Inspire the Next

Deploying an 8,000-user Exchange Server 2010 Environment with Hitachi Compute Blade 2000 and the Hitachi Adaptable Modular Storage 2300

Reference Architecture Guide

By Jeff Chen and Leo Nguyen

February 2011

Microsoft GOLD CERTIFIED Partner

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 an 8,000-user Exchange Server 2010 Environment with Hitachi Compute Blade 2000 and the Hitachi Adaptable Modular Storage 2300

Reference Architecture Guide

It can be difficult to design, deploy, manage and support Microsoft Exchange 2010 environments with hardware components from multiple vendors. This solution reduces the complexity of Exchange environments by using servers and storage from Hitachi that work together seamlessly.

This white paper describes a reference architecture that provides system reliability and high availability in a Microsoft Exchange 2010 environment by using the Hitachi Compute Blade 2000 with N+1 cold standby and logical partitioning technologies, the Hitachi Adaptable Modular Storage 2000 family, and Exchange 2010 Database Availability Groups (DAGs).

The Hitachi Compute Blade 2000 is an enterprise-class platform that offers the following features:

- Balanced system architecture that eliminates bottlenecks in performance and throughput
- Embedded Hitachi logical partitioning (LPAR) virtualization
- Unprecedented configuration flexibility
- Eco-friendly power-saving features and capabilities
- Fast recovery from server failures due to N+1 cold standby design that allows you to replace failed servers within minutes instead of hours or days

With its unique combination of power, efficiency and flexibility, you can now extend the benefits of virtualization to new areas of the enterprise data center — including mission-critical application servers and database servers like Exchange 2010 — with minimal cost and maximum simplicity.

The 2000 family is ideal for a demanding application like Exchange and delivers enterprise-class performance, capacity and functionality at a midrange price. It’s a midrange storage product with symmetric active-active controllers that provide integrated, automated, hardware-based, front-to-back-end I/O load balancing.

This white paper describes a reference architecture for an Exchange 2010 deployment that supports 8,000 users on the Hitachi Compute Blade 2000 and the Hitachi Adaptable Modular Storage 2300. The Exchange databases are protected against failure with a Database Availability Group (DAG).

Figure 1 shows the design of the reference architecture documented in this white paper. Note that the active and passive databases are shown on the servers to which they are mapped from the 2300.

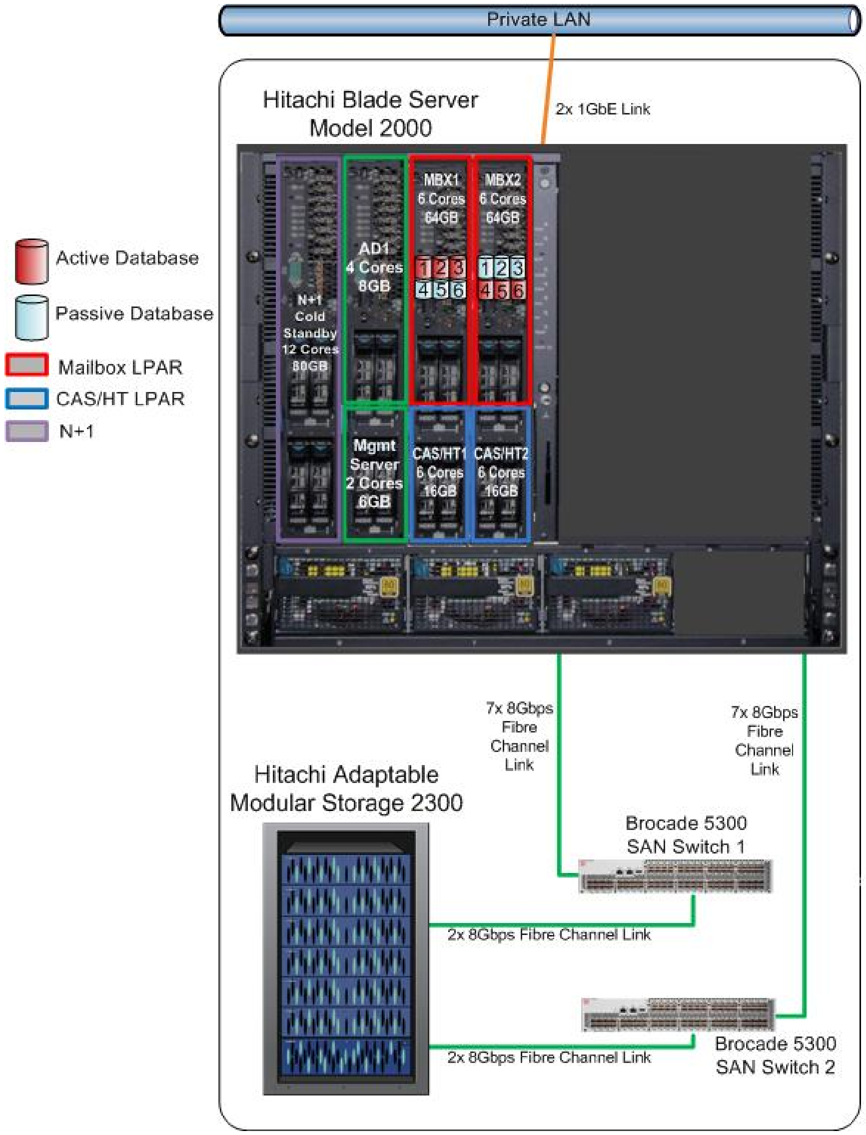


Figure 1

This reference architecture guide is intended for IT administrators involved in data center planning and design, specifically those with a focus on the planning and design of Microsoft Exchange Server 2010. It assumes some familiarity with Hitachi Adaptable Modular 2000 family, Hitachi Storage Navigator Modular 2 software, Microsoft Windows Server 2008 R2 and Exchange Server 2010.

Key Solution Components

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

Hitachi Compute Blade 2000

The Hitachi Compute Blade 2000 features a modular architecture that delivers unprecedented configuration flexibility, as shown in Figure 2.

