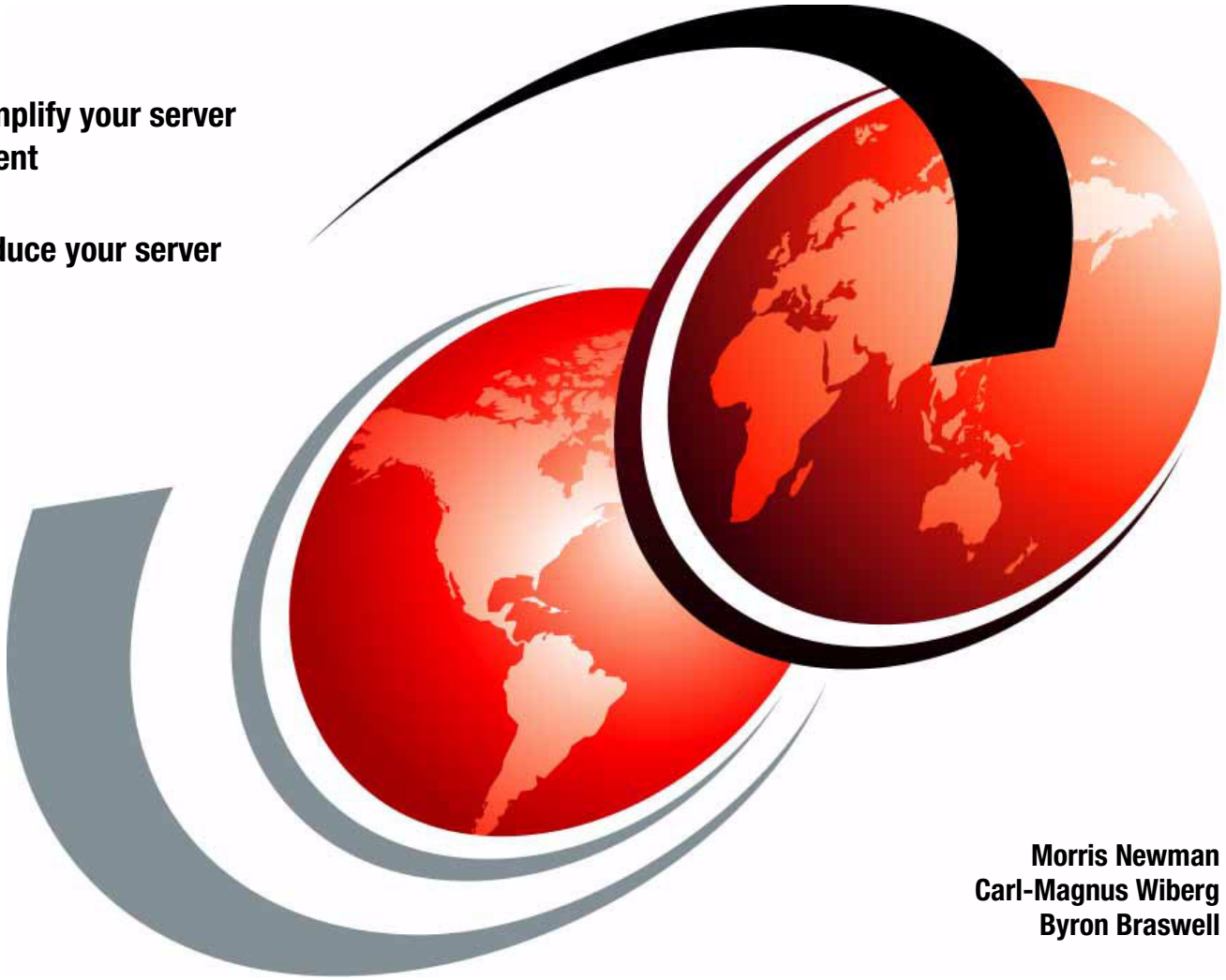
IBM

Server Consolidation with VMware ESX Server

- How to consolidate your production servers
- How to simplify your server management
- How to reduce your server costs



Morris Newman
Carl-Magnus Wiberg
Byron Braswell

Redpaper

ibm.com/redbooks

International Technical Support Organization

Server Consolidation with VMware ESX Server

January 2005

Note: Before using this information and the product it supports, read the information in “Notices” on page vii.

