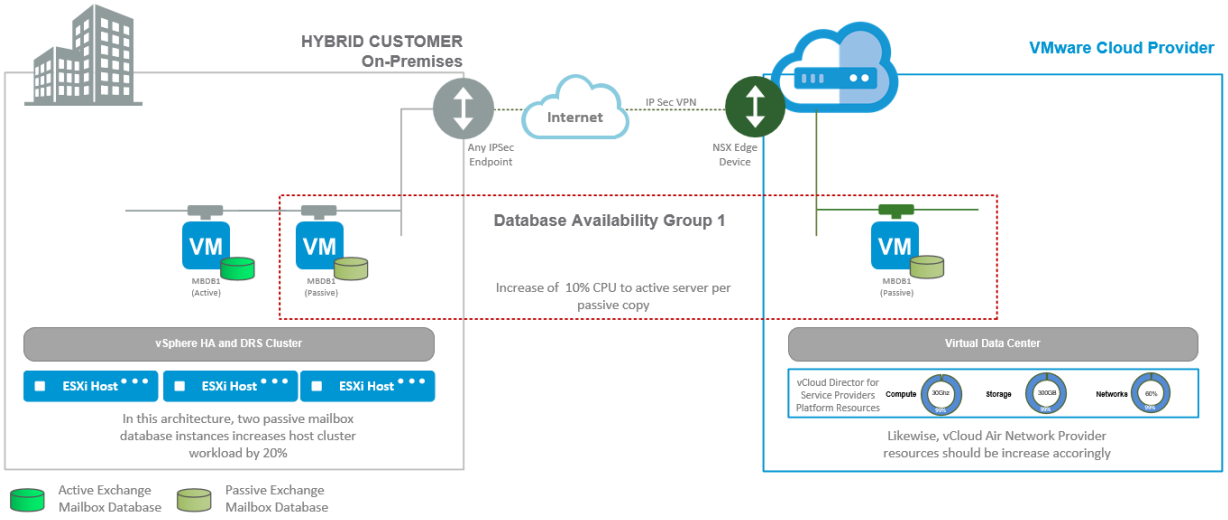
- Format the guest operating system database and log volumes at a 64Kb allocation unit size, as recommended by Microsoft.
- Isolate Microsoft Exchange database and log files from other disk intensive application workloads to avoid performance conflicts. Sharing the storage with other applications might negatively impact Microsoft Exchange I/O performance.
- Isolate Microsoft Exchange database and log disk I/O on separate physical disk arrays. Although Microsoft Exchange Server 2013 DAG configuration allows logs and databases on the same LUN, this isolation enables separate log and database disk tuning with backend RAID levels.
- For most Microsoft Exchange Server 2013 environments, Microsoft recommends hardware level RAID 5 for databases and RAID 10 for the logs. RAID 10 can be used for both the transaction logs and database LUNS, as the performance and resilience that it offers is suitable. However, a RAID 10 array makes it quite an expensive proposition for large amounts of data. Therefore, RAID 5 for databases offers a better space verses performance balance when providing users with larger mailboxes (1 GB plus), provided there are adequate disks for the I/O requirements.

7.2.1 DAG Process Requirements

The DAG feature in Exchange 2013 requires additional considerations when sizing the mailbox server role for a customer design, forcing the architect to account for both active and passive mailboxes. As we have seen in this design, when a mailbox server is a member of a DAG, that virtual server can also host one or more passive databases in addition to any active databases for which they might be responsible. When calculating compute resources, note that each passive database is likely to add an additional 10 percent to the CPU requirements of the mailbox server already hosting an active copy of a mailbox database.

The following figure illustrates this principle. In the figure, there are three Microsoft Exchange mailbox servers, each with an active database (MSDB1a denotes database 1 active) and two passive databases from the other two mailbox servers (MSDB1p denotes database 1 passive). Each passive copy of MSDB1a requires 10 percent extra processing on the server hosting MSDB1a, for a total of 20 percent extra CPU overhead.

