Microsoft Windows Server System

Server Consolidation Using Exchange Server 2003

Product Version:	Exchange Server 2003
Reviewed By:	Exchange Product Development
Latest Content:	www.microsoft.com/exchange/library
Author:	Kay Unkroth

Server Consolidation Using Exchange Server 2003

Kay Unkroth

Published: February 2004

Applies To: Exchange Server 2003

Copyright

Acknowledgments

Project Editor: Susan Bradley

Contributing Editors: mediaService, Amy Groncznack

Technical Reviewers: Robert Quimbey, Matt Gossage, and Nicole Bonilla

Graphic Design: Kristie Smith

Production: Joe Orzech

Table of Contents

Introduction

As e-mail messaging continues to grow in both volume and business importance, organizations are looking for new options to manage future demand in a reliable and cost-effective way. One option is to build a messaging strategy based on advanced technologies available in Microsoft® Windows Server™ 2003 and Microsoft Exchange Server 2003. Enterprise-scale messaging servers are the key to reducing the overall number of servers, as well as the number of server sites in the organization, thus streamlining messaging infrastructure and lowering administrative overhead. Information Technology analysts suggest that server and site consolidation can help an organization to realize significant savings in operational costs. In a recent study, META Group concludes that by accommodating more users per Exchange server, organizations can reduce operational costs by approximately $240,000 to $600,000 annually. For more information, see *Exchange 5.5 Migration to Exchange 2003 Cost-Saving Scenarios* (http://go.microsoft.com/fwlink/?LinkId=23453).

For example, Microsoft upgraded to Exchange Server 2003, reducing 113 mailbox servers in 75 sites to 38 mailbox servers in seven geographical locations worldwide. The organization has approximately 85,000 users and manages approximately 3,500,000 internal e-mail messages and 2,500,000 Internet e-mail messages per day. This example is not unique. Siemens is upgrading to Exchange 2003 and consolidating more than 150 Exchange organizations into a single organization. The window and door manufacturer, Pella, upgraded to Exchange 2003 and consolidated 16 Exchange servers to six Exchange servers. Timex upgraded to Exchange 2003 and consolidated its nine e-mail servers to only two e-mail servers. Each of these examples demonstrates that organizations can benefit from site and server consolidation using Exchange 2003. If your organization includes multiple messaging systems, servers, or locations, you can benefit from a consolidation based on Exchange 2003.

Improvements in Performance and Scalability

Improving performance and scalability has always been an important design goal for the Exchange product development team. The first Exchange product, Microsoft Exchange Server 4.0, was designed to replace MS Mail for PC Networks. MS Mail could host no more than 500 mailboxes per post office. Most organizations, however, required a more enterprise-ready messaging platform. The information store in Exchange Server 4.0 and Microsoft Exchange Server 5.0, in contrast, could hold up to 16 GB of data and more than 500 mailboxes.

Storage capacity was further increased in Microsoft Exchange Server 5.5. Exchange Server 5.5 Enterprise Edition featured an unlimited Exchange information store. Now organizations could place any number of mailboxes on a single server, and Exchange databases could quickly grow to enormous sizes of several hundred GB. With its large database sizes, however, the messaging system was difficult to maintain. Although the number of mailboxes that could be contained within the Exchange store was theoretically unlimited, 1,000 mailboxes per server was the approximate practical limit.

Increasing the practical number of mailboxes per server was a major design goal for Microsoft Exchange 2000 Server. Exchange 2000 Server featured support for multiple storage groups and could hold up to 20 mailbox and public folder databases. When mailboxes were distributed over a number of mailbox stores, servers could host 2,000 or more mailboxes. The individual databases nevertheless remained manageable.

Exchange 2003 improves upon Exchange 2000 in the areas of performance, scalability, and security. Exchange 2003 continues to support multiple storage groups and messaging databases, features an additional Recovery Storage Group, and supports the Volume Shadow Copy service of Windows Server 2003. Exchange 2003 uses improved caching algorithms so that directory lookups can be completed

more efficiently, resulting in a 60 percent reduction of Microsoft Active Directory® directory service queries compared with Exchange 2000. Virtual memory management is improved also, and the propagation of link state information between servers is optimized to reduce network traffic. Also noteworthy is the high level of virus and spam protection that you can achieve by using Exchange 2003. Upgrading to Exchange 2003 from previous versions of Exchange, especially Exchange Server 5.5, offers clear incentives.

Improvements in Client/Server Communication

The more mailboxes an organization places on a single server, the more client connections the server must manage. Consequently, optimizing client/server communication is an important aspect of server consolidation. Microsoft Office Outlook® 2003 is an ideal messaging client for Exchange 2003. This client contains numerous features that are specifically designed to reduce network bandwidth consumption and improve the e-mail experience of users over remote connections in an Exchange 2003 organization.

The following are key improvements that you can make in client/server communication by using Outlook 2003. These improvements enable you to place a large number of mailboxes on a single server running Exchange 2003.

- **MAPI compression and buffer packing** When you use Outlook 2003 in an Exchange 2003 organization, mailbox content is compressed on the Exchange server before being sent to the Outlook 2003 client. In addition, the data is packaged in large, optimized buffer packets, thereby reducing the number of requests that must be transferred over the network between the Outlook client and the server running Exchange 2003. Both MAPI compression and buffer packing significantly lower the network bandwidth requirements for client/server communication and enable an Exchange server to manage an increased number of users with acceptable response times.
- **Exchange cached mode** This feature enables Outlook 2003 to download all items from the server-based mailbox and keep them synchronized in a cache on the local client computer. After a full copy of the mailbox is downloaded, the client performs most e-mail related tasks from the local computer. Communication with the server is only required during offline folder synchronization, when downloading new items to the client computer, when uploading added or changed items to the server, or when sending messages. Figure 1 illustrates the Exchange cached mode principle.

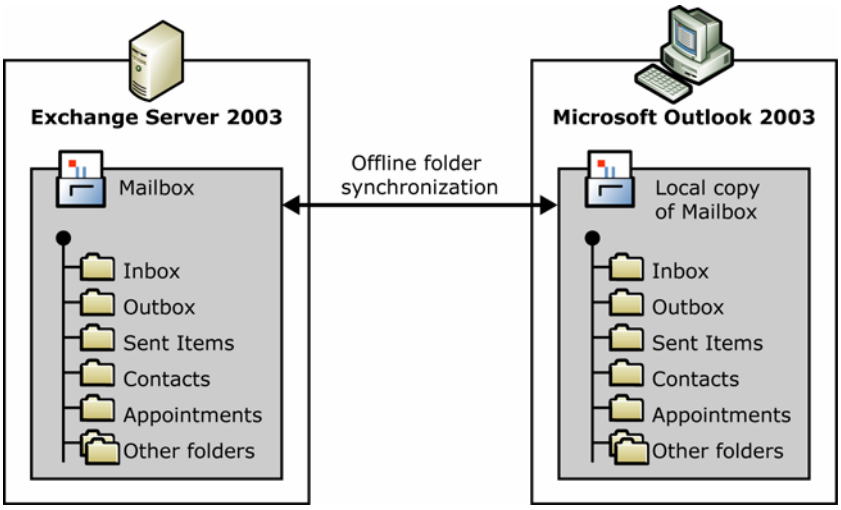

Figure 1 Synchronizing server-based mailbox content in Exchange cached mode

Exchange cached mode can potentially reduce both the workload of servers and the consumption of network bandwidth. Exchange cached mode also improves the remote or branch office user's experience of Outlook 2003. Access to items stored in the local cache is always fast, no matter how slow or unreliable the network connection, and access to the items stored in the local cache does not depend on server availability. Because the client works with a local copy of the data, users do not need to restart Outlook 2003 to an offline profile when network interruptions occur or when the server is shut down temporarily for maintenance reasons.