First Edition (January 2005)

This document created or updated on January 11, 2005.

Contents

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:
This information contains sample application programs in source language, which illustrates programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. You may copy, modify, and distribute these sample programs in any form without payment to IBM for the purposes of developing, using, marketing, or distributing application programs conforming to IBM's application programming interfaces.

Trademarks

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

AIX®
AIX 5L™
BladeCenter™
Chipkill™
DB2®
e-business on demand™
Electronic Service Agent™
@server®
@server®
eServer™
i5/OS™
IBM®
ibm.com®
IntelliStation®
iSeries™
Netfinity®
NetView®
NetVista™
Notes®
OnForever™
OS/2®
OS/400®
PowerPC®
Predictive Failure Analysis®
pSeries®
Redbooks (logo) ™
Redbooks™
RS/6000®
S/390®
ServerProven®
ServeRAID™
SurePOS™
ThinkCentre™
ThinkPad®
Tivoli®
Tivoli Enterprise™
TotalStorage®
Virtualization Engine™
WebSphere®
X-Architecture™
xSeries®

The following terms are trademarks of other companies:

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Intel, Intel Inside (logos), and Pentium are trademarks of Intel Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Other company, product, and service names may be trademarks or service marks of others.

Preface

Today, IT infrastructure organizations are working hard to solve the problems created by the explosion in the scope and complexity of IT platforms adopted in the 1990s. The migration of application architectures to thin-client multi-tier architectures, the rapid introduction of four generations and multiple editions of Microsoft® Windows® servers and the growth of Linux® have swept across IT organizations in successive waves over the last ten years. These waves caused explosive growth in server counts, network complexity and storage volumes throughout geographically distributed IT organizations. The policies and procedures adopted to gain back control of the infrastructure have often introduced their own challenges. Some of the resulting symptoms reported by IT organizations include:

- Large numbers of under-utilized one-application for one box Intel®-based servers
- Pervasive over-provisioning caused by policies that size all servers for worst-case workload scenarios
- Long provisioning cycle times for new servers, storage and networking
- Long delays between change request submissions and operational changes
- Narrow scheduled downtime windows over-subscribed with maintenance activities
- Inconsistent, irreproducible server builds due to the absence of build policies, or an inability to enforce them
- Patch roll outs that break application functionality or performance because the patch-testing systems do not match production systems
- Patchworks of management systems for Linux, Windows and NetWare servers.
- Incomplete information about equipment counts, status and ownership

IT has started to regain the upper hand in the battle against costly, inflexible and disorderly infrastructure. As a first step, IT organizations have generally centralized their IT infrastructure into fewer locations for better visibility. As a second step, they are adopting a new generation of infrastructure technologies and methodologies including server consolidation.

A traditional description of server consolidation is the process of merging the uses of two or more servers to a single server. However, server consolidation is more than simply replacing smaller servers with fewer bigger servers. It is about simplifying and optimizing existing end-to-end IT infrastructures, including servers, databases, applications, networks and systems management processes. The goal is to reduce cost and complexity and provide a stable foundation for growth and new solution deployment, while at the same time provide a higher service-level infrastructure that enables them to respond faster to business unit demands.

This IBM® Redpaper discusses server consolidation options and considerations using VMware ESX Server on IBM @server® BladeCenter™ and IBM eServer™ xSeries® 445 hardware. In addition, systems management options using VMware VirtualCenter, VMotion, VMware P2V Assistant, and IBM Director with VMM are also discussed.

The team that wrote this Redpaper

This Redpaper was produced by a team of specialists from around the world working at the International Technical Support Organization (ITSO), Raleigh Center.

The team Byron, Carl, Morris

Byron Braswell is a networking professional at the ITSO, Raleigh Center. He received a BS degree in Physics and an MS degree in Computer Sciences from Texas A&M University. He writes extensively in the areas of networking and middle ware software. Before joining the ITSO four years ago, Byron worked in IBM Learning Services Development in networking education development.