Figure 6. DAG Layout with Overhead for Passive Databases



Therefore, each mailbox server in the sample architecture requires an additional 20 percent of processing power to account for passive database copies. The sizing process begins with understanding and applying Microsoft guidelines for each server role, as represented by the following high-level processes:

- Design the mailbox server DAG nodes:
  - Apply Microsoft guidelines to determine CPU and memory requirements. Considerations include number of mailboxes, mailbox profile, number of servers in the DAG, number of passive database copies, and several other custom parameters.
  - Use the *Exchange 2013 Mailbox Server Role Requirements Calculator* (https://gallery.technet.microsoft.com/Exchange-2013-Server-Role-f8a61780) from Microsoft to determine storage requirements. (Third-party Web sites are not under the control of VMware, and the content available at those sites might change.)
- Design the peripheral server roles:
  - The *Exchange 2013 Mailbox Server Role Requirements Calculator* also recommends CPU and memory for the CAS and ETS roles.
- Allocate one or more virtual machines for each server role to satisfy the previously calculated number of processor cores and amount of memory.
- Determine how the virtual machines will be distributed across ESXi hosts on site and the resources required for the provider environment.
- Aggregate virtual machine requirements plus some overhead to size each ESXi host. This overhead is important if you want to minimize the performance impact during the loss of an ESXi host. A typical guideline when choosing the number of required hosts is n+1, where n is the number of hosts required to run the workload at peak use. n+1 allows you to design for the possibility of losing one host from your VMware cluster without taking a performance hit during failover.

7.2.2 Sample Environment Sizing

By using the *Microsoft Mailbox Server Role Calculator,* Rainpole.com's IT organization has determined that the following virtual machine requirements will be sufficient to meet the needs of the customer's environment.

Table 4. Environment Sizing

Virtual Machine Role	vCPU	Memory
Exchange Mailbox VM	8 vCPU	48 GB RAM
Exchange Client Access VM	4 vCPU	16 GB RAM
Exchange Edge Transport VM	2 vCPU	8 GB RAM

Now that physical resource requirements of the customer's environment and associated virtual hardware configuration are understood, physical resources can be planned to meet those requirements both on- and off-premises. To build infrastructure availability into the architecture, virtual machines will be distributed across multiple hosts on-premises and rely on the provider's VDC resources to provide the same level of availability at the VMware Cloud Provider data center facility.

Architecting a Robust Technical Platform

When deploying Microsoft Exchange 2013 DAGs on a vSphere platform, consider the guidelines in this section to increase the solution availability and performance by architecting a robust technical platform.

8.1 Virtual Machine Placement

When configured, vSphere HA and DRS can automatically perform virtual machine placement for initial power on or when the cluster is under resource contention. When running DAGs across virtual machines in a vSphere HA or DRS cluster, to avoid a single point of failure, WSFC-configured virtual machines must be kept apart, on different physical hosts. The following practices can help achieve this requirement:

- Create a DRS anti-affinity rule to keep virtual machines on different hosts.
- Enable strict enforcement of affinity rules.
- Set the DRS automation level for WSFC virtual machines to partially automatic.

In addition, when the `ForceAffinePoweron` option for vSphere DRS is set to 1, strict enforcement of anti-affinity rules is enforced. When possible, VMware also recommends using multiple server racks for the ESXi hosts, and distributing the vSphere clusters across the server cabinets to minimize the impact of a single component failure.

Refer to *Setup for Failover Clustering and Microsoft Cluster Service* (https://pubs.vmware.com/vsphere-60/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-60-setup-mscs.pdf) for detailed setup procedures. To set up anti-affinity rules to keep virtual machines on separate hosts, follow the instructions at http://kb.vmware.com/kb/2003128.

8.2 Disk Provisioning

Virtual machine disk files (VMDKs) can be deployed in three different formats—thin, thick, or eagerzeroedthick. Both thin and thick disk files use lazy zeroing, where the initial zeroing of the disk blocks is delayed until the first write. Eagerzeroedthick disk blocks are pre-allocated with zeros at the time of disk provisioning, making it unnecessary to zero the disk on a first write basis during normal running operations. This provides approximately a 10–20 percent performance improvement over the other disk formats.

Most Microsoft Exchange 2013 high availability features are highly sensitive to system response time. The additional overhead of disk zeroing during normal operations might cause unnecessary disk latency and potentially cause a false cluster failover event. If you are deploying DAGs on a vSphere platform, VMware highly recommends defining an operational standard to only use eagerzeroedthick disks for Microsoft Exchange mailbox databases and log files.

8.3 Network Considerations

The network is a critical component required for cluster node communication both locally and across sites. In addition to normal network communications, Microsoft Exchange Server in a non-shared disk, high availability solution, such as DAGs, also uses the network for data replication between replicas. With cross data center interconnect replication being key to a successful high availability and disaster recovery solution, the performance of the design is highly dependent on network bandwidth and latency.

