Deploying Microsoft Exchange Server 2010 with Hyper-V on the Hitachi Virtual Storage Platform

Reference Architecture Guide

By Jeff Chen and Leo Nguyen

October 2010

Feedback

Hitachi Data Systems welcomes your feedback. Please share your thoughts by sending an email message to SolutionLab@hds.com. Be sure to include the title of this white paper in your email message.

Table of Contents

Deploying Microsoft Exchange Server 2010 with Hyper-V on the Hitachi Virtual Storage Platform

Reference Architecture Guide

IT administrators need email solutions that provide data protection and simple management in environmentally friendly data centers. Using Microsoft Exchange Server 2010 with SAN storage like the Hitachi Virtual Storage Platform accomplishes those business-critical objectives. The reference architecture described in this white paper leverages changes in Exchange Server 2010 and the advantages of the Virtual Storage Platform to lower administration costs, improve system efficiency and enable virtualization.

The Hitachi Virtual Storage Platform can help you leverage your information, which is the new currency in today's data-driven economy. Information exists in many forms and must be protected and readily accessible to ensure business survival and success. The Virtual Storage Platform maximizes cost efficiency and return on investment by creating an agile storage infrastructure that reduces costs and increases performance, availability, scalability and reliability.

The Hitachi Virtual Storage Platform is the industry's only 3D scaling storage platform. With the unique ability to concurrently scale up, scale out and scale deep in a single storage system, the Virtual Storage Platform flexibly adapts for performance, capacity, connectivity and virtualization. No other enterprise storage platform can dynamically scale in three dimensions. Scaling up allows you to increase virtual server consolidation, improve utilization of resources, and reduce costs. Scaling out is required when you add new physical or virtual servers to your environment to meet business demands. Scaling deep extends the advanced functions of the Virtual Storage Platform to external multivendor storage.

Exchange Server 2010 can be deployed for many types and numbers of users in a wide variety of infrastructure topologies. By implementing Exchange Server 2010 with the Virtual Storage Platform storage systems, you can effectively scale an environment from a few thousand users to hundreds of thousands of users. Hitachi Dynamic Provisioning software also allows you to combine different workloads in a single storage frame for greater flexibility.

Because Exchange is often one of the most mission-critical application in an organization's IT environment, and because every IT organization is constantly trying to reduce costs and management complexity while increasing flexibility and reliability, deploying Exchange on a virtual platform like Hyper-V is a logical choice. You can realize additional benefits deploying this virtual environment on a Hitachi Virtual Storage Platform system using Hitachi Dynamic Provisioning.

This reference architecture is for a deployment of 20,000 users on a Virtual Storage Platform storage system with the Exchange 2010 servers virtualized using Hyper-V. The Exchange databases are protected by using a Database Availability Group (DAG).

This reference architecture focuses on planning and deploying the Exchange Server 2010 mailbox role and is intended for use by IT administrators responsible for Exchange and storage. It assumes some familiarity with Hitachi Storage Navigator software, Microsoft Windows 2008 R2 and Exchange Server 2010. For information about deploying other Exchange server roles, see the Microsoft TechNet library "Exchange Server 2010."

Solution Overview

Deploying all Exchange components including Active Directory, mailbox, hub transport and client access servers in a virtualized environment offers the greatest overall benefit. It increases agility and availability of Exchange while lowering costs. Fewer physical servers can house multiple Exchange server roles with better resource utilization.

The reference architecture described in this paper uses three physical servers with Windows 2008 R2 Datacenter edition installed and Hyper-V enabled to house the following virtual machines.

- Four Exchange 2010 mailbox servers; each server supports three active and three passive mailbox databases
- Four Exchange 2010 servers with the client access and hub transport roles installed
- One Windows 2008 domain controller

Figure 1 shows the layout of the virtual machines on the three physical hosts and the connections to the Hitachi Virtual Storage Platform.

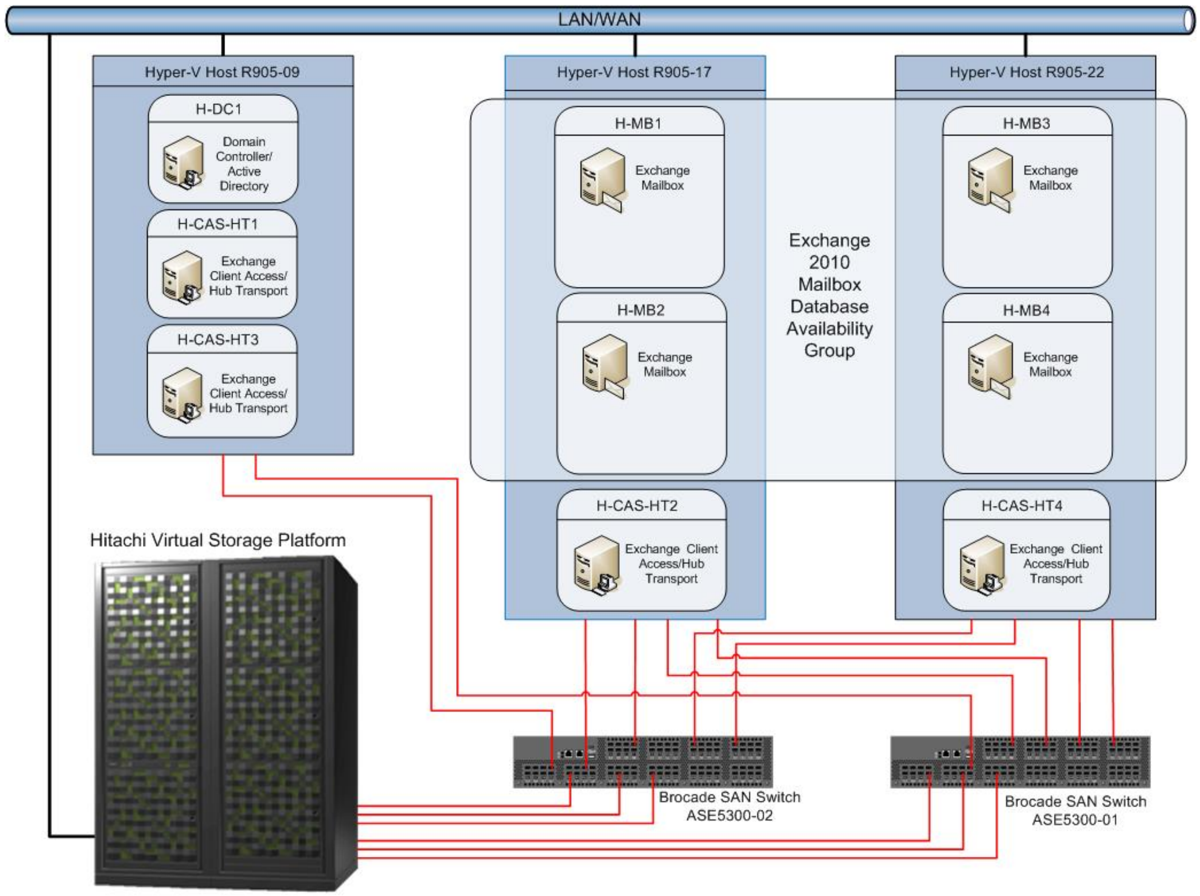

Figure 1

This reference architecture describes an Exchange 2010 environment designed for 20,000 users with 1GB mailboxes. Based on Microsoft recommendations for mailbox database sizing in DAG environments, each mailbox database contained the mailboxes for 1,667 users. This is based on a 2TB maximum database size and allowing a 20 percent overhead as recommended by Microsoft:

(1,667 x 0.001TB) x 1.2 = 2TB

Based on this calculation, Hitachi Data Systems used 12 databases to house 20,000 users. You can use the 1.2 multiplier to get an estimate of the space required, but for a production environment you use the E2010 Mailbox Server Role Requirements Calculator to determine accurate requirements for your environment. For more information about the calculator, see the "Exchange 2010 Mailbox Server Role Requirements Calculator" entry on the Microsoft Exchange Team Blog.