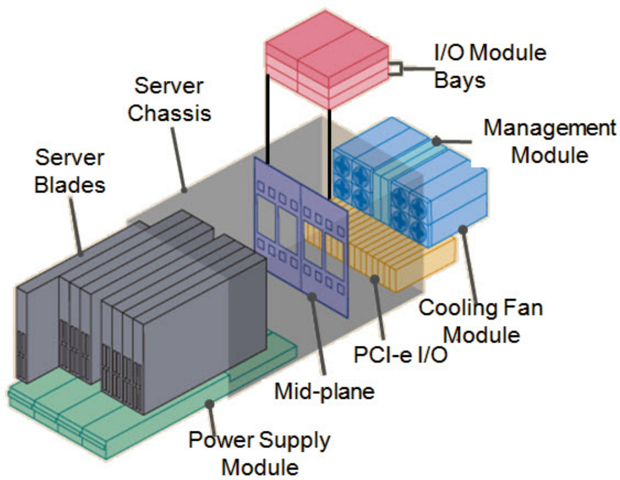


Figure 2

The Hitachi Compute Blade 2000 combines all the benefits of virtualization with all the advantages of the blade server format: simplicity, flexibility, high compute density and power efficiency. This allows you to take advantage of the following benefits:

- Consolidate more resources
- Extend the benefits of virtualization solutions (whether Hitachi logical partitioning, VMware vSphere, Microsoft Hyper-V, or all three)
- Cut costs without sacrificing performance

Hitachi Compute Blade 2000 enables you to use virtualization to consolidate application and database servers for backbone systems, areas where effective consolidation was difficult in the past. And by removing performance and I/O bottlenecks, Hitachi Compute Blade 2000 opens new opportunities for increasing efficiency and utilization rates and reduces the administrative burden in your data center.

No blade system is more manageable or flexible than the Hitachi Compute Blade 2000. You can configure and administer the model 2000 using a web-based HTML browser that supports secure encrypted communications, or leverage the optional management suite to manage multiple chassis using a unified GUI-based interface.

Chassis

The Hitachi Compute Blade 2000 chassis is a 19-inch rack compatible, 10U-high chassis with a high degree of configuration flexibility. The front of the chassis has slots for eight server blades and four power supply modules, and the back of the chassis has six bays for I/O switch modules, eight fan modules, two management modules, 16 half-height PCIe slots and connections for the power cables, as shown in Figure 3.

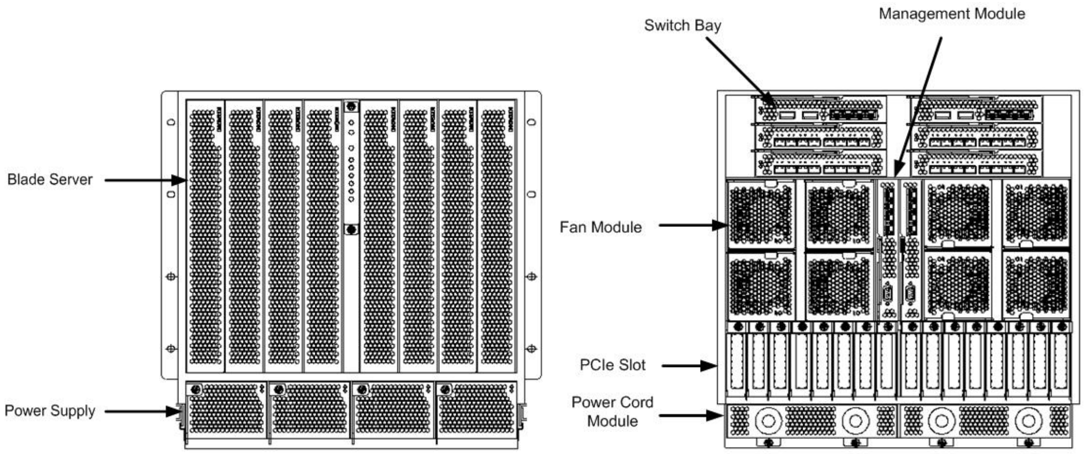


Figure 3

All modules, including fans and power supplies, can be configured redundantly and hot swapped, maximizing system uptime.

The Hitachi Compute Blade 2000 accommodates up to four power supplies in the chassis and can be configured with mirrored power supplies, providing backup on each side of the chassis and higher reliability. Cooling is provided by efficient, variable-speed, redundant fan modules. Each fan module includes three fans to tolerate fan failures within a module, but in addition, if an entire module fails, the other fan modules will continue to cool the chassis.

Server Blades

The Hitachi Compute Blade 2000 supports two blade server options that can be combined within the same chassis. Table 1 lists the specifications for each option.

Table 1. Hitachi Compute Blade 2000 Specifications

Feature	*X55A2*	*X57A1*
Processors (up to two per blade)	Intel Xeon 5600 - 4 or 6 cores	Intel Xeon 7500 – 6 or 8 core
Processor cores	4,6,8 or 12	6, 8, 12 or 16
Memory slots	18	32
Maximum memory	144GB (with 8GB DIMMS)	256GB (with 8GB DIMMS)
Hard drives	Up to 4	N/A
Network interface cards (on-board)	Up to 2 1Gb Ethernet	Up to 2 1Gb Ethernet
Other interfaces	2 USB 2.0 port and 1 serial port	2 USB 2.0 and 1 serial port
Mezzanine slots	Up to 2	Up to 2
PCIe 2.0 (8x) expansion slots	Up to 2	Up to 2

Up to four X57A1 blades can be connected using the SMP interface connector to create a single eight-socket SMP system with up to 64 cores and 1024GB of memory.

I/O Options

The connections from the server blades through the chassis' mid-plane to the bays or slots on the back of the chassis consist of the following:

- The two on-board NICs connect to switch bays one and two
- The optional mezzanine card in the first mezzanine slot connects to switch bays three and four
- The optional mezzanine card in the second mezzanine slot connects to switch bays five and six
- Two connections to dedicated PCIe slots

The I/O options supported by the optional mezzanine cards and the switch modules are either 1Gb Ethernet or 8Gb Fibre Channel connectivity.