Carl-Magnus Wiberg is a Software Specialist in IBM ITS/TSS Sweden. He has 13 years of experience in the PC SW Technical Support field. His areas of expertise include VMware ESX Server, VMware Workstation, Virtual PC, DB2®, OS/2® and xSeries HW. He has been deeply involved in several ground-breaking projects in the virtualization areas. He conducts local entry-level workshops on VMware ESX Server upon request, as well as providing local L1/ L2 VMware ESX Server technical support.

Morris Newman is a Senior IT Specialist with the IBM Systems and Technology Group in Australia. He has extensive experience across many areas of IT, including 13 years of experience in the server consolidation field and four years with VMware products. In addition to being a VMware Certified Professional. He is a Microsoft Certified Systems Engineer and IBM Certified Professional, and has previously held Citrix and Microsoft Certified Trainer certifications. His areas of expertise include consolidation, server management and deployment, storage, xSeries, Microsoft technologies and Linux. He has been involved in several VMware ESX Server and Infrastructure OnDemand projects.

Thanks to the following people for their contributions to this project:

Margaret Ticknor
Denice Sharpe
Jeanne Tucker
Tamikia Barrow
Linda Robinson
David Watts
ITSO, Raleigh Center

David Hardman
IBM RTP

Alex Candelaria
IBM Seattle

Jan Strage
IBM Sweden

Massimo Re Ferre'
IBM Italy

Luis Aguilar Mateo
IBM Spain

Lesley Bain
Andreas Groth
Steve Weeks
IBM UK

Doug Pearson
Systems Group, IBM Bethesda

Joke Koning
IBM Netherlands

Anne Catambay
John Hawkins
Eric Horschman
VMware, Inc.

Become a published author

Join us for a two- to six-week residency program! Help write an IBM Redbook dealing with specific products or solutions, while getting hands-on experience with leading-edge technologies. You'll team with IBM technical professionals, Business Partners and clients.

Your efforts will help increase product acceptance and customer satisfaction. As a bonus, you'll develop a network of contacts in IBM development labs, and increase your productivity and marketability.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our papers to be as helpful as possible. Send us your comments about this Redpaper or other Redbooks™ in one of the following ways:

- Use the online **Contact us** review redbook form found at:

  ibm.com/redbooks

- Send your comments in an email to:

  redbook@us.ibm.com

- Mail your comments to:

IBM Corporation, International Technical Support Organization
Dept. HZ8 Building 662
P.O. Box 12195
Research Triangle Park, NC 27709-2195

1

Server Consolidation

In this chapter, we discuss server consolidation as a solution to server *sprawl*, which is the result of companies buying servers to solve immediate local server needs without considering long-term implications. We also discuss the benefits of consolidation, different consolidation solutions, proper candidates for consolidation, and the criteria to consider before and during consolidation.

Topics covered in this chapter are:

1.1 Server sprawl

Starting in the late 1980s, distributed computing and client/server architecture were the guiding principles. As businesses grew and their IT demands increased, they had to acquire and deploy solutions fast while keeping costs low. Distributed servers were cheap to buy and deploy for the solutions needed. But as organizations deployed single servers for individual solutions, the cost of maintaining these systems multiplied.

Along with the application needs, storage needs also rose. Departments and IT groups connected direct-attached storage devices to isolated servers across the enterprise. This resulted in inefficient use of storage, including duplication of data stored across the enterprise. Data that was vital to the businesses was distributed and duplicated in storage devices in departments across the enterprise without much security and was hard to access when needed, which often resulted in business failures or customer dissatisfaction. Duplication also meant that data integrity was not guaranteed.

The departments administering these systems were spending more time performing system administration and support than doing their real job, which, in turn, affected business growth.

Departments that wanted autonomy were happy with this solution, since they were in full control of their IT resources. IT departments bought servers from varying vendors with different operating systems, which sometime meant they had to hire new IT staff to support those specific servers and operating systems.