Key Solution Components

The following sections describe the key hardware and software components used to deploy this solution.

Hitachi Virtual Storage Platform

The Hitachi Virtual Storage Platform is the industry's only 3D scaling storage platform. With the unique ability to concurrently scale up, scale out and scale deep in a single storage system, the new Virtual Storage Platform flexibly adapts for performance, capacity, connectivity and virtualization. No other enterprise storage platform can dynamically scale in three dimensions. The Virtual Storage Platform provides virtual storage that meets the growing demands of server virtualization.

The trend in server virtualization is to consolidate the I/O workload of many servers onto a single storage system. As more virtual machines are consolidated onto a physical host, storage systems must be able to dynamically add more storage resources to keep up with I/O demand. The 3D scaling capability of the Virtual Storage Platform meets that requirement.

Scaling up allows you to increase virtual server consolidation, improve utilization of resources, and reduce costs. With the Hitachi Virtual Storage Platform, you can increase performance, capacity and connectivity by adding cache, processors, connections and disks to the base system. A virtual server that accesses the storage system can use all these resources, which act as one system managed as a common pool of resources.

Scaling out allows you to meet increasing demands by combining multiple chassis into a single logical system with shared resources. By scaling out you can support increased resource needs in virtualized server environments.

Scaling deep extends the advanced functions of the Virtual Storage Platform to external multivendor storage. By dynamically virtualizing new and existing storage systems, those systems become part of the Virtual Storage Platform's pool of storage resources. Once virtualized, external data can then be migrated, tiered, replicated and managed by the Virtual Storage Platform. In this manner, older data storage systems can gain a longer useful life. You can extend distance replication for business continuity to lower-cost, lower-function storage systems by virtualizing them behind a Virtual Storage Platform.

The switch matrix architecture of the Virtual Storage Platform makes all of this possible. It connects the basic components, front-end directors, back-end directors, global cache modules and virtual storage directors. You can add redundant pairs of directors and cache modules as required without disruption to connected host servers. All these resources are tightly coupled through a global cache that creates a common pool of storage resources. These resources can include external storage that is connected through front-end director initiator ports.

Virtual Storage Platform Architecture

The Virtual Storage Platform offers an entirely new level of scalable enterprise storage, capable of handling the most demanding workloads while maintaining great flexibility. The Virtual Storage Platform offers much higher performance, higher performance scalability, higher reliability and greater flexibility than any storage system on the market today.

The Virtual Storage Platform offers these features:

- The HiStar-E PCI Express Switched Grid acts as the interconnection among front-end directors, back-end directors, data cache adapter boards and virtual storage director boards.
- Data accelerator processors on the front-end directors and back-end directors work with central processor boards called virtual storage directors that manage all I/O by sets of assigned logical devices (LDEVs).
- Dual SAS controllers on back-end director boards contain eight 6Gbps SAS links per board.
- The control memory function resides in global cache and each VSD board contains a local copy with information for its LDEVs. Most control memory accesses are lookups to the local copy.
- Global cache is backed up to solid state drives (SSDs) on the cache boards.
- Each virtual storage director board controls all I/O operations for a discrete group of LDEVs. LDEVs are assigned round-robin across the installed virtual storage director boards as they are created. If necessary, you can manually reassign LDEV ownership to a different virtual storage director.
- Each virtual storage director board executes the code for initiator mode (hosts), external mode (virtualization), back-end director mode, or the copy products send and receive modes. Code execution is done on a per-job basis.
- A Virtual Storage Platform can be scaled from a single-chassis system to a dual-chassis system. Each chassis has a control rack and a logic box.
- Up to 1,280 3.5-inch large form factor (LFF) drives or 2,048 2.5-inch small form factor (SFF) drives can be installed in a dual-chassis system. If you install both LFF and SFF disk containers and drives in a storage system, the limits change based on the configuration you choose.

The Virtual Storage Platform is built as a single-chassis or dual-chassis storage system. Each chassis has one control rack and up to two disk expansion racks. The control rack has the logic box that holds all of the control boards for a chassis and one or two disk containers. The disk expansion racks can hold three disk containers each. Disk containers come in two types: small form factor (up to 128 2.5-inch drives) and large form factor (up to 80 3.5-inch drives). When using two chassis as a single integrated storage system, the two units are cross connected at the grid switch level. The storage system behaves as a single unit, not as a pair of units operating as a cluster.

Figure 2 shows the two types of racks available for a Virtual Storage Platform.

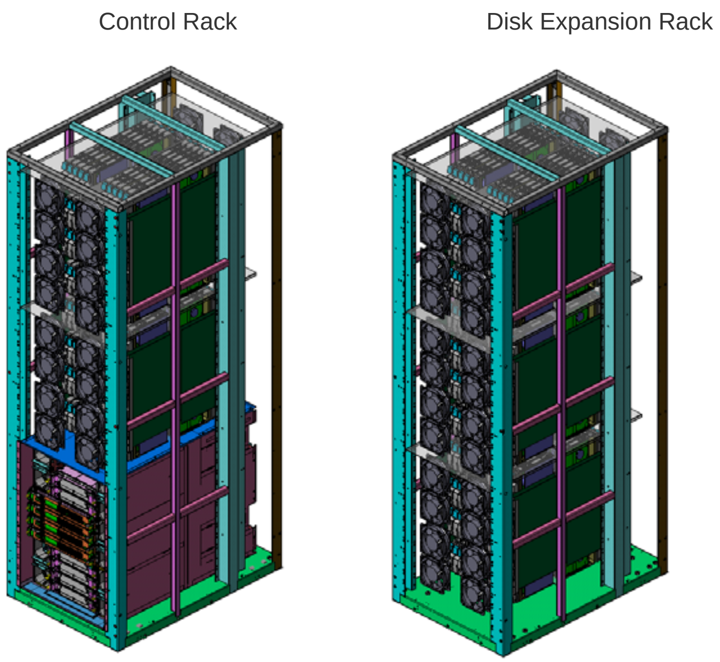

Figure 2

Figure 3 shows the logic boards in a fully populated single-chassis Virtual Storage Platform.

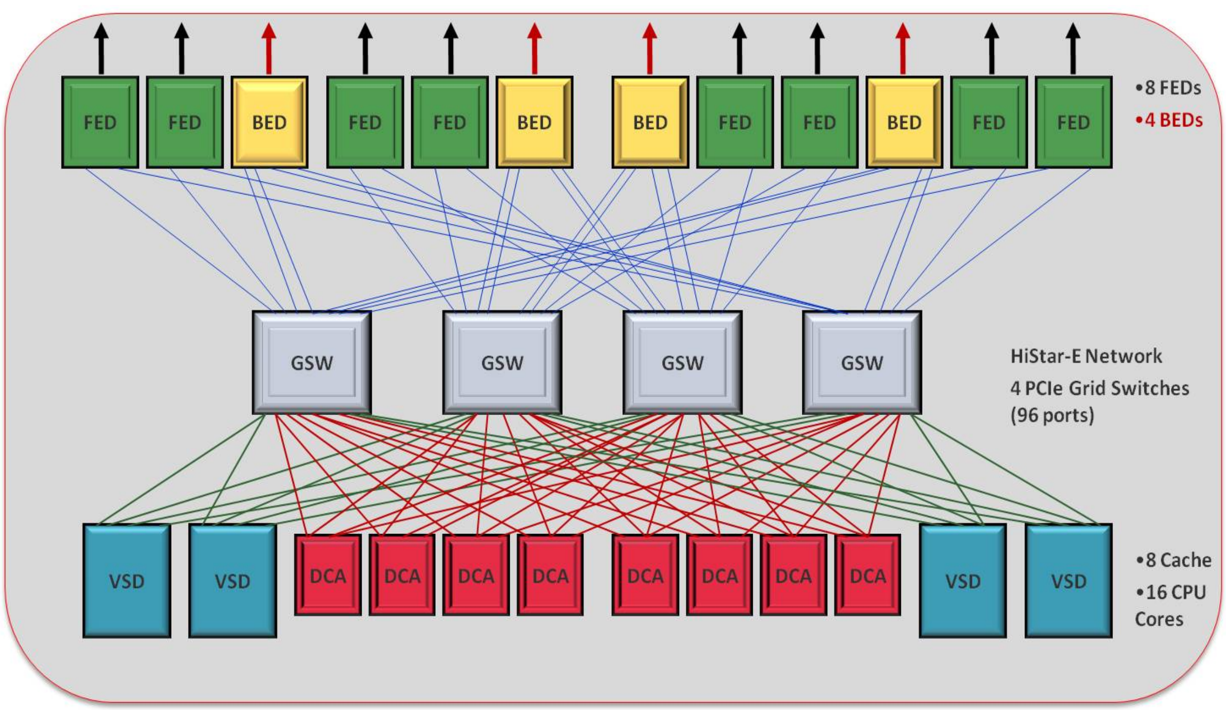

Figure 3

Figure 3 shows the following chassis components (note that a feature is a pair of boards on two separate power domains):