Consider the following network configuration guidelines to achieve a robust technical platform with optimal performance:

- Configure appropriate network adapters on the host and virtual machines to separate networks used for different vSphere, virtual machine, and application purposes. For instance, a separate network for data replication, vSphere vMotion, management VMkernel, and so on.
- When using iSCSI at the host level or in the guest's operating system, the network adapters must be dedicated.
- Use the VMXNET3 paravirtualized NIC with all Windows Server 2012 R2 instances. VMXNET3 is optimized for virtual environments and designed to provide the best performance.
- Enable jumbo frames for the iSCSI and/or vSphere vMotion network.
- Always employ static IP addresses for network interfaces in a Windows Server cluster. Using dynamic configuration through DHCP is not recommended because the failure to renew a DHCP lease might disrupt cluster operations.
- When deploying DAGs across a data center interconnect, a high-speed network must be used for replication traffic. Confirm that the bandwidth and latency of the network is sufficient to support the amount of Microsoft Exchange Server replication traffic

8.4 Other Considerations

Typically, Microsoft Exchange aggressively uses all of the provided memory available in a guest operating system. While vSphere can support memory over commitment, this must be considered with caution to avoid performance impact due to resource contention. For virtualized business critical applications, such as production Microsoft Exchange systems, VMware recommends setting a memory reservation on the amount of memory configured for the virtual machine. Also consider:

- Using the latest processor generations for their enhanced virtualization features and support.
- VMware recommends at least two HBA ports per ESXi host for storage redundancy.

The exact number of NIC/HBA ports might vary on each ESXi host. More ports might be required for a successful design depending on the number of virtual machines and customer-specific network and storage requirements. The number and exact design must be determined by a detailed sizing analysis with the VMware Cloud Provider solution architect.

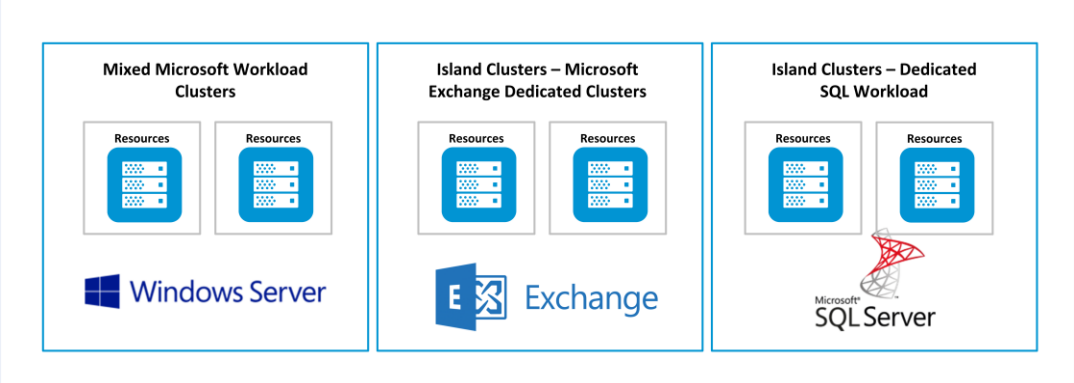
8.5 Dedicated "Island" Application Clusters

Running enterprise applications like Microsoft Exchange 2013 on vSphere might require a different operational approach. For instance, a dedicated vSphere "island" cluster can be configured with a different configuration than typical vSphere configurations, such as including rules to avoid over-commitment, limiting vSphere HA / DRS / vSphere vMotion, or dedicating storage resources to performance-intensive mailbox virtual machines.

The concept of island clusters is fairly simple. An island cluster (also referred to as a dedicated application cluster) hosts workloads with special license, performance, availability, or other configuration requirements. Some software vendors apply licensing policies on their applications, middleware, and databases that are not conducive to virtualization, and especially DRS, where the application can potentially touch a high number of physical CPUs. Island clusters are one approach to dealing with this challenge.

Some service providers also use island clusters of operating systems such as Windows or RHEL. This helps them save money on data center Windows socket licenses, which are typically the most cost effective way of licensing large numbers of Windows VMs running on a host. Another benefit of this approach is that it helps ESXi take advantage of the memory management technique of Transparent Page Sharing (TPS), which is now disabled by default by VMware. While this can be more efficient due to the higher chances of duplicate pages being spawned by VMs in physical memory when multiple instances of the same guest OS virtual machines are running, this approach is not recommended for island clusters supporting virtualized business critical applications. Some use cases for island clusters are represented in the following figure.