Over time, IT infrastructures became complicated, distributed, harder to manage, and harder to support, resulting what the industry calls server sprawl. Server sprawl resulted in inefficient use of floor space, and, with the increased cooling requirements, also increased the operating costs.

These servers were not optimized for running different applications or multiple copies of the same application running on an operating system instance. IT departments typically deployed one application for each operating system instance. As a result, most of the servers deployed in the field had a utilization of 10-35%.

In addition to this, mergers and acquisitions added to the server sprawl. IT departments now had to manage two or more such server sprawls and provide efficient service to the user and customer.

Deploying a simple solution across the enterprise, such as a security or anti-virus patch, became a nightmare. Servers became targets when these were not deployed at appropriate times, which resulted in lost data or service to customers. Backups were done in a random and inconsistent manner since the servers were totally distributed. This resulted in loss of data and time.

Enterprises that experienced server sprawl had no options but to add new servers to support business growth, since their servers are not capable of accommodating the growth.

1.1.1 The solution: server consolidation

Server consolidation, the process of merging the uses of two or more servers to a single server, became the dominant worldwide trend in IT deployment in the late 1990s to solve the problem of server sprawl.

Server consolidation can more efficiently:

- Use IT resources
- Use staff

- Use floor space
- Increase the security of data
- Provide easier management
- Remove hidden costs
- Reduce operational costs
- Add flexibility for future growth

Companies are frequently looking to cut costs, and IT budgets are scrutinized more than ever. IT spending is often limited to projects that are sure to save money and improve business values. IT organizations have realized that distributed computing is costly and carries with it a lot of hidden costs. These organizations have now moved towards a more centralized environment.

1.2 What is server or IT consolidation?

Server or IT consolidation is more than simply replacing smaller servers with fewer bigger servers. It is about simplifying and optimizing existing end-to-end IT infrastructures, including servers, databases, applications, networks and systems management processes, with the goal of reducing cost and complexity and providing a rational, stable foundation for growth and new solution deployment.

Server or IT consolidation facilitates the evolution to an on-demand environment by providing a framework for assessing and transforming IT infrastructure to enable the new IT business model.

Improved availability, cost management, disaster recovery, and security are some of the top cited reasons for consolidation given recent events and the current economic environment.

Figure 1-1 shows a phased approach to consolidation. The easier forms of consolidation represent a significant portion of the total cost of ownership (TCO) savings.

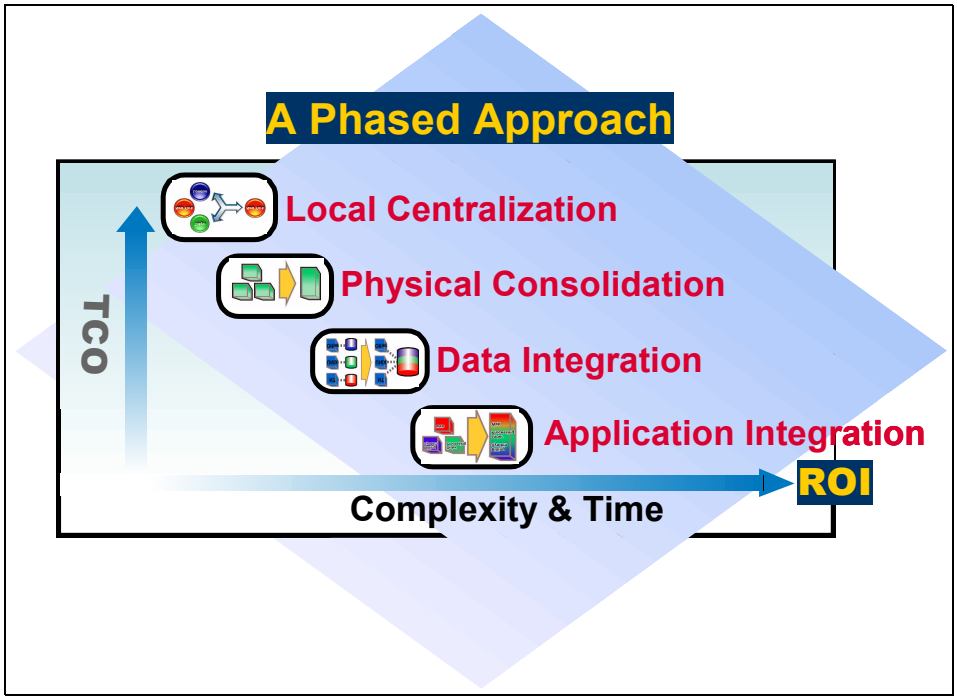


Figure 1-1 A phased approach to consolidation

1.3 The benefits of consolidation

Server consolidation offers the following benefits:

- Reduced total cost of ownership
- Improved efficiency
- Improved service levels
- Infrastructure for growth
- Single point of control

In this section we discuss the benefits of server consolidation in detail.

1.3.1 Reduced total cost of ownership

There are several costs involved in deploying and maintaining an IT infrastructure. This is the most important reason why organizations consolidate.

- Hardware costs

  Server consolidation reduces hardware costs due to the following reasons:

  - Fewer servers
  - Fewer storage devices because duplication is avoided
  - Fewer peripherals to perform backup

  For example, if you are planning to implement a storage area network (SAN) configuration. If, you have 100 servers that you need to connect to a SAN, the Fibre Channel host adapters, switches and fiber cabling alone will be a significant cost. By consolidating to fewer servers, you reduce the cost of that infrastructure considerably.

- Software costs

  Software costs may or may not reduce with consolidation. Software licensing costs may be reduced by consolidating workloads from applications running on multiple servers to an application running on a single server.

  Licensing costs may not reduce proportionately especially when licenses are based on workload. If an application has a licensing cost for each installation, then consolidating multiple operating system instances to a single operating system instance would reduce costs.

  Note that 1:1 consolidation of physical servers onto a server that has been partitioned into multiple virtual machines may not reduce licensing costs. However in some cases, a vendor will license software to a physical server regardless of how many virtual machines are deployed on that server.

- IT support staff

  This is one of the major areas where costs are reduced through the use of server consolidation. The size of IT support staff will most likely be reduced.

  Consolidation allows uniform and consistent system administration, which should reduce unnecessary downtime and associated overhead. The staff will be able to use their time more efficiently.

- Operational costs

  Operational costs involve the cost of floor space, power running the systems, cooling the lab space, and so on. With consolidation, there are fewer servers, which, in turn, reduces all of the operational costs.

- Hidden costs

Hidden costs are those costs incurred by wasting resources. Consolidation helps organizations reduce costs by providing the following advantages:

- Consistent hardware which means no retraining staff
- Consistent backup and recovery
- Reliable servers, improved productivity of staff
- Easy to audit, since there are fewer consolidated servers
- Secure, consistent data
- Failures avoided sue to stable consistent infrastructure

1.3.2 Improved efficiency

As we discussed in 1.1, “Server sprawl” on page 2, the average use of a server’s capacity is only 10-35%. With consolidation, server usage can be increased to an optimum level without increasing user response time.

With server consolidation, a centralized staff of IT professionals, not the users, perform the administration of the servers. Server consolidation provides a stable, robust, simple and consistent infrastructure with uniform system administration tools. Because the newer generation of hardware is more stable and robust, it reduces hardware failures. Because the newer hardware is simple, consistent, and has uniform system administration tools, it reduces human error. Both of these, in turn, increase the efficiency of the user.

1.3.3 Improved service levels

With a consolidated infrastructure, it is much easier to implement 24x7 support. Response time is much better than in a distributed environment, and the data is more easily accessible, while being highly protected. The control procedures are simpler. Data sharing is improved, giving users increased data consistency.

The latest servers with new technology provide a higher level of stability, redundancy, and dependability, which increases the level of service.

A consolidated infrastructure provides consistent backup and recovery procedures, which improve the probability of getting the data recovered in case of failure.

1.3.4 Infrastructure for growth

Consolidation, together with the latest scalable technologies, offers organizations the flexibility to increase the capabilities of servers. It is easy to add more processing power, more memory, or more input and output (I/O) bandwidth.