- **GSW** — Grid Switch PCI Express Switch. One or two features (two or four boards) per control unit with 24 2GB/sec HiStar-E ports each can be installed
- **DCA** — Data cache adapter cache memory. One to four features (two, four, six or eight boards) per control unit with up to 32GB of RAM each can be installed
- **VSD** — Virtual storage director processor module. One or two features (two or four boards) per control unit can be installed.
- **FED** — Front-end director host port module. One to four features (two, four, six or eight boards) per control unit of four or eight 8Gbps Fibre Channel ports can be installed.
- **BED** — Back-end director disk controller module. One or two features (two or four boards) per control unit with eight 6Gbps SAS links per board can be installed. If the back-end director options are not installed (available for the single-chassis configuration only), two additional front-end director options can be used in those chassis slots.

3D Scaling Architecture

The Hitachi Virtual Storage Platform allows for optimal infrastructure growth in all dimensions by scaling up, scaling out and scaling deep.

Scale Up

Scale up to meet increasing demands by dynamically adding processors, connectivity and capacity in a single unit, providing the highest performance for both open and mainframe environments.

In the basic single chassis configuration, the number of logic boards, disk containers, and drives is highly scalable. You can start with the minimum set of logic boards,10, and one disk container, then add more boards (up to a total of 28 boards in a single chassis) and disk containers (up to a total of eight disk containers in a single chassis). Disk container types may be intermixed within a chassis.

Scale Out

Scale out to meet multiple demands by dynamically combining multiple units into a single logical system with shared resources, support increased demand in virtualized server environments, and ensure safe multitenancy — that is, the ability to run multiple servers simultaneously without the risk of corruption or modification of data from one server to another — and quality of service through partitioning of cache and ports.

You can double the scalability of the Virtual Storage Platform with a dual-chassis system with up to six racks. The logic box in each chassis is the same, using the same types and numbers of logic boards. Any front-end port can access any back-end RAID group; no division within the storage system exists between the chassis.

A dual-chassis Virtual Storage Platform can manage up to 247PB of total storage capacity.

Table 1 lists the capacity differences between a single-chassis and a dual-chassis Virtual Storage Platform storage system.

Table 1. Virtual Storage Platform Chassis Capacity Comparison

Maximum Capacity	*Single Chassis*	*Dual Chassis*
Data cache	256GB	512GB
Raw cache bandwidth	64GB/s	128GB/s
Solid state drives	128	256
2.5” SFF drives	1,024	2,048
3.5” LFF drives	640	1,280
Logical volumes (LDEVs)	65,280	130,560

Scale Deep

Scale deep to extend storage value by dynamically virtualizing new, existing external storage systems, extend Hitachi Virtual Storage Platform advanced functions to multivendor storage and offload less demanding data to external tiers to optimize the availability of your tier one resources.

The Virtual Storage Platform provides the virtualization mechanisms that allow other storage systems to be attached to some of its front-end director Fibre Channel ports and accessed and managed via hosts that are attached to the host ports on the Virtual Storage Platform. As far as any host is concerned, all virtualized logical units passed through the Virtual storage Platform to the hosts appear to be internal logical units from the Virtual Storage Platform. The front-end ports on the Virtual Storage Platform that attach to the external storage system’s front-end ports are operated in external or SCSI initiator mode (attached to servers), rather than the usual SCSI target mode (attached to hosts).

For more information about the Hitachi Virtual Storage Platform, see the Hitachi Data Systems web site.

Hitachi Dynamic Provisioning Software

On the Virtual Storage Platform, Hitachi Dynamic Provisioning software provides wide striping and thin provisioning functionalities. In the most basic sense, Hitachi Dynamic Provisioning software is similar to the use of a host-based logical volume manager (LVM), but with several additional features available within the Hitachi Virtual Storage Platform and without the need to install software on the host or incur host processing overhead. Hitachi Dynamic Provisioning software provides for one or more pools of wide striping across many RAID groups within a Virtual Storage Platform. One or more Dynamic Provisioning virtual volumes (DP-VOLs) of a user-specified logical size of up to 60TB (with no initial physical space allocated) are created against each pool.

Primarily, you deploy Hitachi Dynamic Provisioning software to avoid the routine issue of hot spots that occur on logical devices (LDEVs) from individual RAID groups when the host workload exceeds the IOPS or throughput capacity of that RAID group. By using many RAID groups as members of a striped Dynamic Provisioning pool underneath the virtual or logical volumes seen by the hosts, a host workload is distributed across many RAID groups, which provides a smoothing effect that dramatically reduces hot spots. By reducing hot spots in an Exchange environment, you can reduce the need to move mailboxes as the user load varies.

Hitachi Dynamic Provisioning software also carries the side benefit of thin provisioning, where physical space is only assigned from the pool to the DP-VOL as needed using 42MB pages, up to the logical size specified for each DP-VOL. A pool can also be dynamically expanded by adding more capacity or reduced by withdrawing pool capacity. Either operation is performed without disruption or requiring downtime. Upon expansion, a pool can be rebalanced so that the data and workload are wide striped evenly across the current and newly added RAID groups that make up the pool.

Hitachi Dynamic Provisioning software's thin provisioning and wide striping functionalities provide virtual storage capacity to eliminate application service interruptions, reduce costs and simplify administration, as follows:

- Optimizes or "right-sizes" storage performance and capacity based on business or application requirements. This can reduce underutilized storage capacity in an Exchange environment.
- Supports deferring storage capacity upgrades to align with actual business usage.
- Simplifies the storage administration process. The smoothing effect reduces the need to move mailboxes to balance the user load.
- Provides performance improvements through automatic optimized wide striping of data across all available disks in a storage pool.
- Eliminates hot spots across the different RAID groups by smoothing the combined workload.
- Significantly improves capacity utilization.

For more information, see the Hitachi Dynamic Provisioning software datasheet.

Exchange Server 2010

Exchange Server 2010 introduces several architectural changes that need to be considered when planning deployments on a Hitachi Virtual Storage Platform system.

Storage Groups

Exchange Server 2010 no longer uses the concept of storage group objects. This change from previous versions of Exchange facilitates database mobility and means that Exchange data is now protected at the database level instead of at the server level. This results in a simpler and faster failover and recovery process. Limits on the number of databases per server still exist; the maximum for the standard edition of Exchange Server 2010 is five and the maximum for the enterprise edition is 100.

Database Availability Groups

To support database mobility and site resiliency, Exchange Server 2010 introduced Database Availability Groups (DAGs). A DAG is an object in Active Directory that can include up to 16 Mailbox servers that host a set of databases; any server within a DAG has the ability to host a copy of a mailbox database from any other server within the DAG. DAGs support mailbox database replication and database and server switchovers and failovers. Setting up a Windows failover cluster is no longer necessary for high availability; however, the prerequisites for setting up a DAG are similar to that of a failover cluster. Hitachi Data Systems recommends using DAGs for high availability and mailbox resiliency. Figure 4 shows 12 active databases and 12 passive databases placed in a single DAG, and they are deployed on a single Hitachi Virtual Storage Platform system.