Logical Partitioning

The Hitachi logical partitioning (LPAR) virtualization feature is embedded in the firmware of Hitachi Compute Blade 2000 server blades. This is a proven, mainframe-class technology that combines Hitachi LPAR expertise with Intel VT technologies to improve performance, reliability and security.

Unlike emulation solutions, the embedded LPAR virtualization feature does not degrade application performance, and unlike third-party virtualization solutions, it does not require the purchase and installation of additional components, keeping total cost of ownership low. A blade can operate in one of two modes, basic or Hitachi Virtualization Manager (HVM), with two licensing options available in the HVM mode, as follows:

- **Basic** — The blade operates as a standard server without LPAR support.
- **HVM with essential license** — The blade supports two LPARs. No additional purchase is required for this mode.
- **HVM with enterprise license** — The blade supports up to 16 LPARs. The enterprise license is an additional cost.

You can use the embedded LPAR feature alone or combine it with Microsoft Hyper-V, VMware vSphere or both in a single system, providing additional flexibility.

Hitachi Compute Blade 2000 Management Modules

The Hitachi Compute Blade 2000 supports up to two management modules for redundancy. Each module is hot-swappable and supports live firmware updates without the need for shutting down the blades. Each module supports an independent management LAN interface from the data network for remote and secure management of the chassis and all blades. Each module supports a serial command line interface and a web interface. SNMP and email alerts are also supported.

N+1 or N+M Cold Standby Failover

The Hitachi Compute Blade 2000 maintains high uptime levels through sophisticated failover mechanisms. The N+1 cold standby function enables multiple servers to share a standby server, increasing system availability while decreasing the need for multiple standby servers or costly software-based high-availability servers. It enables the system to detect a fault in a server blade and switch to the standby server, manually or automatically. The hardware switching is executed even in the absence of the administrator, enabling the system to return to normal operations within a short time.

The N+M cold standby function has “M” backup server blades for every “N” active server blade, so failover is cascading. In the event of multiple hardware failures, the system automatically detects the fault and identifies the problem by indicating the faulty server blade, allowing immediate failure recovery. This approach can reduce total downtime by enabling the application workload to be shared among the working servers.

Hitachi Adaptable Modular Storage 2300

The Hitachi Adaptable Modular Storage 2300 is a member of the Hitachi Adaptable Modular Storage 2000 family, which provides a reliable, flexible, scalable and cost-effective modular storage system for this solution. Its symmetric active-active controllers provide integrated, automated, hardware-based, front-to-back-end I/O load balancing. Both controllers in a 2000 family storage system are able to dynamically and automatically assign the access paths from the controller to the LU. All LUs are accessible regardless of the physical port or the server from which the access is requested. Utilization rates of each controller are monitored so that a more even distribution of workload between the two controllers can be maintained. Storage administrators are no longer required to manually define specific affinities between LUs and controllers, simplifying overall administration. In addition, this controller design is fully integrated with standard host-based multipathing, thereby eliminating mandatory requirements to implement proprietary multipathing software.

The point-to-point back-end design virtually eliminates I/O transfer delays and contention associated with Fibre Channel arbitration and provides significantly higher bandwidth and I/O concurrency. It also isolates any component failures that might occur on back-end I/O paths.

For more information about the 2000 family, see the *Hitachi Adaptable Modular Storage 2000 Family Overview* brochure.

Hitachi Dynamic Provisioning Software

On Hitachi Adaptable Modular Storage 2000 family systems, Hitachi Dynamic Provisioning software provides wide striping and thin provisioning that dramatically improve performance, capacity utilization and management of your environment. By deploying your Exchange Server 2010 mailbox servers using volumes from Hitachi Dynamic Provisioning storage pools on the 2300, you can expect the following benefits:

- An improved I/O "buffer" to burst into during peak usage times
- A smoothing effect to the Exchange workload that can eliminate hot spots, resulting in reduced mailbox moves related to performance
- Minimization of excess, unused capacity by leveraging the combined capabilities of all disks comprising a storage pool

Hitachi Dynamic Link Manager Software

The Hitachi Dynamic Link Manager software was used for SAN multipathing, configured with the round-robin multipathing policy. Hitachi Dynamic Link Manager software's round-robin load balancing algorithm automatically selects a path by rotating through all available paths, thus balancing the load across all available paths and optimizing IOPS and response time.

Hitachi Server Conductor

Hitachi Server Conductor is suite of programs for centralized management of multiple servers. Server Conductor provides functions for managing servers efficiently, including functions to manage server software configurations and monitor server operating statuses and failures.

The Blade Server Manager Plus component of Server Conductor is required to implement an N+1 or N+M cold standby configuration.

Compared to a cluster configuration, an N+1 cold standby configuration reduces operating costs because a single standby server is shared by multiple active servers.

Exchange Server 2010

Exchange Server 2010 introduces several architectural changes that need to be considered when planning deployments on a Hitachi Adaptable Modular Storage 2000 system.

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.

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.

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

This reference architecture describes a solution designed for 8,000 users. Each user has a 1GB mailbox, and the user profile is based on sending and receiving 100 messages per day. This solution has six databases, with 1,333 users per database. With a single-site, two-mailbox-server configuration, each server holds 4,000 users. Each server houses the following:

- Three active databases
- Three passive databases that are copies of the active databases on the other server

If a server fails, the other server makes the three passive database copies active and handles the user load for all 8,000 users.

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

Table 2. Hardware Components

Hardware	Description	Version	Quantity
Hitachi Adaptable Modular Storage 2300 storage system	Dual controller 16x 8GB Fibre Channel ports 16GB cache memory 80 SAS 2TB 7.2K RPM disks	0897/A-Y	1
Brocade 5300 switch	8Gb Fibre Channel ports	FOS 6.4.0E	2
Hitachi Compute Blade 2000 chassis	8-blade chassis 8 x 8GB dual-port HBAs 8 x 1GB network ports 2 x management modules 8 x cooling fan modules 3 x power supply modules	A0154-E-5234	1
Hitachi Compute Blade 2000 E55A2 blades	Full blade 2 x 6-Core Intel Xeon X5670 2.93GHz 80GB memory	58.22	4
Hitachi dual-port HBA	Dual port 8Gbps Fibre Channel PCIe card	4.2.6.670	8

Table 3 lists the software components used in the Hitachi Data Systems lab.