In Exchange cached mode, users can choose from the following download options.

- **Download full items** This is the default setting and the best option when you use Exchange cached mode over fast and reliable network connections. The client downloads all items to the client immediately.
- **Download headers first and then full items** In this option, the client first performs a header download, so that users can immediately see a list of new items. Following the header download, the client performs a full download to bring all data to the client. This option might be a good choice if users want immediate access to new messages, but note that this mode requires an extra cycle of client/server communication compared with downloading full items.
- **Download headers only** This is the best option when you use Exchange cached mode over slow and unreliable network links, such as dial-up connections. The client downloads the message headers only, so that users can see the size of e-mail messages and attachments, and then decide which items to download or delete directly on the server. The client downloads a message when users open it in Outlook 2003.

> Note
> Because of its many advantages, it is recommended that you enable Exchange cached mode over local as well as remote network connections.

- **Improved Outlook 2003 synchronization performance** Outlook 2003 performance in Exchange cached mode is further improved by reducing the number of change notifications between client and server. The server informs the client about the number and size of messages to be downloaded, and when items are marked read or unread, are flagged, or are modified in other ways, only the header listing the change is sent back to the server. When full items are downloaded, MAPI compression ensures efficient data flow.

  It is important to understand that the synchronization in Outlook 2003 is incremental. That is, if a synchronization cycle is interrupted, the synchronization process resumes where the interruption occurred, instead of resuming the entire synchronization process. Items marked as corrupted or conflicting are moved to the Sync Items folder, allowing synchronization to continue. Users see a

synchronization progress meter in the lower right corner of the Outlook 2003 client. The progress meter shows detailed synchronization information, such as the total amount of data left to synchronize, and whether the mailbox folders are up-to-date.

- **Outlook 2003 performance monitoring** Outlook 2003 collects performance information and sends it to Exchange 2003 so that an Information Technology administrator can identify performance bottlenecks relating to poor bandwidth conditions or poor connectivity. The performance data is stored in the Exchange store, as well as reported in the event log and through performance counters. To monitor response times on a continuous basis and to create performance reports over any period of time, you can use either non-Microsoft tools or Microsoft tools, such as Exchange 2003 Management Pack (Exchange Management Pack.akm) for Microsoft Operations Manager. Analyzing the performance data on a centrally located server can help an Information Technology department identify when local and remote clients are experiencing connectivity issues.

Server Consolidation Strategies

By using Exchange 2003, an organization can improve its level of service through increased system reliability. In addition, an organization can create a highly flexible infrastructure so that future growth can happen in a controlled way.

Organizations that deploy Exchange 2003 can achieve these objectives by using the following general strategies:

- **Consolidate multiple small servers on fewer large servers** You can use Exchange 2003 to accommodate more users per server than by using any previous version of Exchange, while maintaining usable backup windows. Exchange 2003 runs on Windows Server 2003 clusters with up to eight nodes. Backup agents can use the new Volume Shadow Copy service available in Windows Server 2003 to greatly reduce the time required for Exchange 2003 backup operations. Exchange 2003 also features a Recovery Storage Group and a Mailbox Recovery Center. You can use the Recovery Storage Group to restore individual mailboxes from a database backup quickly and conveniently. The Mailbox Recovery Center, on the other hand, provides bulk reconnection of mailboxes to the appropriate users in Active Directory, which supports disaster recovery scenarios.
- **Physically move servers from multiple geographic locations to one central location** This approach is often referred to as site consolidation and entails placing all servers in a data center. Outlook 2003 enables site consolidation in an Exchange 2003 organization. With over-the-wire MAPI compression and client-side caching in Exchange cached mode, Outlook 2003 is more resilient under unstable network conditions and requires less network bandwidth than any previous version of Outlook 2003. Site consolidation helps to lower administrative overhead, and makes it easier for information workers to access messaging resources when they work remotely.
- **Automate the maintenance of multiple servers using a reduced number of management interfaces** This approach is sometimes referred to as virtual consolidation, which means managing multiple servers as one virtual server system. Individual servers may reside in a data center or in different geographic locations. Integration with Active Directory and new administration tools, such as Exchange 2003 Management Pack (Exchange Management Pack.akm) for Microsoft Operations Manager, are key catalysts for virtual consolidation in Exchange 2003 organizations. For example, Exchange Management Pack.akm, included in Exchange 2003, enables organizations to monitor the performance, availability, and security of all Exchange servers in the organization from a single Microsoft Operations Manager console or Web page.

Making the Server Consolidation Decision

The decision to consolidate messaging systems should be based on an assessment of an organization's existing infrastructure. This assessment should include a graphical representation of the environment that illustrates the various regions that are administered separately, as well as the servers within these regions that store mailboxes, run messaging connectors, manage incoming browser connections, or perform other tasks. The assessment should also include a basic network analysis because network reliability and available bandwidth significantly influence the consolidation strategy.

Figure 2 illustrates a recommended approach to identifying consolidation opportunities.