Figure 7. Dedicated Island Application Clusters



8.6 Client Connectivity

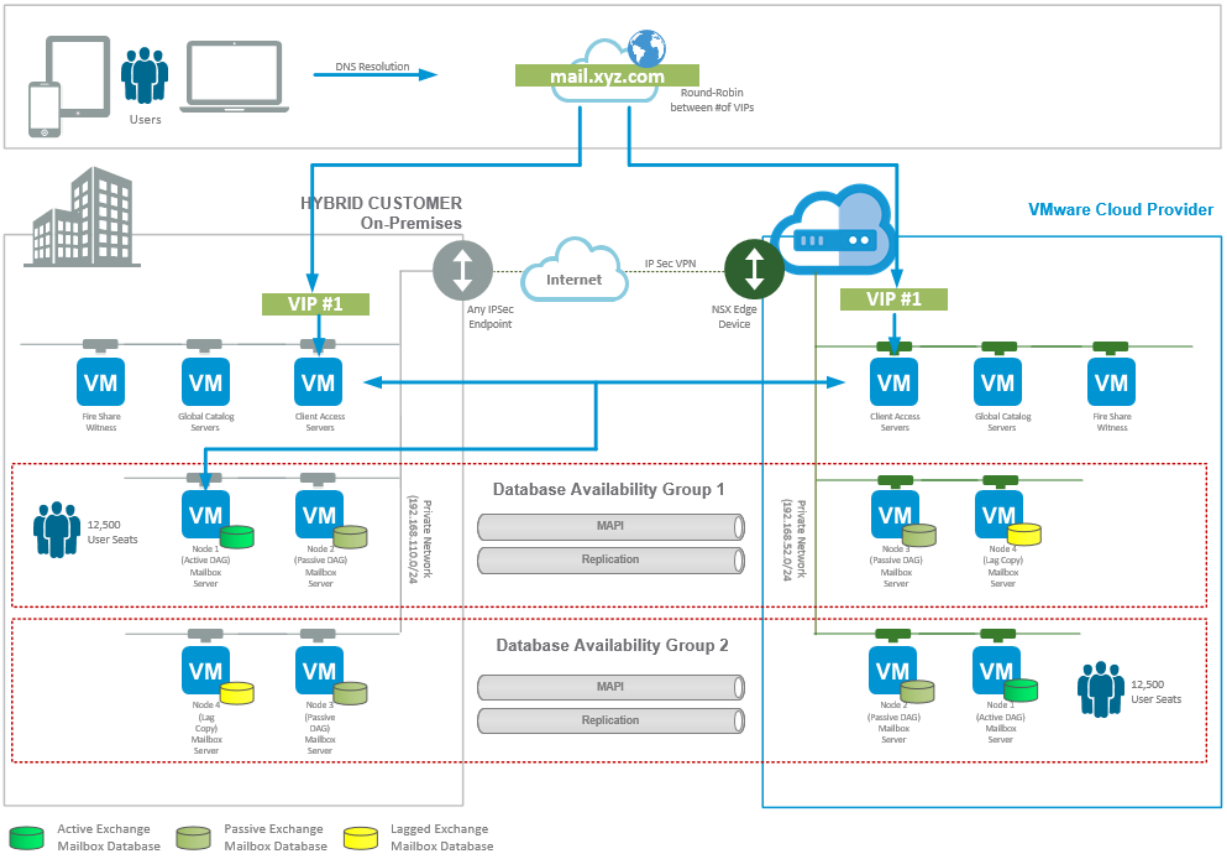
With a dual site architecture with a VMware Cloud Provider, another benefit of Microsoft Exchange 2013 is that the namespace model can be simplified.

This sample solution architecture uses the Client Access Server to proxy requests to the mailbox server hosting the active database copy. This allows the CAS in one site, for instance the VMware Cloud Provider Program data center, to proxy a session to a mailbox server that is located in the on-premises data center. Assuming that network use, latency, and throughput are not a design concern, this allows us to employ a single namespace across both sites for redundancy, which also simplifies the namespace architecture.