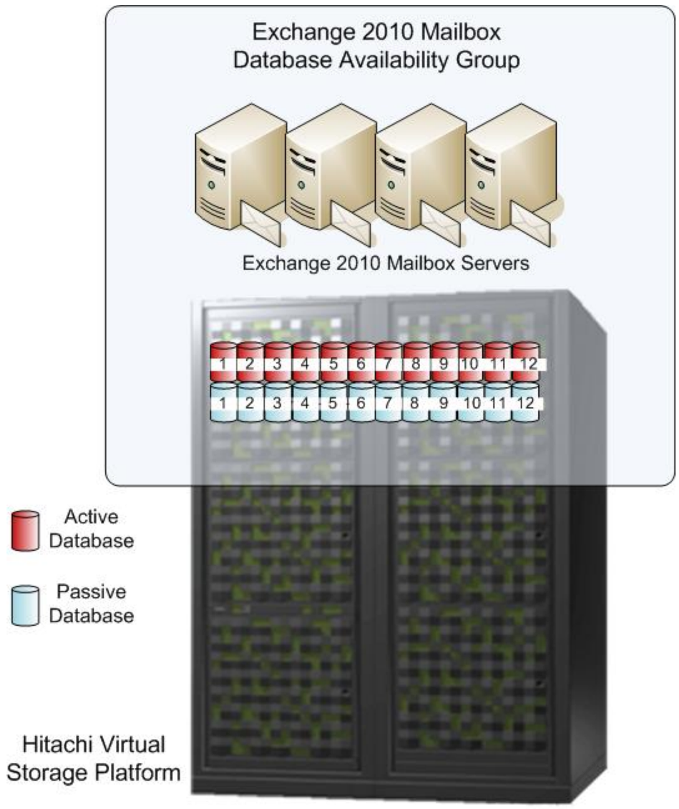

Figure 4

Databases

In Exchange Server 2010, the changes to the Extensible Storage Engine (ESE) enable the use of large databases on larger, slower disks while maintaining adequate performance. Exchange 2010 supports databases up to approximately 16TB but Microsoft recommends using databases of 2TB or less. When using databases larger than 1TB, always configure online database scanning to run continuously by enabling the **Enable background database maintenance (24 x 7 ESE scanning)** option for each database.

The Exchange Store's database tables make better use of the underlying storage system and cache and the store no longer relies on secondary indexing, making it less sensitive to performance issues.

Solution Design

Testing performed to validate this 20,000-user reference architecture used a Hitachi Virtual Storage Platform and a user profile that assumes each user sends and receives an average of 100 messages per day, which gives an I/O profile of 0.12 IOPS per mailbox. This section provides information regarding how this solution is designed, what factors need to be considered when designing the storage infrastructure as well as how to size it appropriately.

Hardware Components

Table 2 lists the detailed information about the hardware components used in the Hitachi Data Systems lab.

Table 2. Hardware Components

Hardware	*Detail Description*	*Version*	*Quantity*
Hitachi Virtual Storage Platform storage system	Single frame 6 8GB Fibre Channel ports used 128GB cache memory 240 300GB, 10K RPM, SAS disks	70-00-50-00/00	1
Brocade 5300 switch	SAN switch with 8Gb Fibre Channel ports	FOS 6.4.0b	2
Dell PowerEdge R905 server	4 x Quad-Core AMD Opteron processor 1.9 GHz, 128GB RAM. Equipped with 2 x Emulex LPe12002 8GB HBAs	BIOS firmware 4.2.1	2
Dell PowerEdge R905 server	4 x Quad-Core AMD Opteron processor 1.9 GHz, 64GB RAM. Equipped with 2 x Emulex LPe12002 8GB HBAs	BIOS firmware 4.2.1	1

Software Components

Table 3 lists the software components, including Exchange validation software, used in the Hitachi Data Systems lab.

Table 3. Software Components

Software	*Version*
Hitachi Storage Navigator	Dependent on microcode version
Hitachi Dynamic Provisioning	Dependent on microcode version
Windows Server	2008 R2, Datacenter for Hyper-V hosts, Enterprise for VMs
Microsoft Exchange Server	2010
Microsoft Exchange Load Generator 2010	14.01.0180.003
Microsoft Exchange Server Jetstress	14.01.0180.003

Solution Infrastructure

To meet the 20,000 user requirement and to use the hardware resources efficiently, the VM servers listed in Table 4 were deployed on the three physical Hyper-V hosts.

Table 4. VM Servers

VM Server Name	*Hyper-V Host Name*	*Role*	*Number of vCPUs*	*Virtual Memory (GB)*
H-DC1	R905-09	Active Directory domain controller	1	2
H-CAS-HT1	R905-09	Exchange client access and hub transport	4	8
H-CAS-HT3	R905-09	Exchange client access and hub transport	4	8
H-MB1	R905-17	Exchange mailbox	4	48
H-MB2	R905-17	Exchange mailbox	4	48
H-CAS-HT2	R905-17	Exchange client access and hub transport	4	8
H-MB3	R905-22	Exchange mailbox	4	48
H-MB4	R905-22	Exchange mailbox	4	48
H-CAS-HT4	R905-22	Exchange client access and hub transport	4	8

This reference architecture uses four mailbox servers that are placed in a single DAG, where two high-availability copies of the mailbox databases (one active and one passive) are configured. The decision to use a single DAG was based on Microsoft's recommendation to minimize the number of DAGs and to consider using more than one DAG only if one of the following conditions applies to your environment:

- You deploy more than 16 mailbox servers.
- You have active mailbox users in multiple sites.
- You require separate DAG-level administrative boundaries.
- You have mailbox servers in separate domains.

Each mailbox server contains three active databases and three passive copies. Figure 5 shows the topology of this Exchange 2010 implementation using four Hyper-V VM mailbox servers configured in a single DAG.