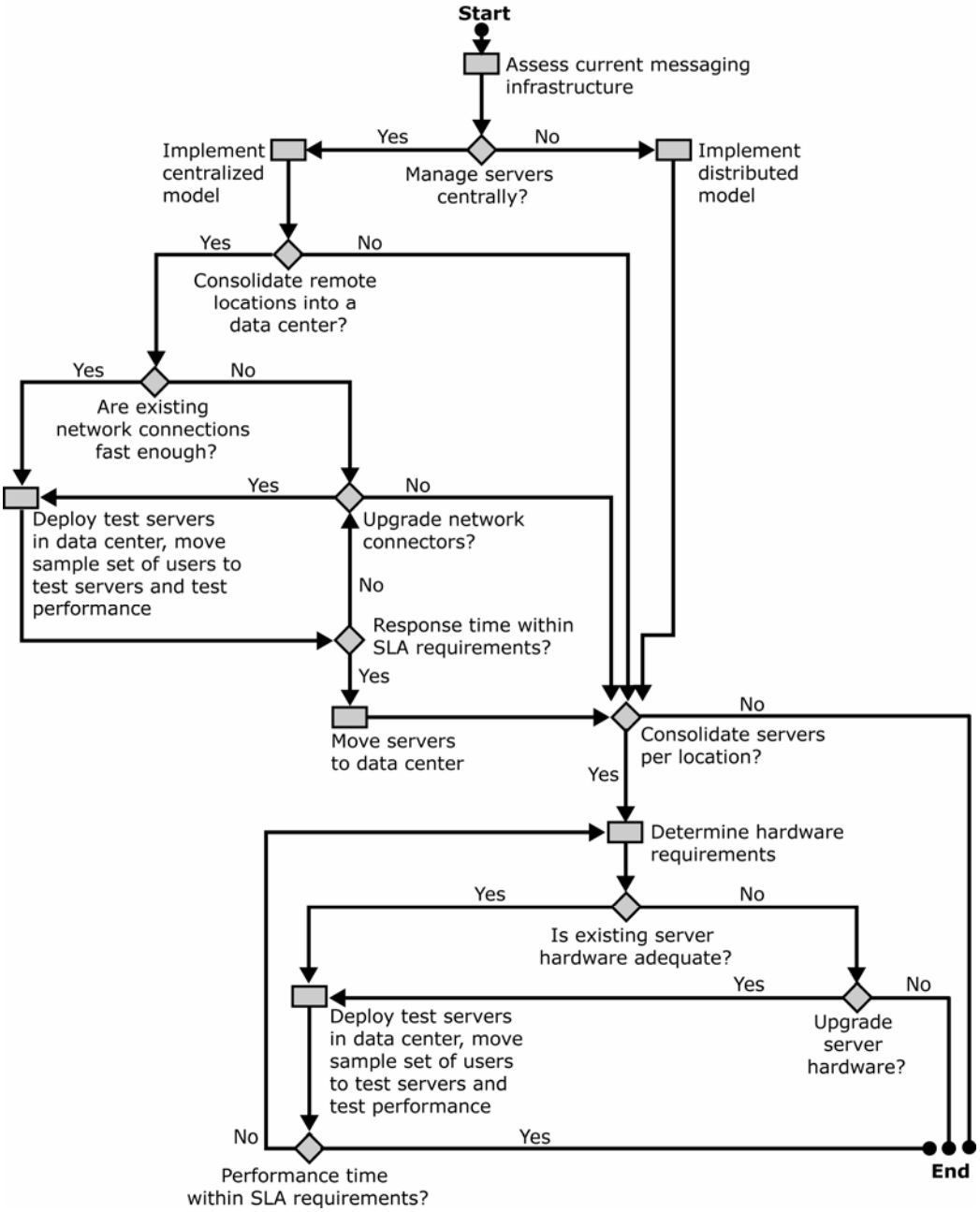

Figure 2 Determining server consolidation opportunities

You should address server consolidation strategy questions in the following order:

- **Is it possible to implement a central administration model?** The administration of Exchange 2003 is not bound to the physical network structure, so an organization can implement a centralized administration model for Exchange Server 2003—even across multiple geographical locations. A centralized administration model is better suited to server consolidation than is a decentralized model, because it is usually impossible to consolidate servers across regions that require independent administration. If you cannot implement a completely centralized administration, consider implementing a hybrid model in which a different Information Technology department is responsible for each individual geographical region in the organization. Each department is then centrally responsible for all those remote locations that belong to its assigned region. It might then be possible to consolidate servers into regional data centers.
- **Is it possible to consolidate remote locations into a data center?** Remote locations that fall under the same administrative authority are good candidates for site consolidation if sufficient network bandwidth is available to support client/server communication with acceptable response times. As mentioned previously, Outlook 2003 in Exchange cached mode can work over slow or unreliable network connections. However, you should conduct performance tests and pilot projects to determine whether the existing bandwidth is sufficient. An organization might have to upgrade network connections that cannot accommodate the workload, or refrain from eliminating mailbox servers in regional offices.

  > **Note**
  > To estimate bandwidth requirements, consider the number of users that connect to Exchange 2003 over the network link, the client type used to access mailboxes (for example, Outlook 2003 in Exchange cached mode or Microsoft Office Outlook Web Access), and user habits (for example, e-mail users who communicate frequently compared with e-mail users who communicate infrequently). You must also take into consideration processes that do not relate to Exchange-based client/server communication; for example, file and printer sharing or Active Directory replication.

- **Is it possible to consolidate servers within each location?** To identify servers that can be consolidated within a data center or geographical location, you must analyze the various tasks that the existing servers currently manage. Some server types are more eligible for an actual physical consolidation than others. For example, some servers might be responsible only for storing mailboxes (mailbox servers), while others might serve only the purpose of message transfer (bridgehead servers). You should consolidate mailbox servers, but it is not necessarily a good idea to consolidate bridgehead servers or Web servers onto mailbox servers in large organizations. Bridgehead servers and Web servers, due to their need for high network bandwidth, are more suited for server farms that are based on load-balancing technologies, such as Microsoft Application Center 2000.

  > **Note**
  > Server consolidation is primarily a way to concentrate user data onto a small number of servers by using a sophisticated storage subsystem. Servers that hold mailboxes and public folders are most suited for server consolidation.

- **Is it necessary to replace the existing server hardware in order to manage the expected workload?** After you identify the servers that can be consolidated, calculate the final number of users that each of these servers must support. Based on these figures, estimate each server's approximate workload to determine its hardware requirements.

  Consider whether the existing server hardware can accommodate future demands. Reusing existing hardware might help to reduce costs, but do not upgrade existing Pentium Pro or Pentium II multi-processor computers to Windows Server 2003 and Exchange 2003. Intel discovered that an upgrade to Windows Server 2003 on servers that have two or more Pentium Pro or Pentium II processors might cause instability, data corruption, or other unpredictable results. Upgrading these platforms will force Windows Server 2003 into single-processor mode, which is a supported configuration. However, a single processor that is operating at 200 MHz is not a good choice for server consolidation. When you consolidate onto fewer servers, consider replacing existing server hardware and storage technology. Table 1 lists basic processor and memory configuration recommendations for mailbox servers.

Table 1 Processor and memory configurations for mailbox servers

Number of Users	CPUs	Memory
Fewer than 500	1 - 2	512 MB – 1 GB
500 – 1,000	2 - 4	1 GB – 2 GB
1,000 – 2,500	4 - 8	2 GB – 4 GB
2,500 or more	4 - 8	4 GB

> **Note**
> The figures in Table 1 are recommendations for an optimal configuration. It is possible to support a greater number of users and at the same time, use less hardware. Microsoft, for example, is currently operating a number of four-processor Exchange 2003 servers that have 4,500 mailboxes each. Server performance can be improved by using additional processors, but the four-processor configuration works.

The following factors influence your Exchange server hardware design:

- **Connected users versus total users** When you design server hardware, estimate how many users will connect to the server concurrently. For example, performance requirements might be lower if not all users connect at the same time; for example, when users work in shifts. However, the server background activity increases as the number of mailboxes increases, even when users have not logged on to the system. For example, database defragmentation activity increases on a server storing thousands of mailboxes. Always consider background activities when you design server hardware.
- **Location and use of message stores** Exchange users can store their messages in server-based mailboxes or in personal folders (.pst files) on their local client computers or on a network drive. Users can reduce a server's storage and processing workload by setting their default delivery locations for messages to their personal folders. For example, when a user opens or saves an item in a personal folder, the Microsoft Exchange Information Store service on the server is not involved. However, when you decide whether to use personal folders in this way, consider the fact that storing all messages in personal folder stores on local client computers makes it impossible to back up messages centrally.

  > **Note**
  > To require users to store messages in .pst files on their client computers, configure mailbox quotas for the mailbox stores. If a mailbox exceeds its size limits, the user must reduce the amount of data by downloading messages to the .pst file or deleting them. Otherwise, the user cannot continue to receive and send e-mail messages.

- **User habits and messaging clients** When you design server hardware, consider the number of messages users send and receive per day. Also consider the types of clients that they use. Outlook 2003 and Outlook Web Access 2003 place different workloads on a server. If your users primarily use Outlook Web Access 2003, assume additional processing requirements for Internet Information Services (IIS).
- **Public folder usage and replication** Public folders can have a significant impact on server performance, depending on folder size, frequency of access, number of different views for each folder, number of replicas, replication schedule, and frequency of content changes. Place large public folder repositories on a dedicated Exchange server that does not store any user mailboxes. Exchange servers that store public folders are sometimes called public servers.
- **Connectors and gateways** Connectors and gateways are messaging components that can substantially increase a server's workload. Message transfer involves server-to-server communication and might require message conversion. As mentioned previously, do not configure connectors or gateways on large mailbox servers. An organization can benefit from servers that

are dedicated to message transfer. Another option is to migrate all messaging systems to Exchange 2003, thereby eliminating the need for messaging gateways.