For instance, in our sample design scenario where two data centers with low latency, throughput, and use between the on-premises and off-premises facilities exist, the design can simplify the namespace so that users only have to use a single namespace name for Internet access, regardless of where their mailbox is located. With the architecture shown in the following figure, the CASs in both data centers can be used to route and proxy traffic to the mailbox server hosting the user's active database copy.

Because network traffic is not a concern in this architecture, for an optimal solution, DNS is configured to round-robin between the VIPs of the load balancers located in each data center. The result is that there is a site-resilient namespace design as long as you can accept that 50 percent of the client connections will be proxied from the alternate site.

Figure 8. Microsoft Exchange 2013 Single Namespace Architecture



Operational Model

This section describes the operational aspects of the sample solution architecture in a technology- and product-focused manner.

VMware vRealize® Operations Manager™ provides a unified view and deep insights into the health, risk, and efficiency of the infrastructure and applications to help provide quality of service and early detection of performance, capacity, and configuration issues. For the solution architecture presented in this paper, visibility into the Microsoft Exchange 2013 business critical application is possible with the "out-of-the-box" adapter for Microsoft Exchange. With endpoint agents, vRealize Operations Manager can provide a performance management solution that can extract counters from the guest operating systems, applications, and mailbox databases.

vRealize Operations Manager ships as a single virtual appliance (with multiple HA options) that can be quickly installed and used to improve the performance and health of your vSphere infrastructure. vRealize Operations Manager functionality includes:

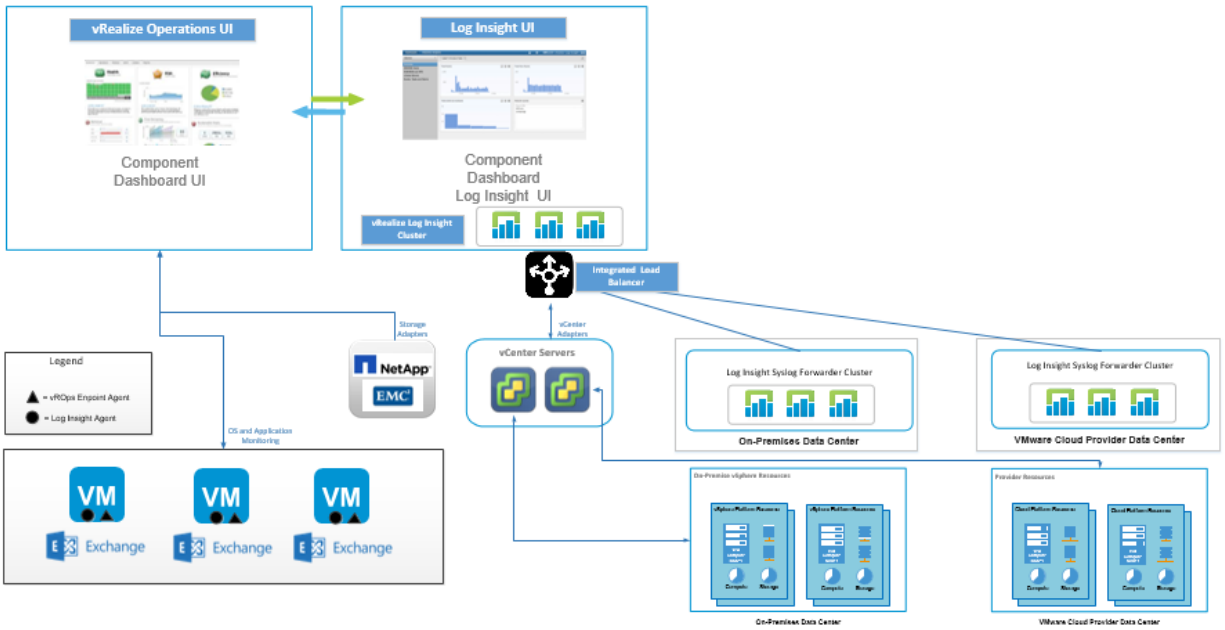
- Operations dashboard – Provides at-a-glance views into the health, risk, and efficiency of the virtual infrastructure.
- Health and workload views – Helps identify anomalies, faults, and stressed workloads that can impact the performance and health of the infrastructure.
- Workload details view – Provides in-depth analysis on what is impacting the performance and health of virtual machines, hosts, data stores, and vSphere clusters.
- Visibility into other application products through adapters.
- Uses adapters to collect data from a variety of data sources, including third-party products such as Microsoft Exchange. Adapters work with the vRealize Operations Manager collector to collect and process data. The collector acts as a gateway between vRealize Operations Manager and the adapters. The adapters connect to and collect data from data sources, transform the data into a format that vRealize Operations Manager can consume, and pass the data to the collector for final processing. Depending on the data source and adapter implementation, an adapter might collect data by making API calls, using a command-line interface, or sending database queries. The VMware vCenter® adapter collects metrics and events from VMware vCenter Server®. This is part of the base vRealize Operations Manager installation.