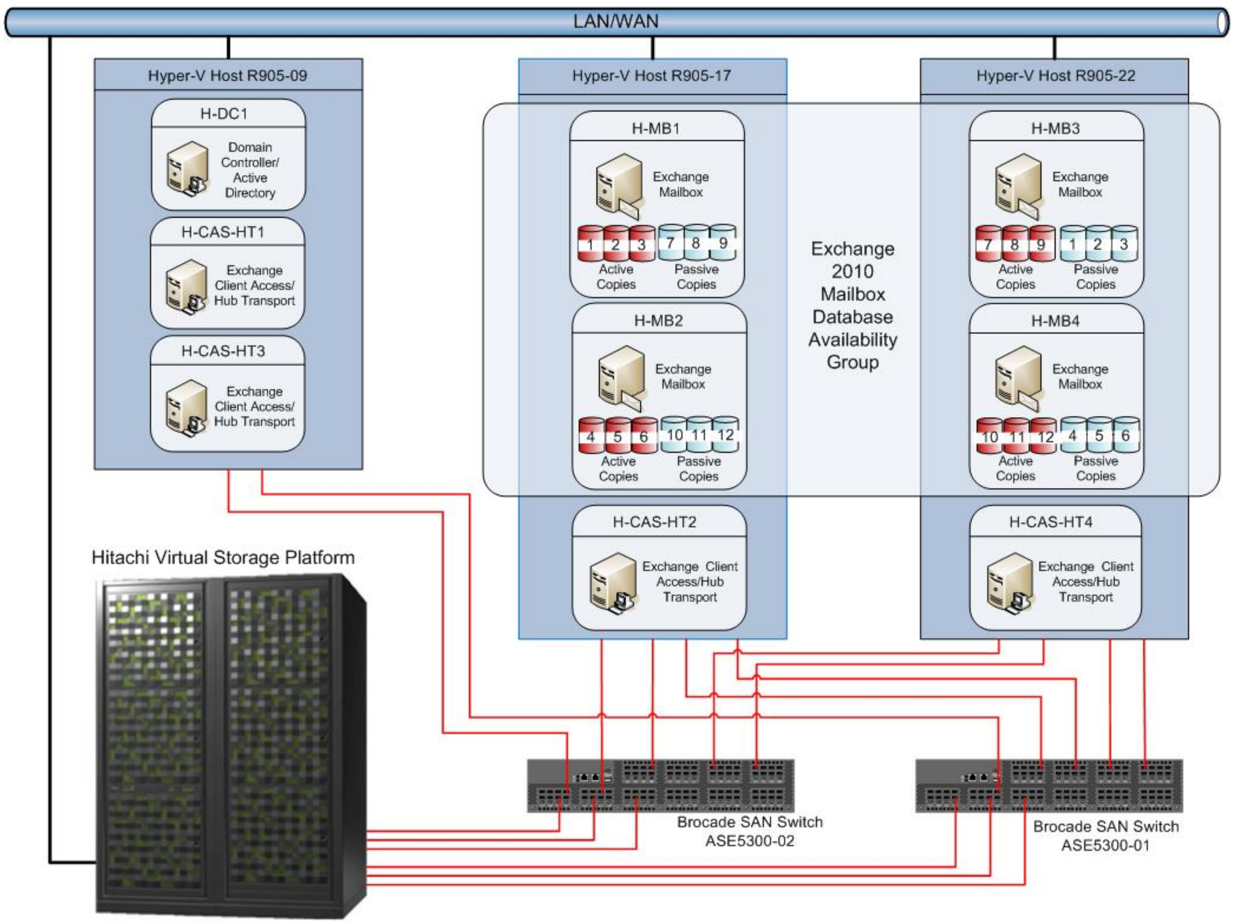

Figure 5

Designing Exchange Server Roles

Do not combine mailbox roles with any other roles in a Hyper-V environment. This is because you have only four virtual processers on a Hyper-V virtual machine. If you combine the mailbox role with other roles, you significantly limit the number of active mailboxes that can be hosted by the virtual machine.

When virtualizing the client access server role and hub transport server role, you can consider deploying both roles in the same virtual machine. This reduces the number of virtual machines to manage and the number of server operating systems to patch and maintain. The other benefit is simplification of the design process.

When deploying roles in isolation, Microsoft recommends a processor core ratio of one hub transport server logical processor for every five mailbox server logical processors, and three client access server logical processors for every four mailbox server logical processors. This can become confusing, especially when you factor in having to provide sufficient client access and hub transport servers during multiple VM server failures or maintenance scenarios. You can simplify the configuration by deploying one combined client access and hub transport server for every one mailbox server with each server having the same number of virtual CPUs.

For more information, see the Microsoft TechNet article "Understanding Client Access and Hub Transport Combined Role Configurations in Capacity Planning."

Mailbox High Availability

A key aspect of this Exchange 2010 reference architecture on the Hitachi Virtual Storage Platform is mailbox resiliency. Mailbox resiliency allows you to deploy your mailbox servers so that they are protected by automatic failover at the individual database level. This is accomplished by using a DAG. In this reference architecture, because the Exchange mailbox databases reside on intelligent, RAID-protected disks, only two copies of the databases are deployed: one active and one passive.

Microsoft recommends having at least three database copies (one active and two passive) when deploying with direct-attached storage (DAS) or just a bunch of disks (JBOD). This is because both server failure and storage (hard drive) failure need to be taken in consideration. However, in this reference architecture, only two copies of the databases are deployed (one active and one passive) because the Exchange mailbox databases reside on Virtual Storage Platform. The Virtual Storage Platform provides high performance and the most intelligent RAID-protected storage system in the industry, which reduces the possibility of storage failure. The only consideration in this case is to protect against server failure.

Reducing database copies to two instead of three provides number of benefits, including these:

- Uses less storage
- Requires less server resources
- Consumes less network traffic to replicates passive databases
- Reduces the number of databases to manage

The number of database copies required in a production environment also depends on factors such as the use of lagged database copies or backup and recovery methodologies.

SAN Architecture

For high availability purposes, the storage area network (SAN) configuration for this reference architecture uses two Fibre Channel switches. These are connected to four HBA ports on each of the two physical servers that contain virtual servers with the Exchange mailbox role installed. Four redundant paths from the switches to the Virtual Storage Platform are configured for Exchange storage volumes, and two redundant paths from the switches to the Virtual Storage Platform are used for Hyper-V VM boot volumes. Note that the Hyper-V host labeled R905-09 only has two paths from the physical server to the switches because the Virtual Storage Platform is only used for the virtual server boot volumes. Port CL-1A, CL-5A, CL-2A, and CL-6A on the Virtual Storage Platform are used for Exchange databases, and port CL-3B and CL-4B are allocated for VM OS boot volumes. Windows native MPIO is used for multipathing with the round-robin load balancing algorithm. Figure 6 illustrates the SAN design used on this reference architecture.