Table 3. Software Components

Software	Version
Hitachi Storage Navigator Modular 2	Microcode dependent
Hitachi Dynamic Provisioning	Microcode dependent
Hitachi Dynamic Link Manager	6.5
Microsoft Windows Server	Windows 2008 R2 Enterprise
Microsoft Exchange Server 2010	SP1
Hitachi Server Conductor	80-90-A

Hitachi Compute Blade 2000 Chassis Configuration

This reference architecture uses four standard X55A2 blades, two standard 1Gb LAN switch modules, and eight Hitachi dual port 8Gbps HBA cards. Each blade has two on-board NIC, and each NIC is connected to a LAN switch module. Each blade has two PCIe slots available, and all are populated with Hitachi dual port 8Gbps HBA cards. Hitachi HBAs are required for HVM mode. With Hitachi HBA, when HVM is enabled, eight virtual WWNs are created for LPARs. Figure 4 shows the front and back view of Hitachi Blade Server Model 2000 used in this solution.

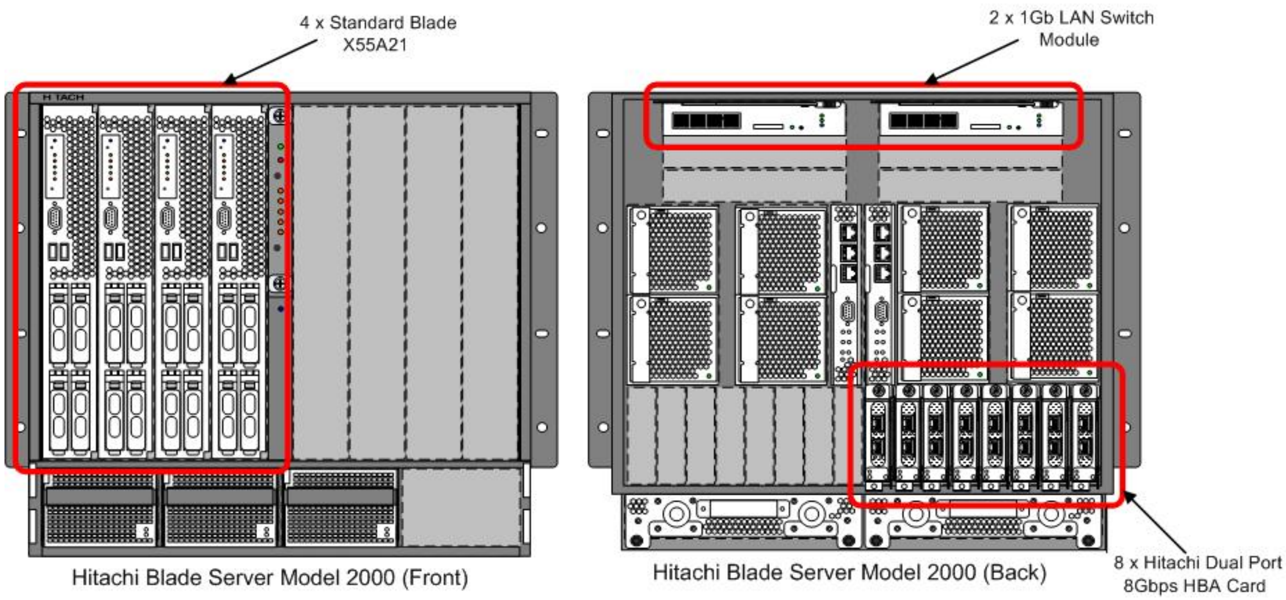


Figure 4

Hitachi Compute Blade 2000 Blade Configuration

When designing a server infrastructure for an Exchange 2010 environment, user profiling is the first and most critical piece in the design process. Hitachi Data Systems evaluated the following factors when designing this reference architecture:

- Total number of user mailboxes
- Mailbox size limit
- Total send and receive per day
- Average message size
- Deleted item retention
- Single item recovery

With this information, Hitachi Data Systems determined the correct sizing and number of Exchange mailbox roles to deploy using the Exchange 2010 Mailbox Server Role Requirements Calculator, which is an Excel spreadsheet created and supported by the Exchange product group. The goal of the calculator is to provide guidance on the I/O and capacity requirements and a storage design. This complex spreadsheet follows the latest Exchange product group recommendations on storage, memory, and mailbox sizing. Table 4 lists the optimized configurations for the Exchange roles and servers for the environment described in this white paper.

Table 4. Hitachi Compute Blade 2000 Configuration

Blade Server	LPAR	Server Name	Role	Number of CPU Cores	Memory (GB)
Blade 0	N/A	N/A	N+1 failover standby	12	80
Blade 1	LPAR1	BS-Mgmt1	Blade management server	2	4
	LPAR2	BS-AD1	Active Directory DNS	4	8
Blade 2	LPAR1	BS-CASHT1	Client access Hub transport	6	16
	LPAR2	BS-MBX1	Mailbox server	6	64
Blade 3	LPAR1	BS-CASHT2	Client access Hub transport	6	16
	LPAR2	BS-MBX2	Mailbox server	6	64

For more information about the Microsoft Exchange sizing calculator, see the "Exchange 2010 Mailbox Server Role Requirements Calculator" entry on the Microsoft Exchange Team Blog.

Processor Capacity

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 this reference architecture, the following formulas were used to calculate the CPU requirement for the Exchange roles, taking failover into consideration:

Megacycles per mailbox = (average CPU usage × speed of processors in megacycles) × (number of processors ÷ number of mailboxes)

CPU usage = (number of users × current megacycles per mailbox) ÷ (number of processors x speed of processors in megacycles)

Physical Memory

This reference architecture supports 8,000 users who send and receive 100 messages per day. Based on Table 5, the mailbox resiliency estimate IOPS per mailbox was 0.10 with a 20 percent overhead for a total of 0.12 IOPS and 6MB for database cache per mailbox.