- **Is the server hardware correctly designed according to service level agreements?** You can use Exchange 2003 stress and performance tools on a test server to verify whether its hardware is designed to accommodate the number of mailboxes you want to place on the server. For example, you can use Load Simulator (LoadSim) (LoadSim.exe) to test how an Exchange 2003 server responds to a large number of Outlook 2003 clients. Another useful tool is Jetstress (JetStress.exe), which you can use to simulate disk input/output (I/O) load to verify the performance and stability of your disk subsystem. You can download these tools from the "Exchange Server 2003 Tools" website (http://go.microsoft.com/fwlink/?LinkId=23490). To determine how many mailboxes Exchange 2003 servers of a given class can support, you might also want to read the performance benchmark studies at the "Performance and Scalability for Exchange Server 2003" website (http://go.microsoft.com/fwlink/?LinkId=23430). Another source of information on this topic is *Troubleshooting Exchange Server 2003 Performance* (http://go.microsoft.com/fwlink/?LinkId=23454).

Recommendations for Server Consolidation

When you identify server consolidation opportunities, it is a good idea to consider best practices that might apply to your specific situation. For example, it is highly recommended to use hardware from the Windows Server Catalog when you design servers running Exchange 2003. The Windows Server Catalog lists the hardware that Microsoft has certified for Windows Server 2003. The catalog is available at http://go.microsoft.com/fwlink/?linkid=17219. If you want to deploy multiple servers in a Windows cluster, ensure that your hardware is listed in the catalog's "Cluster Solutions" section.

> **Note**
> Microsoft supports hardware for Windows Server 2003 clusters only if it has passed the Windows Hardware Quality Labs (WHQL) test for the Windows Server 2003 family.

When you plan a server consolidation, keep the following recommendations in mind:

- **Design server hardware generously** Design your server hardware according to current and future requirements to prepare for future growth. You might want to consider additional processors, 2 GB or more of memory, and a reliable storage subsystem that has a capacity of at least two or three times the estimated size of your messaging databases. Note that hardware technology evolves at a rapid pace. Within a relatively short period of time, upgrade options might not be available for your server platform, which can pose a serious problem if future demands require you to increase system performance; for example, in the event that you need additional processors.

  > **Note**
  > To enable the Microsoft Exchange Information Store service to use up to 3 GB of virtual address space on a server that has more than 1 GB of physical memory, modify the Boot.ini file on your server and add the /3GB and /USERVA parameters to the startup line for the operating system. For more information about these parameters, see Microsoft Knowledge Base Article 316739, "How to Use the /USERVA Switch in the Boot.ini File to Tune /3GB Configurations" (http://go.microsoft.com/fwlink/?linkid=3052&kbid=316739).

- **Avoid single points of failure** By using fewer servers, you have fewer points of potential failure; however, the impact of failures increases. You can implement redundancy to address potential points of failure. If one component fails, another component can take over. Performance might be degraded temporarily during an outage, but users can continue their work.

  You should provide redundancies for the following components:

  - **General server components** For processor, memory, motherboard, and other components, consider implementing a Windows Clustering solution if high availability is a critical requirement for your mailbox servers. Exchange 2003 supports clusters running Windows Clustering. If one

out of a maximum of eight nodes fails, another node in the cluster assumes the processing until the original node is repaired.

- **Storage subsystem** Windows clusters do not provide redundancies for the storage subsystem. For this reason, it is important to provide redundancies within the storage subsystem itself; for example, you can implement a redundant array of independent disks (RAID). RAID configurations can also lead to an increased server performance when multiple disks perform I/O operations concurrently. Table 2 lists typical RAID configurations for Exchange 2003.

Table 2 Typical RAID configurations for Exchange 2003

RAID Technology	Description	Used For	Comments
RAID 0	An array of striped disks without parity.	Possibly client computers' disk drives or servers. Used in conjunction with RAID 1.	RAID 0 achieves very high performance because all disks can read or write data concurrently, but a significant disadvantage of this RAID configuration is lack of fault tolerance and a high risk of disk failure. If only one disk in the array breaks, all data is lost and must be restored from backup. For this reason, RAID 0 without RAID 1 is seldom used in server systems.
RAID 1	An array of mirrored disks.	Operating system, paging file, and transaction logs.	RAID 1 achieves very high performance and very high fault tolerance because all data is mirrored after it is written, and both disks contain the complete data. This configuration is a perfect choice for data that requires the highest reliability. Its most significant disadvantage is its high cost. RAID1 is the recommended technology for the transaction logs on mailbox and public folder servers.
RAID 0+1	A striped array of mirrored disks without parity.	Exchange databases that require high I/O performance and very high fault tolerance.	This configuration is a combination of the mirrored disks included in a RAID 0 drive. Each disk in the array is mirrored to guarantee a very high level of fault tolerance. This RAID configuration is becoming the configuration of choice for many organizations. For individual disk sizes larger than 18 GB, you should use RAID 0+1 instead of RAID 5. RAID 0+1 offers the highest performance and very high reliability.

RAID Technology	Description	Used For	Comments
RAID 5	An array of striped disks that have parity.	Exchange databases that require mid-range performance and fault tolerance.	RAID 5 works similarly to RAID 0, but includes a mechanism to write a checksum of the data on each stripe to one of the disks. If one disk in the array fails, the system can reconstruct the data from the remaining hard disks. Note that in a RAID 5 array, multiple small disks perform better than fewer large disks. For example, if you have three 27 GB disks, you can use only 54 GB of storage space. If you instead use nine 9 GB disks, you have 72 GB of storage space. However, the more disks you include in the array, the higher the chance that two disks will break at the same time. This requires you to re-create the RAID drive and restore the data from backup. It is safer to use fewer large disks instead of multiple small disks. As previously mentioned, RAID 0+1 is a recommended alternative to RAID 5 in cases where you have large disk capacities.

> **Note**
> Storage Area Networks are an alternative to traditional RAID configurations. A Storage Area Network is a high-speed special-purpose network that you can use to connect individual servers with external data storage devices. In fact, the sophisticated use of Storage Area Networks in mailbox server configurations is one of the main reasons that large organizations embrace server consolidation in Exchange 2003 deployment projects. Storage Area Networks have significant advantages over traditional RAID configurations, because they can increase reliability and fault tolerance, simplify backup and restore procedures, and make it easier to replace outdated server hardware with a new system. By integrating mailbox servers running Windows Server 2003 and Exchange 2003, organizations with Storage Area Network solutions can decrease the complexity of managing expanding volumes of messaging data.

- **Network infrastructure** Redundant networking equipment can provide load balancing of data traffic between physical network components, and can provide automatic failover in the event that a network component fails. Equip your servers with multiple network adapters, and provide redundant network connections to servers. Your network infrastructure design should also include redundancies for systems that provide important application-layer services, such as DNS servers, Active Directory domain controllers, and global catalog servers.

- **Implement a powerful backup solution** For servers that store a large amount of data, implement a high-performance backup solution that processes multiple backup operations concurrently. Current backup solutions can achieve data transfer rates of more than 100 GB per hour. If your backup solution is not as fast, consider performing concurrent operations to back up Exchange 2003 in a timely way. You can back up multiple storage groups at the same time in separate backup sessions.

  When you implement a backup solution for Exchange 2003, do not rely on the backup tool included in Windows Server 2003 because this basic tool is unable to use the Volume Shadow Copy service. Instead, evaluate non-Microsoft backup solutions that include an Exchange 2003 Volume Shadow Copy service requestor. The Volume Shadow Copy service uses requestors to create shadow copies of

Exchange 2003 databases. Based on these shadow copies, the Volume Shadow Copy service in Windows Server 2003 can greatly reduce the time that it takes to back up and restore Exchange 2003. Non-Microsoft backup solutions that are based on the Volume Shadow Copy service infrastructure can also use the shadow copies to almost instantaneously restore one or more databases.