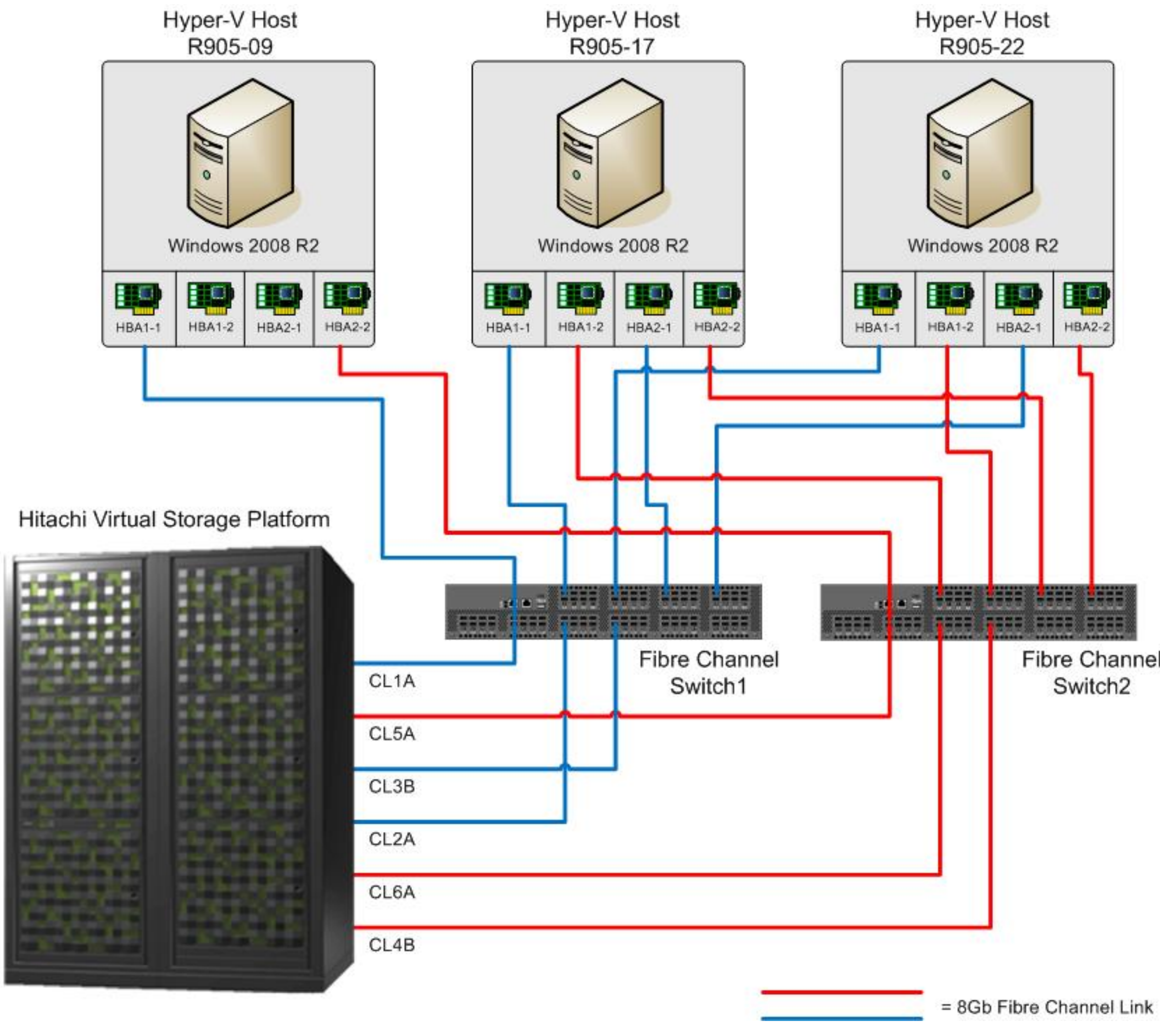

Figure 6

Network Architecture

When deploying Exchange mailbox servers in a DAG, Hitachi Data Systems recommends having two separate local area network subnets available to the members: one for client access and the other for replication purposes. The configuration is similar to the public, mixed and private networks used in previous versions of Exchange. In Exchange 2010, the two networks are now referred to as the MAPI network, which is used for communication among the DAG members and other Exchange servers, and the replication network, which is dedicated to log shipping and seeding. Using a single network is a Microsoft-supported configuration, but it is not recommended by Hitachi Data Systems. Having at least two networks connected to two separate network adapters in each server provides redundancy and enables Exchange to distinguish between a server failure and a network failure. Each DAG member must have the same number of networks and at least one MAPI network. If your environment has members of the same DAG located in different geographic sites follow Microsoft's guidance regarding network latency for DAG members. For more information about network planning for Exchange Server 2010, see the Microsoft TechNet article "[Planning for High Availability and Site Resilience](#)."

Figure 7 shows the network configuration used for this reference architecture.

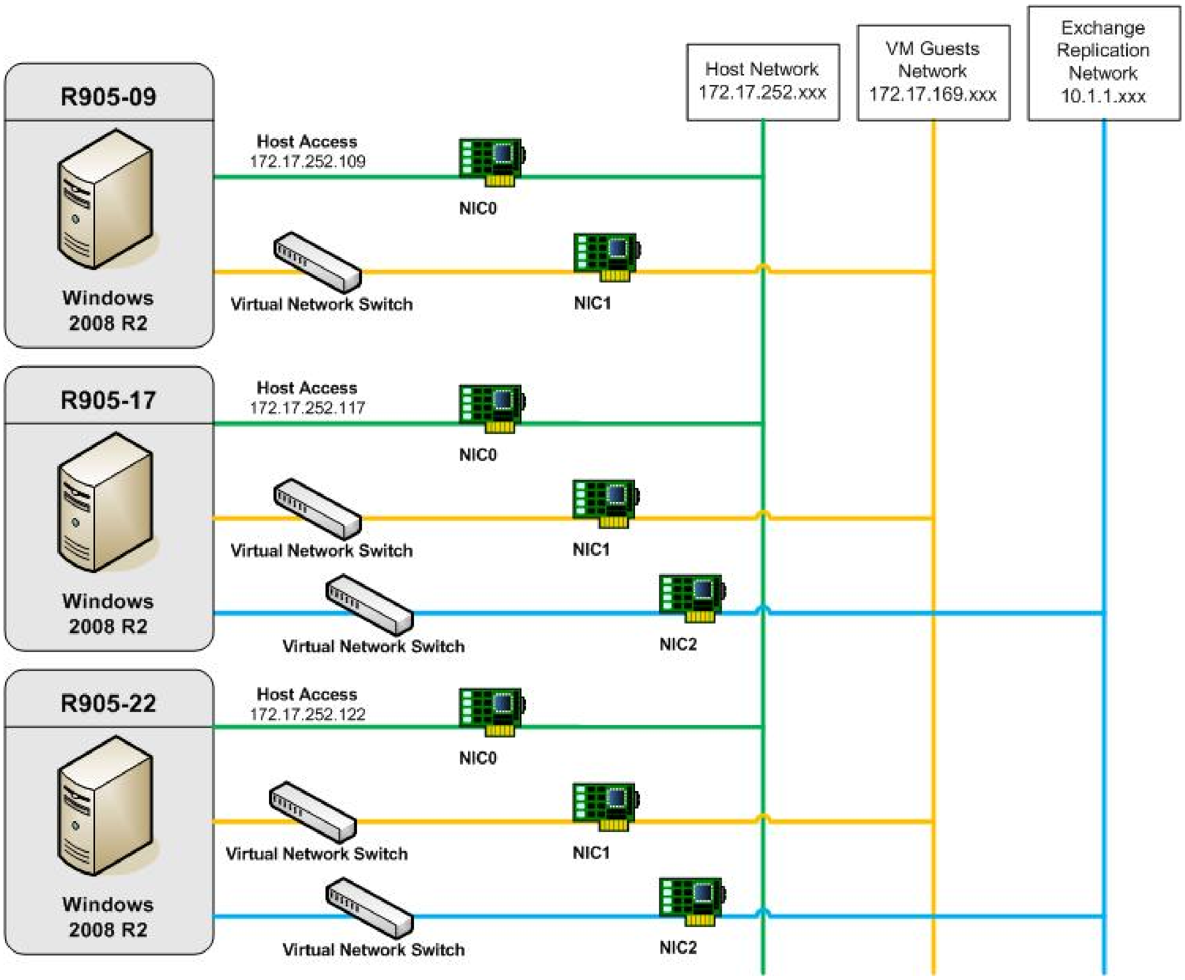

Figure 7

Storage Architecture

To satisfy 20,000 users needing 1GB of mailbox capacity and an I/O profile of 0.12 IOPS, this reference architecture uses four Dynamic Provisioning pools each created from multiple RAID-5 (7D+1P) groups on the Hitachi Virtual Storage Platform storage system. Pool 10 contains 12 active Exchange databases and Pool 11 contains 12 passive databases. Pool 12 contains 12 active Exchange logs and Pool 13 contains 12 passive logs. Hitachi Data Systems used this layout of pools and databases to test the storage in a worst-case scenario: all active databases on a single pool. However, in a production environment, Hitachi Data Systems recommends that each of the database pools contain half of the active and half of the passive databases. This allows the load to be spread evenly across both pools. Other designs are possible, but always make sure that the active and passive databases for a specific database are in separate pools.

Table 5 shows configurations for each of the Dynamic Provisioning pools used in the Hitachi Data Systems lab.

Table 5. Dynamic Provisioning Configuration

Dynamic Provisioning Pool	*RAID Configuration*	*Number of Drives*	*Drive Capacity (GB)*	*Drive Type*	*Usable Pool Capacity (TB)*
Pool 10	RAID-5 (7D+1P)	104	300	SAS	24.42
Pool 11	RAID-5 (7D+1P)	104	300	SAS	24.42
Pool 12	RAID-5 (7D+1P)	16	300	SAS	3.75
Pool 13	RAID-5 (7D+1P)	16	300	SAS	3.75
Pool 00	RAID-5 (3D+1P)	4	2,000	SATA	5.37

Pool 00 consists of 2TB SATA drives configured as RAID-5 (3D+1P) and is used for virtual server OS volumes. This pool was shared with other test environments within the lab. The configuration of the storage used by the Hyper-V hosts in a production environment depends on the requirements of the specific environment.

Table 6 lists LUN and LDEV configuration. Each LDEV on the storage is presented as a LUN to the hosts. One database is stored in each LUN, and each Hyper-V host has one LDEV on Pool 00 assigned to the host for VM OS volumes.