Table 5. Database Cache Guidelines

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

Based Table 5, Hitachi Data Systems calculated the database cache size as 8,000 * 6 MB = 48GB. After determining the database cache size of 48GB, Hitachi Data Systems used Table 6 to determine physical memory. 64GB is the ideal memory configuration based on this mailbox count and user profile. Table 6 lists Microsoft's guidelines for determining physical memory capacity.

Table 6. Physical Memory Guidelines

Server Physical Memory (GB)	*Database Cache Size (GB)*
2	0.5
4	1.0
8	3.6
16	10.4
24	17.6
32	24.4
48	39.2
64	53.6
96	82.4

N+1 Cold Standby Configuration

Hitachi Compute Blade 2000's N+1 cold standby feature is used in this solution to enhance the reliability of the entire Exchange environment. When combined with Exchange's DAG feature, Hitachi Compute Blade 2000's N+1 cold standby feature allows rapid recovery from server or database failures:

- DAG allows recovery within seconds in the event of a failed database or server but all users are hosted on a single server
- N+1 cold standby allows full recovery within minutes in the event of a failed server with users distributed across two servers

See the "Engineering Validation" section for relevant test results.

In this reference architecture, blade 0 is allocated as a cold standby blade. Server Conductor is required for this feature, and it is installed on the management server on blade 1. Exchange servers are installed on blades 2 and 3, and they are registered as active servers in Server Conductor. When hardware failure occurs, failover to standby blade is performed automatically, and hardware related Information is automatically transferred using the blade management network during the failover. You do not need to change the SAN or LAN configuration after the failover.

Hitachi Data Systems considered the following requirements when configuring the N+1 cold standby feature for this reference architecture:

- The standby blade must be in the same mode (basic or HVM) as the active blades. The active blades in HVM mode cannot failover to basic mode standard blade. Likewise, an active blade in basic mode cannot failover to a blade in HVM mode.
- The operating systems installed on active blades must be configured for SAN boot.

Exchange Server Roles Design

In this reference architecture, the client access server and hub transport server roles are combined, while the mailbox role was installed on a separate LPAR. The primary reasons are to minimize the number of servers, operating system instances, and Exchange servers to manage and to optimize performance for planned or unplanned failover scenarios

Mailbox High Availability Design

In this reference architecture, only two copies of the databases are deployed: one active and one passive. This is because the Exchange mailbox databases reside on intelligent, RAID-protected disks, 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.

Figure 5 shows the solution topology using two mailbox servers configured in a single DAG. Note that the active and passive databases are shown on the servers to which they are mapped from the 2300.

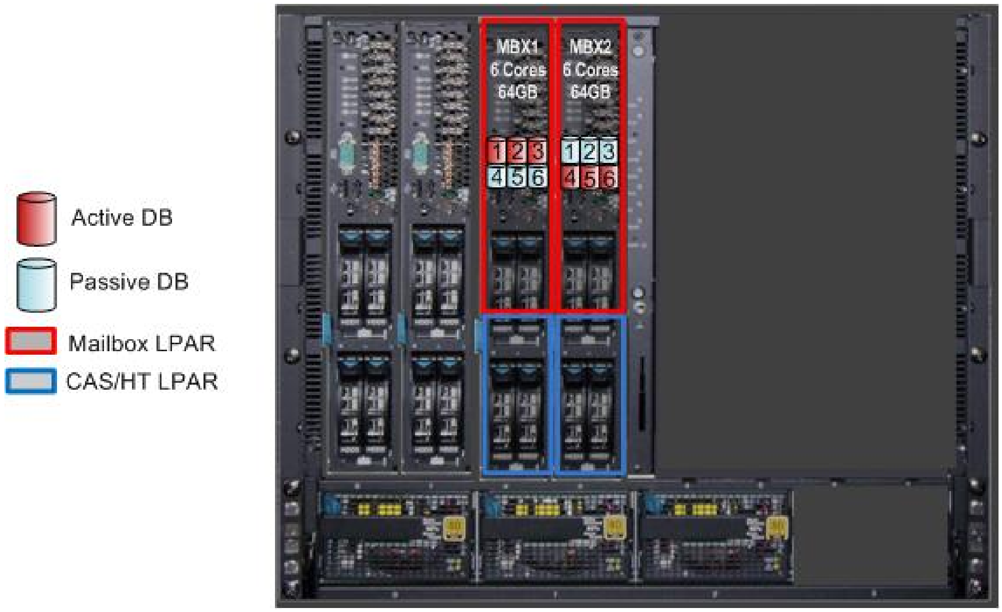


Figure 5

Each mailbox server hosts 4,000 users. If the mailbox server that hosts the active database fails, or if an active mailbox database becomes inaccessible, the passive mailbox database becomes active on the other server. That server is able to handle the load of all 8,000 users.

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 a Hitachi Adaptable Modular Storage 2300 storage system. The Hitachi Adaptable Modular Storage 2300 provides high performance and the most intelligent RAID-protected storage system in the industry, which reduces the possibility of storage failure. Reducing database copies to two instead of three provides a 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 and the backup and recovery methodologies used.

SAN Architecture

For high availability purposes, the storage area network (SAN) configuration for this reference architecture uses two Fibre Channel switches. SAN boot volumes for the OS are connected to two HBA ports. Exchange volumes are connected to four HBA ports. Four redundant paths from the switches to the 2300 are configured for Exchange volumes, and two redundant paths from the switches to the 2300 are used for SAN OS boot volumes. Port 0A, 1A, 0E and 1E on the 2300 are used for Exchange databases and logs, and port 0C and 1C are allocated for SAN OS boot volumes. Hitachi Dynamic Link Manager software is used for multipathing with the round-robin load balancing algorithm.

Figure 6 shows the SAN design used in this reference architecture.