The following sequential steps are performed when you back up Exchange 2003 by using the Volume Shadow Copy service:

1. The backup program starts a manual or scheduled backup process.
2. The Volume Shadow Copy service requestor in the backup program sends a command to the Volume Shadow Copy service to take a shadow copy of the selected Exchange 2003 storage groups.
3. The Volume Shadow Copy service communicates with a Volume Shadow Copy writer component in Exchange 2003 to pause new transactions, to finish current transactions, and to flush all cached data to disk.
4. The Volume Shadow Copy service communicates with the appropriate storage provider to create a shadow copy of storage volumes that contain the Exchange 2003 storage groups.
5. The Volume Shadow Copy service informs Exchange 2003 to resume normal operations.
6. The backup program copies the shadow copies of the storage group databases and logs to the tape backup device.
7. When the tape copy is completed, the Volume Shadow Copy service requestor in the backup program communicates with the Volume Shadow Copy service to delete the shadow copy.

Recommendations for Small Organizations

In the context of Exchange 2003, small organizations are defined as those that have fewer than 75 users, because organizations that have fewer than 75 users can obtain Exchange 2003 as part of Microsoft Windows Small Business Server 2003. Small Business Server 2003 is a complete network solution that you should evaluate if you are certain that your organization will not exceed the 75-user limit. Windows Small Business Server 2003 is available in two editions: Standard and Premium.

The Standard Edition includes the following components:

- **Windows Server 2003** This operating system includes the Active Directory directory service.
- **Microsoft Windows SharePoint™ Services** Use this component to implement team communications and collaboration.
- **Exchange Server 2003** Use this component to implement the messaging and collaboration infrastructure.
- **Outlook 2003** Use this component to manage e-mail, calendars, contacts, and other personal and team information.
- **Shared Fax Service** Use this component to receive faxes through Windows SharePoint Services, e-mail messages, or printers, and to send faxes through shared telephone lines.
- **Routing and Remote Access service** Use this component to implement dial-in and VPN connectivity.

In addition to the previous components, the Premium Edition includes the following:

- **ISA Server 2000** Use this component to implement firewall technology to secure Internet connections.
- **SQL Server 2000** Use this component to implement relational databases that support line-of-business applications.

- **Microsoft Office FrontPage® 2003** Use this component to provide tools for Web site development or to create customized solutions for Windows SharePoint Services.

Table 3 lists recommended and maximum possible hardware configurations for Small Business Server 2003.

Table 3 Recommended and maximum hardware configurations for Small Business Server 2003

Hardware	Recommended	Maximum
Processor	X86-based 550 MHz or higher	2 X86-based 2.4 GHz processors or higher (hyper threading is supported)
Memory	384 MB	4 GB
Hard Disk	6 GB plus space for data storage	2 hard disks in a mirror set (RAID 1) that have 6 GB, plus space for data storage
Network Adapters	1 or 2 depending on topology	2 redundant (teaming) network adapters
Other	Bootable CD or DVD drive Super VGA (SVGA) monitor, and a video adapter that has a 800 x 600 or higher resolution and a minimum of 256 colors Tape or other backup device Modem or fax device Uninterruptible power supply (UPS)	Same as for the recommended configuration

> **Note**
> Small Business Server 2003 is the only platform that can run Exchange 2003, SQL Server 2000, and Windows SharePoint Services on the same computer. For more information about Small Business Server 2003, see the "Microsoft Windows Small Business Server 2003" website (http://go.microsoft.com/fwlink/?LinkId=23456).

Recommendations for Medium Organizations

Organizations that have more than 75 users cannot deploy Small Business Server 2003. Instead, these organizations should run Exchange 2003 on a separate computer. Medium organizations might have up to 1,000 users in a central location, in which case a single Exchange 2003 server can host all mailboxes and public folders. A limited number of remote locations may exist that connect to the data center through virtual private network (VPN) connections over the Internet. Remote users may also use dial-up connections to the Internet to access their mailboxes through VPN connections or RPC over HTTP. Figure 3 illustrates a possible Exchange 2003 configuration for a medium organization that has 500 local and remote users.

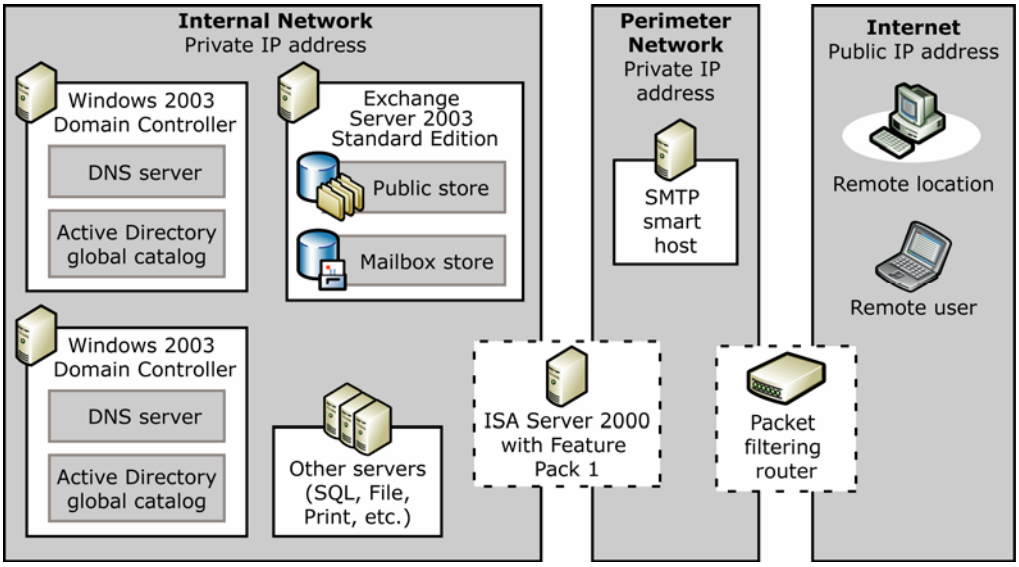

Figure 3 An Exchange 2003 design for a medium organization

Consider the following recommendations when you develop a server consolidation strategy for a medium organization:

- **Deploy two domain controllers running Windows Server 2003, both configured as DNS servers and global catalog servers in a single domain** Exchange 2003 requires a dependable DNS and Active Directory infrastructure, so it is important to deploy at least two global catalog servers. If you shut down one of the global catalog servers, the second global catalog server can continue to provide directory services to the Exchange 2003 organization, and users can continue to access their mailboxes, and send and receive messages.

  Configure both global catalogs as DNS servers to provide sufficient redundancy in the DNS infrastructure. DNS is an extremely critical network service; without it, Active Directory and the Exchange 2003 organization cannot function. Messaging clients will query DNS to locate internal resources, such as domain controllers and mailbox servers. Exchange 2003, on the other hand, relies on DNS to retrieve Internet Protocol (IP) addresses of domain controllers for directory lookups and external Simple Mail Transfer Protocol (SMTP) hosts when it sends messages to the Internet. Running DNS servers on domain controllers enables you to integrate DNS zones with Active Directory. An advantage of Active Directory-integrated zones is that all DNS servers have writable copies, and changes to zone information are replicated between the DNS servers as part of Active Directory replication.

  > Note
  > It is possible and supported to install Exchange 2003 directly on a domain controller. This might be an option for organizations that do not want to deploy two extra servers that run Windows Server 2003 and Active Directory. However, this configuration has performance limitations. When Exchange 2003 runs on a domain controller, it must always use the local domain controller for directory lookups and cannot perform load balancing between multiple domain controllers that might exist on the network. Deploy dedicated domain controllers in the forest of an Exchange 2003 organization, especially if additional servers that also require Active Directory access exist, for example, File, Print, or SQL servers.