Table 6. LUN Configuration

LUN Allocation	*Dynamic Provisioning Pool*	*LDEV Size (GB)*	*LDEVs*	*Storage Ports*
Active database LUNs	Pool 10	2,000	10:01,10:02,10:03,10:04,10:05,10:06,10:07,10:08,10:09,10:0A,10:0B,10:0C	CL1A,CL2A,CL5A,CL6A
Passive database LUNs	Pool 11	2,000	11:01,11:02,11:03,11:04,11:05,11:06,11:07,11:08,11:09,11:0A,11:0B,11:0C	CL1A,CL2A,CL5A,CL6A
Active log LUNs	Pool 12	200	12:01,12:02,12:03,12:04,12:05,12:06,12:07,12:08,12:09,12:0A,12:0B,12:0C	CL1A,CL2A,CL5A,CL6A
Passive log LUNs	Pool 13	200	13:01,13:02,13:03,13:04,13:05,13:06,13:07,13:08,13:09,13:0A,13:0B,13:0C	CL1A,CL2A,CL5A,CL6A
VM OS boot LUNs	Pool 00	600	01:01,01:02,01:03	CL3B,CL4B

Determining I/O Requirements

When designing the storage architecture for Exchange 2010, always start by calculating the I/O requirements. You must determine how many IOPS each mailbox needs. This is also known as determining the I/O profile. Microsoft has guidelines and tools available to help you determine this number. Two factors are used to estimate the I/O profile: how many messages a user sends and receives per day and the amount of database cache available to the mailbox. The database cache (which is located on the mailbox server) is used by the ESE to reduce I/O operations. Generally, more cache means less I/O operations eventually hitting the storage system. Table 7 lists Microsoft's guidelines.

Table 7. Estimated IOPS per Mailbox

Messages Sent and Received per Mailbox per Day	*Database Cache per Mailbox (MB)*	*Standalone Estimated IOPS per Mailbox*	*Mailbox Resiliency Estimated IOPS per Mailbox*
50	3	0.06	0.05
100	6	0.12	0.10
150	9	0.18	0.15
200	12	0.24	0.20
250	15	0.30	0.25
300	18	0.36	0.30
350	21	0.42	0.35
400	24	0.48	0.40
450	27	0.54	0.45
500	30	0.60	0.50

For this reference architecture, an I/O profile of 100 messages a day, or 0.1 IOPS per mailbox, was used. To ensure that the architecture can provide sufficient overhead for periods of extremely high workload, Hitachi adds 20 percent overhead for testing scenarios for a total of 0.12 IOPS.

To calculate the total number of IOPS or transactions per second (TPS) for an Exchange environment use the following formula:

(# of users) x (estimated IOPS per mailbox) = required host IOPS (TPS)

For example:

20,000 users x 0.12 IOPS = 2400 IOPS

This means that each of the four mailbox servers in this reference architecture must be able to support 600 IOPS.

This calculation provides the number of application IOPS required by the host to service the environment, but it does not calculate the exact number of physical IOPS required on the storage side. Additional calculations must be performed to factor in the read-write ratio used by Exchange Server 2010 and the write penalty incurred by the various types of RAID levels. To calculate how many read and write IOPS are required, see the Microsoft TechNet article "Understanding Database and Log Performance Factors."

The transaction logs in Exchange Server 2010 require approximately 10 percent as many I/Os as the databases. For mailbox databases in a resiliency scenario, the log write I/O is 50 percent of the database write I/O. For log read I/O in a mailbox resiliency situation, apply a 10 percent overhead to account for the use of continuous replication. For example, if the databases on a server require 600 I/Os, the logs require 60 I/Os. Of those 60 I/Os, 30 are write I/O and 30 plus 10 percent (or 33) are read I/O. After you calculate the transactional log I/O, Microsoft recommends adding another 20 percent overhead to ensure adequate capacity for busier-than-normal periods.

Determining Capacity Requirements

In addition to the mailbox quota requirement, you must also consider the size of the database dumpster and the amount of white space the database is likely to have. The database always has free pages or white space spread throughout. During online maintenance, items marked for removal from the database are removed, freeing those pages. The amount of white space in the database can be estimated by knowing the amount of mail (in MB) sent and received by the users with mailboxes inside the database.

Each database also has a dumpster that stores items deleted from a user's mailbox. By default, these items are stored for 14 days (calendar items are stored for 120 days). In addition to the dumpster, Exchange Server 2010 includes the single item recovery feature. This feature is disabled by default and it prevents the purging of data before the deleted item retention window has passed. When enabled, this feature increases the size of the mailbox for a two-week period and must be considered when determining the capacity requirements.

You must also consider content indexing, which allows you to perform a quick and easy search of your mail items. This factor contributes about 10 percent of the total database size of additional overhead.

The personal archive feature was not included in this reference architecture; however, if you plan to use it you must increase the storage quota for each mailbox to include the additional data. A personal archive mailbox can be created at any time. Before SP1, the archive mailbox resided on the same mailbox database as the user's primary mailbox. SP1 allows the archive mailbox to be stored in a different database. Special quotas for the archive are configured at creation to help manage the growth of the archive data. By default, the archive quotas are set to unlimited. The personal archive was created to eliminate the need for `.pst` files by allowing users to keep historical data within the Exchange database rather than on their local computers. By bringing the archive data back into the Exchange database, users and administrators can manage this data with features like retention policies and legal holds, they can recall the data easily with multi-mailbox search and this data can now be protected by native Exchange processes.

If you plan to use a recovery database in your environment, you must allocate sufficient capacity for all the databases you plan to recover per server.

The transaction log files maintain a record of every transaction and operation performed by the Exchange 2010 database engine. Transactions are written to the log first then written to the database. The message size and I/O profile (based on the number of messages per mailbox per day) can help estimate how many transaction logs are generated per day. Table 8 provides guidelines for estimating how many transaction logs are generated for a 75K average message size.

Table 8. Number of Transaction Logs Generated per I/O Profile for 75K Average Message

I/O Profile	*Transaction Logs Generated per Day*
50	10
100	20
150	30
200	40
250	50
300	60
350	70
400	80
450	90
500	100

As message size increases, the number of logs generated per day grows. According to Microsoft, if your message size doubles to 150K, the logs generated per mailbox increases by a factor of 1.9. If the message size doubles again to 300K, the factor of 1.9 doubles to 3.8, and so on.

Consider these additional factors when determining transaction log capacity:

- Backup and restore factors
- Move mailbox operations
- Log growth overhead
- High availability factors

If you plan to include lag copies in your Exchange environment, you must determine capacity for both the database copy and the logs. The log capacity requirements depend on the delay and usually require more capacity than the non-lagged copy.

The amount of space required for logs also depends on your backup methodology and how often logs are truncated.

For more information about calculating capacity requirements, see the Microsoft TechNet article "Understanding Mailbox Database and Log Capacity Factors."

You can download Microsoft's Exchange 2010 Mailbox Server Role Requirements Calculator from the Microsoft Exchange Team Blog.

RAID Group and LU Design

This reference architecture uses enterprise-class 300GB 10K SAS drives in the Hitachi Virtual Storage Platform. Mid-tier drives (in the 7200RPM range) can be used for Exchange 2010 as long as the number of spindles can satisfy Exchange's I/O requirements. The Virtual Storage Platform can house both SAS and SATA drives. The decision about which type to use also depends on the I/O and capacity requirements of the Exchange environment.

Because of the reduced I/O requirements of Exchange 2010 and the use of Hitachi Dynamic Provisioning, Hitachi Data Systems recommends the following:

- Use SAS drives when mailboxes are 2GB or smaller with I/O requirements of 0.12 IOPS per user or higher.
- Use SAS or SATA drives when mailboxes are between 2GB and 5GB in size. Your choice is partially dependent on your I/O requirements.
- Use SATA drives when mailboxes are over 5GB and I/O requirements are 0.12 IOPs per user or lower.
- Use RAID-1+0 (2D+2D) for SATA.

When using SAS drives, a RAID-5 (7D+1P) on the Virtual Storage Platform can meet the requirements for most Exchange environments. RAID-5 offers better capacity utilization and is therefore more cost effective than RAID-1+0. Due to the lower I/O requirements and the features of the Hitachi Virtual Storage Platform, a RAID-5 configuration is more than adequate to meet the I/O requirements for this reference architecture.

Hitachi Dynamic Provisioning software is used in this reference architecture and Hitachi Data Systems recommends using it for all Exchange 2010 implementations. Separate the database and log LUs into different pools. .

Server Design

If you need a high availability Exchange solution, use at least two physical servers. Use enterprise-class servers that contain enough processor and memory resources to accommodate a failover situation. The servers must have enough memory and processor power to operate the entire Exchange environment in the event that a server fails.