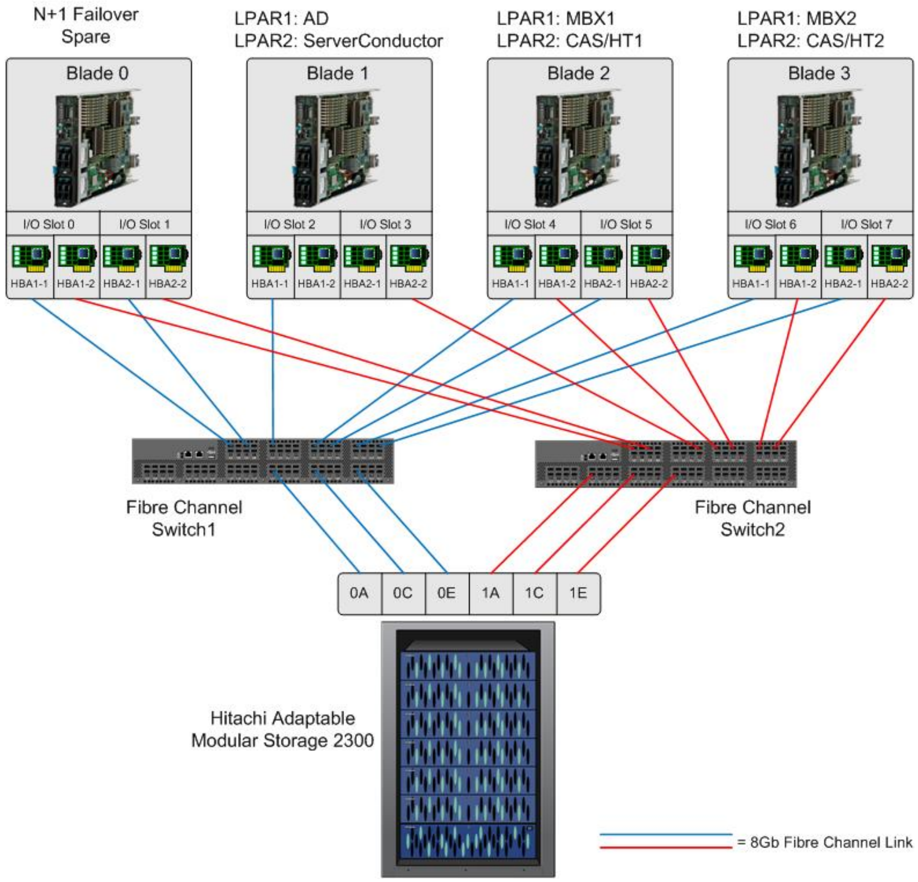


Figure 6

Network Architecture

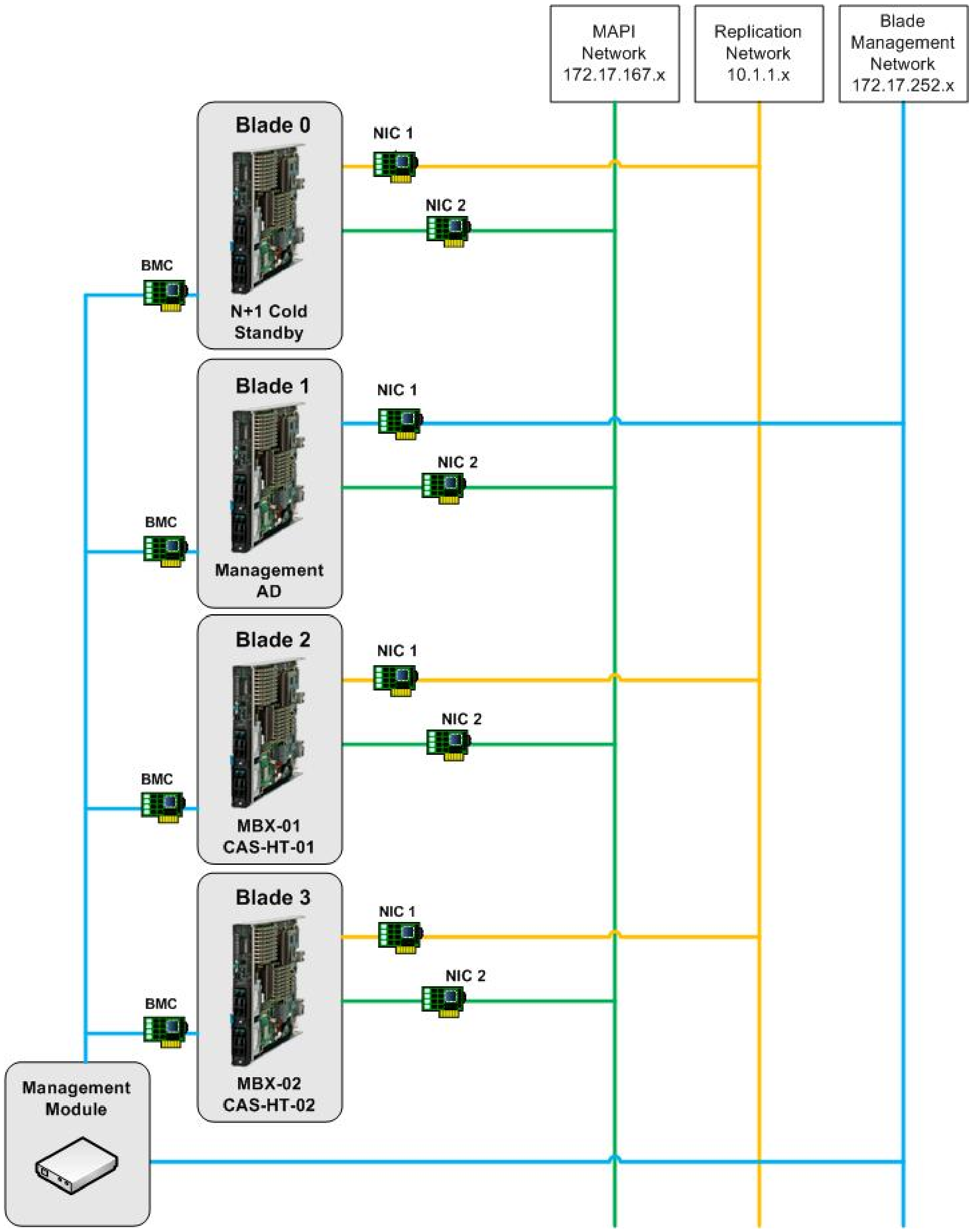
When deploying Exchange mailbox servers in a DAG, Hitachi Data Systems recommends having two separate local area network subnets available to the members: one for client access and the other for replication purposes. The configuration is similar to the public, mixed and private networks used in previous versions of Exchange. In Exchange 2010, the two networks have new names:

- **MAPI network** — Used for communication between Exchange servers and client access
- **Replication network** — Used for 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.