- **Deploy Exchange Server 2003 Standard Edition** The Standard Edition of Exchange 2003 is designed to meet the messaging and collaboration needs of small and medium organizations. It costs approximately six times less than the base price of the Enterprise Edition. However, the Standard Edition does not support Windows cluster configurations, and supports only one mailbox database and one public folder database in a single storage group. The messaging databases are limited to 16 GB. If you find that the capabilities of the Standard Edition do not meet your service level requirements, you can deploy Exchange Server 2003 Enterprise Edition on Microsoft Windows 2000 Advanced Server, Microsoft Windows 2000 Datacenter, or Windows Server 2003 Enterprise Edition. While the

Enterprise Edition supports multiple databases per server without size restrictions and provides Windows Clustering support, keep in mind that its cost, combined with the cost of additional hardware might outweigh the benefits of server consolidation in a medium organization. For example, you might need to buy additional hardware to implement a two-node cluster in an active/passive configuration for the Exchange 2003 mailbox server.

In any case, design the storage subsystem with a RAID configuration to provide sufficient fault tolerance for the Exchange store. By default, all messaging databases and their transaction log files are in the \Program Files\Exchsrvr\Mdbdata directory. Exchange 2003 uses transaction logs to preserve transactions committed in the memory cache, which ensures that committed transactions are not lost if the server shuts down unexpectedly. By separating the transaction logs from the databases and placing them on a separate physical disk, you can increase the fault tolerance of the Exchange store. If the database hard disk breaks, you can replace the disk, create new databases, and then replay the transaction logs, which are still available because they reside on another disk. Your users can continue working as if nothing happened. If the disk on which the transaction logs reside fails, the databases are still available. Only the most recent transactions might be lost, if they were not incorporated into the databases before the problem occurred. Figure 4 illustrates a recommended hard disk configuration for a non-clustered server running Exchange Server 2003 Standard Edition.

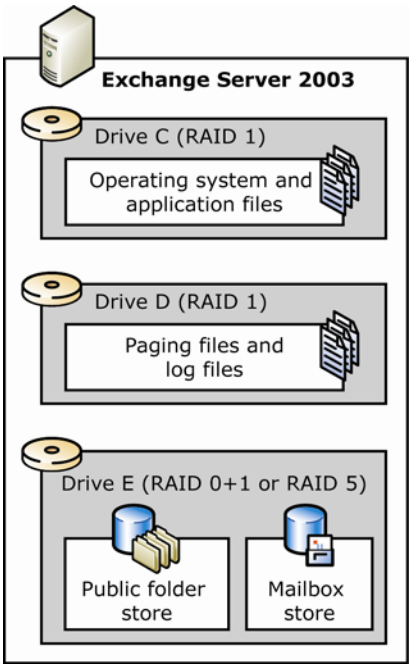

Figure 4 An Exchange 2003 configuration that has databases and transaction log files on separate disks

- **Deploy a perimeter network to protect internal resources from the Internet** You should separate the internal network from the Internet through an arrangement of dedicated, dissimilar firewalls. You should not use an Exchange 2003 server as a firewall. The outer firewall can be a packet-filtering router or a more sophisticated firewall system to provide controlled access to the resources on the perimeter network. If possible, enable network address translation on both the outer and inner firewall to hide actual IP addresses from potential attackers. On the inner firewall, consider deploying Microsoft Internet Security and Acceleration (ISA) Server 2000 because ISA Server and Exchange 2003 are designed to work closely together to provide a more secure messaging environment. If you deploy ISA Server 2000 with Feature Pack 1 on the inner firewall, you can protect the internal Exchange 2003 organization through Secure Sockets Layer (SSL) encryption, two-factor authentication, URL scanning, and SMTP filtering. At the same time, you can provide remote Internet users with VPN or RPC over HTTP connectivity and access to mailboxes through Post Office Protocol

version 3 (POP3) or IMAP4 (Internet Message Access Protocol version 4rev1), Microsoft Outlook Web Access 2003, Microsoft Outlook Mobile Access, and Exchange ActiveSync. For more information about configuring ISA Server 2000 for Exchange Server 2003, see *Using ISA Server 2000 with Exchange Server 2003* (http://go.microsoft.com/fwlink/?LinkId=23455).

> **Note**
> To achieve high security, use dissimilar firewall products on inner and outer firewalls, so that an attacker cannot use the same techniques on outer and inner firewall to penetrate the internal network. If you use ISA server on the inner firewall, use a non-Microsoft product on the outer firewall, or vice versa. Furthermore, do not run Exchange 2003 on a firewall because an Exchange server runs many services that an attacker might be able to exploit.

- **Consider deploying an SMTP smart host on the perimeter network** The Exchange 2003 environment in Figure 3 does not require an explicit messaging connector to send or receive messages over the Internet. If, through DNS, the Exchange server is able to resolve SMTP domain names into IP addresses of external SMTP hosts, the server can establish SMTP connections directly to send messages. You can configure ISA Server to accept incoming SMTP connections and forward accepted messages to Exchange 2003. However, many organizations prefer to deploy an additional SMTP smart host (or an arrangement of smart hosts) on the perimeter network to prevent any direct connections over the Internet to or from internal messaging systems. The SMTP smart host accepts messages from the Internet and relays them through the inner firewall to the Exchange server. In the opposite direction, the SMTP smart host accepts messages from Exchange 2003, queries DNS to find the destination SMTP host, and then sends the messages. You can configure the SMTP service in Exchange 2003 to forward all outbound messages to a smart host on the perimeter network. You can also use the standard SMTP service in Windows Server 2003 to implement a smart host.

  > **Note**
  > To increase the throughput and reliability of message transfer, consider deploying multiple SMTP smart hosts on the perimeter network. It is possible to register multiple mail exchanger (MX) resource records for an SMTP domain in DNS for load balancing and fault tolerance.

Recommendations for Large Organizations

The recommendations for medium organizations are also applicable to large organizations that have more than 1,000 users and possibly multiple server locations or even multiple messaging systems. For example, large organizations should ensure that multiple domain controllers and global catalog servers are available in each server location. However, you must consider several requirements that go beyond the needs of medium organizations. In locations that have many servers, you might deploy a multi-node cluster of Exchange mailbox servers in front of a Storage Area Network to support a very high number of users. You might also deploy dedicated bridgehead servers to optimize message transfer. Figure 5 shows an environment that has clustered mailbox servers and bridgehead servers in several locations.