The following two options provide this design with metrics to formulate custom Microsoft Exchange dashboards in vRealize Operations Manager:

- End Point Operations Management Solutions – Directly collects data from the operating system.
- The Microsoft SCOM adapter – Collects resource metrics such as time series, resource availability, and resource relationship data from Microsoft System Center Operations Manager (SCOM) databases.

The architecture of a vRealize Operations Manager solution for this specific design depends on a number of design factors, and specifically on the demarcation line and access to the cloud data center resources provided to the customer by the VMware Cloud Provider. The following figure shows one high-level design option for the integration of vRealize Operations Manager in this sample solution architecture.

Figure 9. Sample Operational Management Design



As shown, VMware vRealize Log Insight™ provides a monitoring and operational role in this solution. In addition to infrastructure monitoring, the design can also forward logs from Windows Servers or Windows-based applications such as Active Directory or Microsoft Exchange. The vRealize Log Insight Windows agent must be installed on each source operating system, allowing messages from Windows event channels and log files to be forwarded to the vRealize Log Insight server. vRealize Log Insight can be configured to provide a comprehensive monitoring of operating systems and applications, allowing application owners and operating systems operational teams to query, analyze, and audit log data.

Conclusion

This document provides a validated business continuity solution, at a high level, for Microsoft Exchange Server 2013 employing the high availability and disaster recovery capabilities of DAGs across a distributed architecture. The overall aim of this solution architecture paper is to design a highly available and disaster recoverable solution for a virtualized Microsoft Exchange business critical environment spanning on-premises and VMware Cloud Provider data centers.

DAGs are the ideal solution for deploying a highly available Microsoft Exchange Server 2013 environment on a vSphere platform. DAGs provide out-of-the-box protection for hardware, software, and from data failure, as well as additional capabilities to reduce the requirement for backups. With its non-shared disk architecture, DAGs can be used safely with vSphere vMotion, DRS, and vSphere HA to reduce downtime and improve flexibility in the Microsoft Exchange architecture while lowering costs and minimizing the need to perform a Microsoft Exchange DAG failover. This solution can also reduce Microsoft Exchange mailbox server recovery time from hours or days to minutes or even seconds.

Consider the following key architectural points when designing a solution:

- DAGs and database mirroring can be used in combination with vSphere vMotion, vSphere HA, and DRS to maximize Exchange Server availability.
- ESXi 5.1, 5.5, and 6.0 support up to five-node clusters for Windows Server 2008 SP2 and later, but earlier ESXi versions support only two-node clusters.
- To avoid single points of failure, use vSphere anti-affinity rules to run AlwaysOn Availability Group replica virtual machines on separate hosts.
- Microsoft Exchange DAGs on vSphere are supported for non-shared disk configurations, except when the system disk's VMDK is located on a NFS datastore.

As shown in this solution architecture paper, with the combined power of the VMware Cloud Provider Program and virtualized Microsoft Exchange Server 2013, an IT organization can demonstrate business-critical levels of performance and availability of mailbox databases, while accelerating deployment time and simplifying IT operations. For more information, see Section 12, Reference Documents.

Assumptions and Caveats

VMware, Microsoft, and other third-party hardware and software information provided in this document is based on the current performance estimates and feature capabilities of the indicated versions. This information is subject to change by their respective vendors. Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Reference Documents

Table 5. Reference Documents

Document	Link or URL
Microsoft Clustering on VMware vSphere: Guidelines for supported configurations	http://kb.vmware.com/kb/1037959
Exchange 2013 Server Role Requirements Calculator	https://gallery.technet.microsoft.com/Exchange-2013-Server-Role-f8a61780
Exchange Server 2013 – Technet	https://technet.microsoft.com/en-us/library/bb124558(v=exchg.150).aspx
Setup for Failover Clustering and Microsoft Cluster Service	https://pubs.vmware.com/vsphere-60/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-60-setup-mscs.pdf