This reference architecture uses two on-board NICs. The network connections on blade 0, which acts as an N+1 cold standby server, are the same as the network connections on blade 2 and 3. NIC1 on the server hosting Server Conductor on blade 1 is connected to the blade management network. A baseboard management controller (BMC) is built into each blade and it is internally connected to the management module. The management module, BMC and HVM work in coordination to report hardware failures. Server Conductor can monitor the blade through the blade management network.

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



MAPI Network 172.17.167.x
Replication Network 10.1.1.x
Blade Management Network 172.17.252.x
Blade 0
NIC 1
NIC 2
BMC
N+1 Cold Standby
Blade 1
NIC 1
NIC 2
BMC
Management AD
Blade 2
NIC 1
NIC 2
BMC
MBX-01 CAS-HT-01
Blade 3
NIC 1
NIC 2
BMC
MBX-02 CAS-HT-02
Management Module

Figure 7

Storage Architecture

Sizing and configuring storage for use with Exchange 2010 can be a complicated process, driven by many factors such as I/O and capacity requirements. The following sections describe how Hitachi Data Systems determined storage sizing for this reference architecture.

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 fewer 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 host IOPS for an Exchange environment, Hitachi Data Systems used the following formula:

Number of users x estimated IOPS per mailbox = required host IOPS

For example:

8,000 users x 0.12 IOPS = 960 host IOPS

Because this reference architecture uses two servers, the host IOPS calculation is divided by two. This means that each of the two mailbox servers in this reference architecture must be able to support 480 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 were 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.

The transaction logs in Exchange Server 2010 require approximately 10 percent as many I/Os as the databases. After you calculate the transactional log I/O, Microsoft recommends adding another 20 percent overhead to ensure adequate capacity for busier-than-normal periods.

Log Requirements

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

Hitachi Data Systems considered these additional factors when determining transaction log capacity for this reference architecture:

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

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 lists 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

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. For this reference architecture, no lag copy was used.

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

RAID Configuration

To satisfy 8,000 users needing 1GB of mailbox capacity and an I/O profile of 0.12 IOPS, this reference architecture uses a RAID-1+0 (2D+2D) configuration of 2TB SAS drives. RAID 1+0 is primarily used for overall performance. RAID-1+0 (2D+2D) was used for this solution because Hitachi Data Systems testing shows it is an optimal configuration for this disk size and speed. The four Dynamic Provisioning pools used for this solution were created from 20 RAID-1+0 (2D+2D) groups on the Hitachi Adaptable Modular Storage 2300. Table 9 lists configurations for each of the Dynamic Provisioning pools used in the Hitachi Data Systems lab.

Table 9. Dynamic Provisioning Configuration for Exchange Databases and Logs

Dynamic Provisioning Pool	Number of RAID Groups	Number of Drives	Usable Pool Capacity (TB)
0	9	36	30.0
1	9	36	30.0
2	1	4	3.5
3	1	4	3.5

Pool Configuration

Pool 0 and 1 contain a mix of three active and three passive Exchange databases each. Pool 2 and 3 contain a mix of three active and three passive Exchange logs each. Hitachi Data Systems recommends keeping the database and log on separate pools for performance reasons. Table 10 shows LUN allocation for each of the Dynamic Provisioning pools used in the Hitachi Data Systems lab.

Table 10. LUN Allocation for Exchange Databases and Logs

LUN Allocation	Dynamic Provisioning Pool	LUN Size (GB)	LUNs	Storage Ports
Active and passive database LUNs	Pool 0	2000	00,01,02,03,04,05	0A,1A,0E,1E
Active and passive database LUNs	Pool 1	2000	06,07,08,09,10,11	0A,1A,0E,1E
Active and passive log LUNs	Pool 2	100	20,21,22,23,24,25	0A,1A,0E,1E
Active and passive log LUNs	Pool 3	100	26,27,28,29,30,31	0A,1A,0E,1E

Drive and LUN Configuration

Each LUN on the storage is presented as a LUN to the hosts. With a total of 12 LUNs, six LUNs are used for databases and the other six LUNs are used for logs. The Exchange LUNs are configured as RAID-1+0 (2D+2D) and separated into four different pools (P0 – P3). The SAN OS boot is configured as RAID-5 (4D+1P), RG0. Table 11 lists the detailed drive configuration used in this solution for drive slots 0 through 11.

Table 11. Drive Configuration for Slots 0 to 11

Drive Slot	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
RKA 4	P2	P2	P2	P2	P3	P3	P3	P3				
RKA 3	P1	P1	P1	P1	P1	P1	P1	P1	P1	P1	P1	P1
RKA 2	P0	P0	P0	P0	P0	P0	P0	P0	P0	P0	P0	P0
RKA 1	P0	P0	P0	P0	P0	P0	P0	P0	P0	P0	P0	P0
RKA 0	RG0	RG0	RG0	RG0	RG0							

Table 12 lists the detailed drive configuration used in this solution for drive slots 12 to 23.

Table 12. Drive Configuration for Slots 12 to 23

Drive Slot	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*
RKA 4												
RKA 3	P1	P1	P1	P1	P1	P1	P1	P1	P1	P1	P1	P1
RKA 2	P1	P1	P1	P1	P1	P1	P1	P1	P1	P1	P1	P1
RKA 1	P0	P0	P0	P0	P0	P0	P0	P0	P0	P0	P0	P0
RKA 0												

SAN OS Boot Configuration

Each active server blade is divided into two logical partitions. Each LPAR operates as an independent OS. To use LPARs on the blades, SAN OS boot is required. This reference architecture uses a single RAID group, RG0, that consists of five 300GB SAS drives configured as RAID-5 (4D+1P) on the 2300 for the OS boot volumes. Each boot volume is 100GB in size and mapped to storage ports 0C and 1C.