When designing a server infrastructure for an Exchange 2010 environment, follow Microsoft recommendations for sizing of servers and for ratio of CPUs between the servers with the various Exchange 2010 roles installed. For this reference architecture, a one-to-one ratio of virtual CPUs on the mailbox servers and client access/hub transport servers was maintained.

Servers used in a Hyper-V environment must meet certain hardware requirements. For more information, see the *Hyper-V Planning and Deployment Guide* from the Microsoft Web site.

Designing an Exchange 2010 implementation requires consideration of processor and memory requirements for your mailbox servers. Many articles are available on the Microsoft TechNet site to assist with proper planning and implementation of Exchange 2010.

Processor Capacity Planning

With the release of Exchange 2010, Microsoft has new processor configuration recommendations for servers that host the mailbox role. This is due to the implementation of mailbox resiliency. It is now based on two factors: whether the server will host both active and passive database copies and the number of database copies. A passive database copy requires CPU resources to perform the following tasks:

- Check or validate replicated logs
- Replay replicated logs into the database
- Maintain the content index associated with the database copy

For more information about CPU planning for the mailbox server role, see the Microsoft TechNet article "Mailbox Server Processor Capacity Planning."

Physical Memory Sizing

Memory sizing for the Exchange mailbox server role is critical to reducing the I/O workload presented by the server to the storage platform. Increasing the amount of memory on the mailbox server results in fewer I/Os to the storage system.

For more information about memory planning for the mailbox server role, see the Microsoft TechNet article "Understanding the Mailbox Database Cache."

Engineering Validation

The following sections describe the configuration used for building and validating the Exchange environment documented in this guide.

Test Methodology

The following sections describe the test methodology used to validate this reference architecture.

Exchange Jetstress 2010

Exchange Jetstress 2010 is used to verify the performance and stability of a disk subsystem prior to putting an Exchange 2010 server into production. It helps verify disk performance by simulating Exchange database and log file I/O loads. It uses Performance Monitor, Event Viewer and ESEUTIL to ensure that the storage system meets or exceeds your performance criteria. Jetstress generates I/O based on Microsoft's estimated IOPS per mailbox user profiles. The test was performed from four servers against three databases for each server, for total of 12 databases for 24-hour window. The goal was to verify the storage was able handle high I/O load for a long period of time. Table 9 lists the Jetstress parameters used in testing.

Table 9. Jetstress Test Parameters

Parameter	Value
Number of databases	12
User profile	0.12
Number of users per database	1667
Total number of users	20,000
Mailbox size	1GB

Test Results

The following sections provide results of the testing conducted to validate this reference architecture.

Exchange Jetstress 2010

A 24-hour test was run concurrently against the four servers and 12 database instances. All latency and achieved IOPS results met Microsoft requirements and all tests passed without errors.

Individual Server Metrics

Table 10 lists transaction I/O performance for Server1.

Table 10. Transactional I/O Performance (Server1)

Database	I/O Database Reads Average Latency (ms)	I/O Database Writes Average Latency (ms)	I/O Database Reads/sec	I/O Database Writes/sec	I/O Log Writes Average Latency (ms)	I/O Log Writes/sec
1	7.61	1.33	333.60	195.27	1.27	100.52
2	6.61	1.58	334.17	196.06	1.25	99.61
3	6.59	1.91	332.64	195.53	1.25	99.77

Table 11 lists transactional I/O performance for Server2.

Table 11. Transactional I/O Performance (Server2)

Database	I/O Database Reads Average Latency (ms)	I/O Database Writes Average Latency (ms)	I/O Database Reads/sec	I/O Database Writes/sec	I/O Log Writes Average Latency (ms)	I/O Log Writes/sec
4	7.40	1.32	339.98	199.38	1.25	102.36
5	6.52	1.53	341.66	200.77	1.24	101.74
6	6.55	1.86	339.40	199.75	1.24	101.60

Table 12 lists transactional I/O performance for Server3.

Table 12. Transactional I/O Performance (Server3)

Database	I/O Database Reads Average Latency (ms)	I/O Database Writes Average Latency (ms)	I/O Database Reads/sec	I/O Database Writes/sec	I/O Log Writes Average Latency (ms)	I/O Log Writes/sec
7	7.56	1.45	317.72	186.82	1.48	94.58
8	7.06	1.65	318.16	187.01	1.46	94.14
9	7.08	1.96	318.18	187.29	1.46	94.20

Table 13 lists transactional I/O performance for Server4.

Table 13. Transactional I/O Performance (Server4)

Database	I/O Database Reads Average Latency (ms)	I/O Database Writes Average Latency (ms)	I/O Database Reads/sec	I/O Database Writes/sec	I/O Log Writes Average Latency (ms)	I/O Log Writes/sec
10	7.51	1.42	329.05	193.03	1.39	98.08
11	6.80	1.59	327.78	193.08	1.37	97.30
12	6.82	1.90	327.46	192.90	1.37	97.25

Table 14 lists database sizing and throughput for three database files for Server1.

Table 14. Database Sizing and Throughput (Server1)

Metric	Result
Achieved transactional I/O per second	1587.28
Target transactional I/O per second	600.12
Initial database size (bytes)	5265320771584
Final database size (bytes)	5301299511296

Table 15 lists database sizing and throughput for three database files for Server2.

Table 15. Database Sizing and Throughput (Server2)

Metric	Result
Achieved transactional I/O per second	1620.93
Target transactional I/O per second	600.12
Initial database size (bytes)	5265010393088
Final database size (bytes)	5301727330304

Table 16 lists database sizing and throughput for three database files for Server3.

Table 16. Database Sizing and Throughput (Server3)

Metric	Result
Achieved transactional I/O per second	1515.18
Target transactional I/O per second	600.12
Initial database size (bytes)	5261973716992
Final database size (bytes)	5296509616128

Table 17 lists database sizing and throughput for three database files for Server4.

Table 17. Database Sizing and Throughput (Server4)

Metric	Result
Achieved transactional I/O per second	1563.30
Target transactional I/O per second	600.12
Initial database size (bytes)	5262997127168
Final database size (bytes)	5298413830144

Conclusion

This reference architecture guide details a tested robust Exchange Server 2010 resiliency solution capable of supporting 20,000 users with a 0.12 IOPS per user profile and a user mailbox size of 1GB. This reference architecture uses a single DAG configured with four mailbox servers. A Hitachi Virtual Storage Platform with Hitachi Dynamic Provisioning software was used for these tests. Testing confirmed that the reference architecture is more than capable of delivering the IOPS and capacity requirements needed to support the active and replicated databases for 20,000 Exchange mailboxes configured with the specified user profile, while maintaining additional headroom to support peak throughput.

The solution outlined in this document does not include data protection components, such as VSS snapshot or clone backups, and relies on the built-in mailbox resiliency features of Exchange Server 2010 coupled with the RAID technology of the Hitachi Virtual Storage Platform to provide high availability and protection from logical and physical failures. Adding protection requirements can affect performance and capacity requirements of the underlying storage configuration, and as such needs to be factored into your storage design accordingly.

Many factors affect the storage design for an Exchange 2010 environment, so it is critical that you validate the storage design for your environment before deployment.

Hitachi Data Systems Global Services offers experienced storage consultants, proven methodologies and a comprehensive services portfolio to assist you in implementing Hitachi products and solutions in your environment. For more information, see the Hitachi Data Systems Global Services web site.

Live and recorded product demonstrations are available for many Hitachi products. To schedule a live demonstration, contact a sales representative. To view a recorded demonstration, see the Hitachi Data Systems Corporate Resources web site. Click the **Product Demos** tab for a list of available recorded demonstrations.

For more information about Hitachi products, contact your sales representative or channel partner or visit the Hitachi Data Systems web site.

Hitachi Data Systems Corporation

Corporate Headquarters
750 Central Expressway,
Santa Clara, California 95050-2627 USA
www.hds.com

Regional Contact Information
Americas: +1 408 970 1000 or info@hds.com
Europe, Middle East and Africa: +44 (0) 1753 618000 or info.emea@hds.com
Asia Pacific: +852 3189 7900 or hds.marketing.apac@hds.com