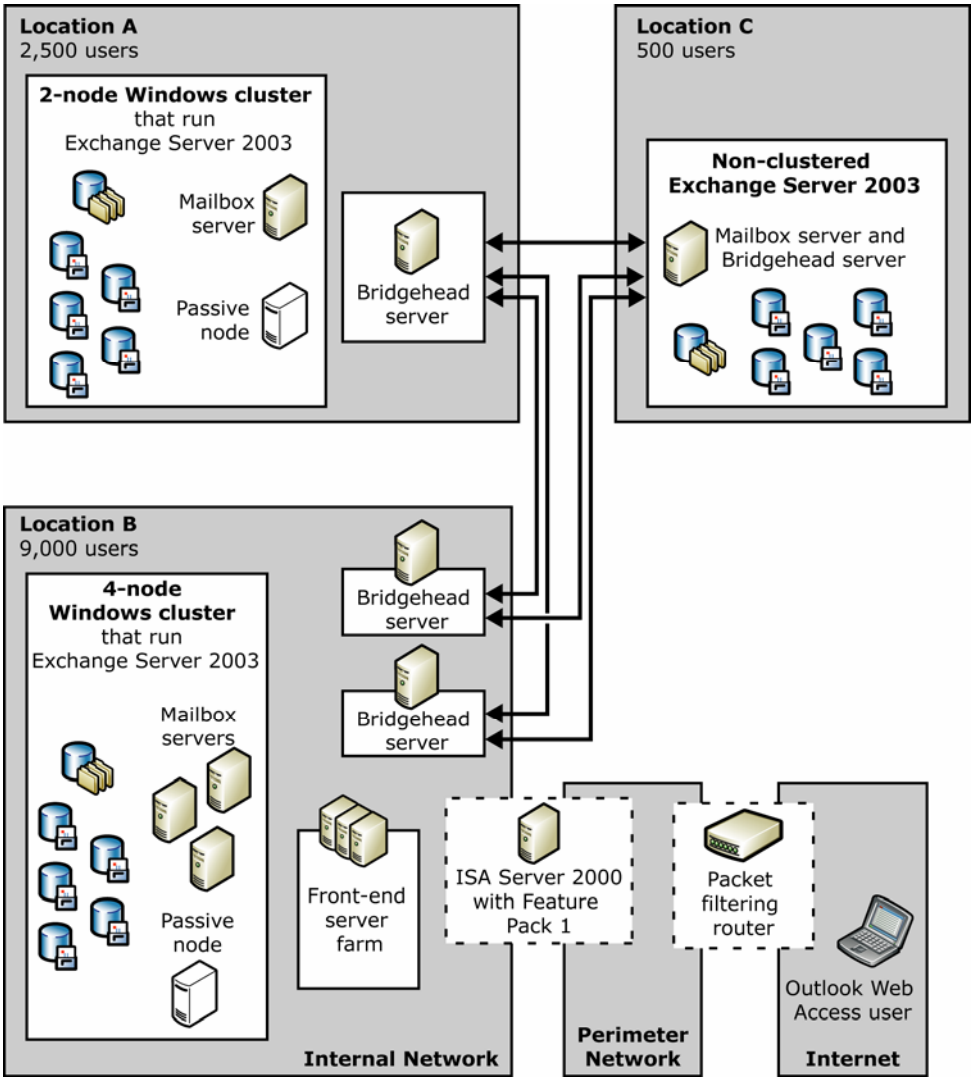

Figure 5 A large Exchange 2003 organization that has three server locations

Consider the following recommendations when you develop a server consolidation strategy for a large organization:

- **Deploy a redundant DNS and Active Directory infrastructure in all server locations** As mentioned in the previous section, your environment requires a dependable DNS and Active Directory infrastructure that provides reliable services to the Exchange 2003 organization. Most critical is the location of domain controllers, global catalog servers, and the directory replication topology. The Active Directory forest of an Exchange 2003 organization should have more than one domain controller in each domain, multiple global catalog servers in each server location, and a well-structured site topology for efficient directory replication.
- **Deploy dedicated mailbox and bridgehead servers in locations that have a large number of users** The sample environment illustrated in Figure 5 shows dedicated mailbox and bridgehead servers in locations A and B. Dedicated bridgehead servers between routing groups help to reduce processing requirements for message transfer on mailbox servers. Server location C does not have a bridgehead server because the mailbox server was deemed powerful enough to function as a bridgehead server, as well. For load balancing and increased fault tolerance, consider deploying multiple bridgehead servers.

  You may also make bridgehead servers responsible for distribution group expansion, which is a useful setting if you want to prevent groups that have thousands of members from expanding on mailbox

servers. If you specify an expansion server for a distribution group, all other Exchange servers in the organization must forward messages addressed to this mail-enabled group to its expansion server first. the expansion server then populates each message header with group membership information before transferring the message to its destinations.

- **Keep the number of routing groups to a minimum** Figure 5 shows a configuration of three routing groups that correspond to server locations. These routing groups are connected to each other through routing group connectors on bridgehead servers. However, if the network connections between the server locations are fast and reliable (for example, a VPN over fast Internet connection), you might be able to eliminate all bridgehead servers by placing all mailbox servers into a single routing group. Within a single routing group, message transfer is always direct from mailbox server to mailbox server. Routing group connectors or bridgehead servers are not required.

  The smaller the number of routing groups in an Exchange 2003 organization, the smaller the link state table that is used to replicate link state information between all Exchange 2003 servers. This means that less bandwidth is required to transfer link state information over the computer network. An Exchange 2003 organization can span up to 1,000 routing groups, but it is best to restrict the number of routing groups to fewer than 150.

- **Migrate the entire organization to Exchange Server 2003** Mixed environments that have multiple messaging systems require additional bridgehead servers and messaging connectors to connect the systems. The additional servers and connectors increase administrative overhead and network requirements. Consequently, if your organization uses Exchange Server 5.5 or non-Exchange messaging systems, you can streamline the messaging infrastructure by migrating all users to Exchange 2003. It might be possible to migrate users from different messaging systems to a single Exchange 2003 mailbox server.

- **Deploy front-end servers in locations that have multiple mailbox servers** If you have multiple mailbox servers and want to provide users with access to mailboxes through Outlook Web Access 2003, POP3, or IMAP4 over the Internet, you might find it helpful to deploy a farm of dedicated front-end servers. A front-end server is a computer that receives incoming client connections and proxies them to an appropriate mailbox server. Front-end servers determine the location of each resource in the organization by using Active Directory. In a location that has multiple mailbox servers, front-end servers provide a single point of access to all messaging resources, regardless of the actual mailbox server where the resources reside. By deploying multiple front-end servers in a server farm, you can provide load balancing and eliminate single points-of-failure for incoming client connections.

  Front-end servers can help to reduce the workload of mailbox servers. SSL encryption can be enabled on the front-end servers, offloading these processing requirements from the mailbox servers. You can place the front-end servers in the perimeter network. However, when you use ISA Server 2000 with Feature Pack 1, it is better to place the ISA servers in the perimeter network instead, and move the Exchange front-end servers to the internal network. This simplifies your configuration and gives front-end servers full access to Active Directory domain controllers. Plus, you benefit from the security features of ISA Server when you provide access to Outlook Web Access 2003 over the Internet.

  > **Note**
  > Front-end servers should not hold any mailboxes or public folders. In server locations that have only one Exchange server holding mailboxes and public folders, you can configure ISA Server to proxy incoming client connections directly to the mailbox server. In this scenario, a front-end server is not necessary.

- **Choose an active/passive configuration for Windows clusters running Exchange Server 2003** Windows clusters are most suited for very large mailbox servers that have a high availability requirement. A cluster is a group of servers connected to each other by means of a public and private network, as well as an external storage system. The physical servers can act as one or many virtual Exchange 2003 servers. A virtual server corresponds to a generic IP address and a network name, and owns a disk resource in the cluster. Any of the cluster nodes can then host the virtual servers, and users can access all of the resources in the cluster without having to know the actual name of the node that currently hosts the virtual server. When you configure a virtual Exchange 2003 server, place the

mailbox and public folder stores on the shared disk system. Only the Enterprise Edition of Exchange 2003 supports the Windows Clustering.

Figure 6 shows the main components of a multi-node cluster, which can support a very large number of users, for example, 9,000 users in Location B of Figure 5. By grouping two or more computers together in a cluster, you minimize system downtime caused by software, network, or hardware failures because another node can automatically take over a virtual Exchange 2003 server if the node that is currently executing the virtual server fails. This process is called a failover.