Table 13 lists LUN allocation for RG0 used in the Hitachi Data Systems lab.

Table 13. LUN Allocation for SAN OS Boot

Server	*Role*	*LUN*
Blade 1 LPAR 1	Active Directory	12
Blade 1 LPAR 2	Management	13
Blade 2 LPAR 1	Client access 1 Hub transport 1	14
Blade 2 LPAR 2	Mailbox 1	14
Blade 3 LPAR 1	Client access 2 Hub transport 2	15
Blade 3 LPAR 2	Mailbox 2	16

Engineering Validation

The following sections describe the testing used for validating the Exchange environment documented in this guide.

Exchange Jetstress 2010 Test

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 on two Hitachi Compute Blade 2000 servers against three databases for each server, for total of six 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 14 lists the Jetstress parameters used in testing.

Table 14. Jetstress Test Parameters

Parameter	*Value*
Number of databases	6
User profile	0.12
Number of users per database	1,333
Total number of users	8,000
Mailbox size	1GB

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

Table 15 lists transaction I/O performance for Server 1.

Table 15. Transactional I/O Performance (Server 1)

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	17.996	3.614	109.844	65.771	0.771	56.411
2	16.714	3.411	109.967	65.835	0.690	56.321
3	17.257	3.524	109.626	65.614	0.769	56.393
4	17.407	5.022	109.759	65.716	0.724	56.332
5	16.326	3.501	109.780	65.725	0.403	57.164
6	17.377	4.738	109.545	65.557	0.732	56.293

Table 16 lists transactional I/O performance for Server 2.

Table 16. Transactional I/O Performance (Server 2)

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	15.273	3.326	107.766	64.607	0.769	55.077
2	17.364	5.421	107.900	64.646	0.573	56.185
3	15.011	3.077	108.004	64.773	0.402	56.281
4	18.884	3.809	107.892	64.634	0.836	55.472
5	18.451	3.700	108.045	64.726	0.835	55.624
6	17.951	3.637	108.273	64.861	0.856	55.660

Table 17 lists aggregate target and achieved throughput for three databases for Server 1.

Table 17. Throughput Test Results (Server 1)

Metric	*Result*
Target transactional I/O per second	960.000
Achieved transactional I/O per second	1052.738

Table 18 lists aggregate target and achieved throughput for three databases for Server 2.

Table 18. Throughput Test Results (Server 2)

Metric	*Result*
Target transactional I/O per second	960.000
Achieved transactional I/O per second	1036.127

Exchange Loadgen 2010 Test

Exchange Load Generator is a pre-deployment validation and stress-testing tool that introduces various types of workloads into a test (non-production) Exchange messaging system. Exchange Load Generator lets you simulate the delivery of multiple MAPI client messaging requests to an Exchange server. To simulate the delivery of these messaging requests, you run Exchange Load Generator tests on client computers. These tests send multiple messaging requests to the Exchange server, which causes a mail load.

The simulation was set for a normal eight hours test run with six databases and an Outlook 2007 Online Mode profile. The goal was for all tests to achieve a CPU utilization rate of less than 80 percent. All tests passed without errors.

Exchange DAG Switchover and Failover Tests

In these tests, the objective was to validate that the two-member DAG design can sustain a single mailbox server failure using a switchover (planned maintenance) or a failover (unplanned outage). The switchover test was conducted by selecting the databases and manually performing a switchover. The failover test was conducted by shutting down one mailbox server to determine if the other mailbox server can handle the full 8,000 users with CPU utilization of less than 80 percent.

Table 19 lists the DAG test results.

Table 19. DAG Test Results

Test Criteria	*Target*	*Result*
Switchover	Sustain 8,000 users	Passed
Failover	Sustain 8,000 users	Passed
Switchover CPU utilization	<80%	69%
Failover CPU utilization	<80%	71%
Switchover time	N/A	15 seconds
Failover time	N/A	18 seconds

N+1 Cold Standby Switchover and Failover Tests

The objective of these tests was to demonstrate that in the case of a failure of an active blade, the N+1 blade can take over for the failed blade.

To perform these tests, Hitachi Server Conductor and Blade Server Manager Plus were installed on the blade management server (BS-Mgmt1). Each blade server was registered to Blade Server Manager Plus, and an N+1 group was created. The standby failover blade and active blades were added into this N+1 group. The manual switchover test was performed from Blade Server Manager. The failover test was performed from the management module by issuing a false hardware failure alert.

Table 20 shows the N+1 cold standby test results.

Table 20. N+1 Cold Standby Test Results

Test	*Recovery Time*
Switchover	6 minutes, 33 seconds
Failover	9 minutes, 2 seconds

Conclusion

This reference architecture guide documents a fully tested, highly available Exchange 2010 deployment for 8,000 users. This solution provides simplified planning, deployment and management by using hardware components from a single vendor.

The solution uses the Hitachi Compute Blade 2000 that extends the benefits of virtualization to new areas of the enterprise data center with minimal cost and maximum simplicity. It also allows rapid recovery — within minutes instead of hours or days — in the event of a server failure.

The Hitachi Adaptable Modular Storage 2300 is ideal for a demanding application like Exchange and delivers enterprise-class performance, capacity and functionality at a midrange price. It's a midrange storage product with symmetric active-active controllers that provide integrated, automated, hardware-based, front-to-back-end I/O load balancing.

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.

Hitachi Data Systems Academy provides best-in-class training on Hitachi products, technology, solutions and certifications. Hitachi Data Systems Academy delivers on-demand web-based training (WBT), classroom-based instructor-led training (ILT) and virtual instructor-led training (vILT) courses. For more information, see the Hitachi Data Systems Academy web site.

For more information about Hitachi products and services, 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