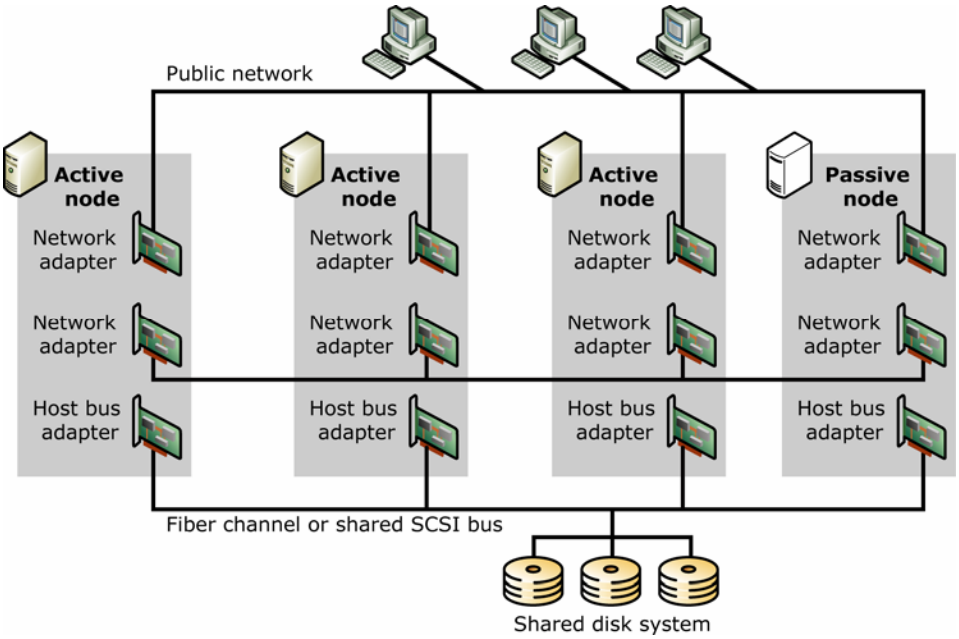

Figure 6 A multi-node active/passive cluster for three Exchange 2003 mailbox servers.

Failover can also be triggered manually. For example, you can intentionally fail over a virtual server to the passive node if the currently active node requires maintenance. Users are disconnected only for a brief period of time during the failover process. The hierarchy of service dependencies was flattened in Exchange 2003, so that virtual Exchange 2003 servers fail over significantly faster than virtual servers running a previous version of Exchange. Users might not notice the short downtime period if they are using Outlook 2003 in Exchange cached mode.

> **Note**
> Simplified hardware and software maintenance is one of the main reasons why organizations deploy Windows clusters. You can move the virtual servers to alternate nodes and then perform hardware or software upgrades on the original node, which is now passive. To upgrade hardware or software in this way is called a rolling upgrade.

You can configure multiple virtual servers on one cluster. However, consider several limitations when you design a server cluster. For example, a server cluster can have only one public folder store associated with the MAPI-based public folder hierarchy. Several components, such as Connector for Lotus Notes or Novell GroupWise, are not cluster-aware at all. Furthermore, a physical Exchange 2003 server cannot manage more than four storage groups. If you specify that virtual Exchange 2003 servers have multiple storage groups, ensure that a node does not have to run more than one virtual server at a time. Otherwise, you might create a situation in which a particular node must host more than four storage groups. You can address this issue in an active/passive cluster configuration by ensuring that the failover happens only to the passive node. It is possible to configure preferred nodes for failover.

You should leave at least one node free of a virtual server so that this passive node can take over a virtual server from a failing node without impacting performance. In an active/active configuration, one of the remaining nodes would have to assume the extra workload, in addition to its own virtual

servers, when a cluster node fails or is failed over manually. It is important to note that Microsoft does not support active/active configurations in cluster configurations that have more than two nodes. For more information about clustering with Windows Server 2003 and Exchange 2003, see *Using Clustering with Exchange 2003: An Example* (http://go.microsoft.com/fwlink/?LinkId=23460).

> **Note**
> Windows Clustering does not provide load balancing or storage redundancies. Virtual Exchange 2003 servers can be moved over from a failed node to another node in the cluster, but it is impossible to run the same virtual server on multiple nodes at the same time, or to specify that multiple virtual servers use replicated copies of the same messaging databases. Each virtual server must have an individual set of databases. Consequently, you must either configure the disk subsystem by using RAID 5 or RAID 0+1, or deploy a high-availability Storage Area Network solution.

- **Use a Storage Area Network to implement the storage system for multiple mailbox servers** Because Exchange 2003 supports Windows clusters that have up to eight nodes, a powerful and reliable shared disk system is required to implement the storage subsystem for all the messaging databases. If the number of users in a server location justifies the investment, consider deploying a Storage Area Network. Server clusters are a perfect choice for front-end Storage Area Network-based data storages. Storage Area Networks rely on Internet SCSI (iSCSI), Fibre Channel switching, or Gigabit Ethernet technology, which connects the storage systems, on which data is stored and protected, to the computer systems running Windows Clustering and the virtual Exchange 2003 servers. Fibre Channel switching or Gigabit Ethernet technology is fast and reliable, and allows hardware vendors to create storage solutions of up to several terabytes (TB). Complete Storage Area Network packages include the hardware, as well as the necessary storage management software. In a reliable Storage Area Network environment, multiple paths exist to the stored data, and the data can be backed up and restored efficiently.

  > **Note**
  > The storage industry is moving toward a seamless Storage Area Network provisioning model, where storage can be added at any time and reallocated to applications, such as Exchange 2003, as needed. Windows Server 2003 supports this model by providing support for emerging Storage Area Network industry standards like iSCSI, as well as established standards like Fibre Channel and Gigabit Ethernet.

- **Configure mailbox servers that have four storage groups and a large number of mailbox stores** Exchange 2003 supports up to four storage groups, each containing up to five messaging stores. Distribute the mailboxes across multiple mailbox stores to keep the size of individual databases manageable. Also, distribute the mailbox stores across multiple storage groups. Databases within a single storage group share a common set of transaction log files and must be backed up together. Separate storage groups use individual sets of transaction logs. They can be backed up in parallel and they support concurrent I/O operations, which improves performance. The Microsoft Exchange Information Store service operates by using multiple threads that can write transactions for separate transaction log files concurrently. For more information about designing storage groups and mailbox databases, see *Planning an Exchange Server 2003 Messaging System* (http://go.microsoft.com/fwlink/?LinkId=23461).

Conclusion

Exchange 2003 addresses the needs of small, medium, and large organizations, and server consolidation based on a deployment of Exchange 2003 can lead to significant cost savings in messaging operations. The decision to consolidate messaging servers should be based on an assessment of the current infrastructure. You must take an inventory of the existing server infrastructure, determine servers' capacity and workload, and document support and management processes. The assessment should also reveal the cost of managing and maintaining the existing infrastructure. Based on this information, you can identify server consolidation opportunities, business and technology priorities, and the required capacity for the consolidated environment. You are then prepared to make trade-offs between schedule, budget, priorities (including high

availability), and infrastructure flexibility. The new environment's design should be based on research on appropriate options for managing and maintaining a consolidated infrastructure in terms of capacity, fault tolerance requirements, and future growth.

By using a basic outline of the new environment, an organization can start developing actual deployment plans. This typically includes identifying the business impact of each of the consolidation alternatives, including an assessment of organizational roles and responsibilities, risks, budget, and desired results during and after the consolidation. Your planning process should include pilot projects that verify the network and infrastructure design, as well as hardware and software requirements. It should also acknowledge any technical limitations and risks. User and data migration plans should detail the procedures for moving users and data to the new environment, and include a detailed deployment schedule and contingency plan to address possible issues. When you implement the new consolidated production environment, do not forget to document the configuration of servers, as well as the post-consolidation management procedures. Before migrating users and data to the new environment, ensure that appropriate backups and contingency plans are in place. It is a good idea to perform the migration in small and manageable steps, adhering to a detailed schedule. You might view each step of the migration as a test of the consolidated environment. When the transition to the new environment is complete, evaluate the results of the consolidation project, including costs and maintenance procedures. A periodic re-evaluation of the consolidation can help to identify opportunities for optimization. Finally, remember that you cannot limit a comprehensive strategy to the process of consolidating an existing environment. Your strategy must also contain a set of standards and policies to maintain a consolidated environment.