Deploying a Redundant Array of Independent Disks (RAID) for reliable data storage is much easier and cheaper in a consolidated environment than in a distributed environment.

High-availability solutions are also easier and cheaper to deploy in a consolidated environment.

1.3.5 Single point of control

Server consolidation by definition provides a single point of control.

With a single point of control, there are inherent standards and discipline that improve the service level to users. Rapidly growing organizations, especially those growing through mergers and acquisitions, frequently see that disparate, distributed systems are so hard to manage that they can lose control, constraining further corporate growth.

A single point of control allows enterprises to:

- Reduce or eliminate department operational costs
- Reduce number of systems, disk storage costs
- Reduce owner operational costs
- Offer better availability of service
- Improve systems management
- Have better version control management
- Have better software distribution
- Reduce risk and increase security

1.3.6 Minimize learning and optimize use of skilled resources

With distributed computing, departments had their own combination of servers, storage solutions, operating systems and applications. The knowledge developed by the IT staff in one department might not have been of use to other departments.

With server consolidation, a full-time administrator can be justified, freeing users from administering the systems.

1.4 Types of consolidation

As shown in Figure 1-1 on page 3, server consolidation can be placed into the following four phases listed in order of increasing difficulty:

- Centralization, see 1.4.1, “Centralization” on page 6
- Physical consolidation, see 1.4.2, “Physical consolidation” on page 7
- Data integration, see 1.4.3, “Data integration” on page 9
- Application integration, see 1.4.4, “Application integration” on page 10

The implementation of consolidation becomes difficult from centralization to application integration.

1.4.1 Centralization

The simplest form of consolidation is centralization in which servers are moved to a common location. The move may be physical or virtual. Centralization provides very few advantages of server consolidation, but is a first step towards future consolidation efforts. This makes future growth easier and more measurable. It also makes future high availability a feasible solution.

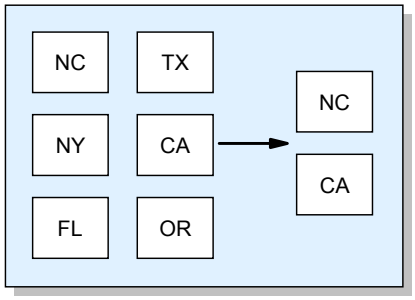


Figure 1-2 Centralization

Physical centralization

Servers and storage from many sites are physically relocated to fewer sites. All servers are brought under a common set of administrators, tools, and so on. This has the following advantages:

- Reduced number of IT staff
- Improved security
- Uniform system management
- Clear understanding of maintenance costs
- Consistent level of service
- Reduced downtime

There are several key considerations for physical centralization:

- What applications and services are used on the servers to be centralized?
- What is the architecture of the applications and services?
- Is the network communication requirement between the client PC and the server considered to be thin or fat?
- Is the wide area network (WAN) link between the client PC and centralized location able to deliver a reliable and responsive service?

If the server being centralized hosts a client/server application, and the network traffic, bandwidth, requirements exceed the capability of the WAN, you need to consider an upgrade to the connection speed or the implementation of a thin client solution.

Thin client solutions represent a consolidation of client systems, and is outside the scope of this document. More information is available on this topic in the following Redpapers:

- *Implementing Windows Terminal Server and Citrix MetaFrame on IBM @server xSeries Servers*, REDP-3629-00
- *Deploying Citrix MetaFrame on IBM @server BladeCenter with IBM FAStT Storage*, REDP-3583-02

Virtual centralization

In this form, servers in multiple sites are logically connected through a network and controlled remotely. Hardware remains in the same location but the servers are brought under a consistent set of system management tools. The advantages of this type of consolidation are:

- Reduced number of IT staff
- Uniform system management
- Consistent level of service

Although this form of consolidation is the simplest, the results could be disastrous if it is done without proper analysis and planning.

1.4.2 Physical consolidation

Physical consolidation is the process of replacing a number of smaller servers with larger servers of the same processor architecture. In many situations, centralization is a prerequisite for physical consolidation. The number of servers that can be consolidated to a single server depends on:

- Capability of the old servers
- Usage of old servers
- Capability of the new servers

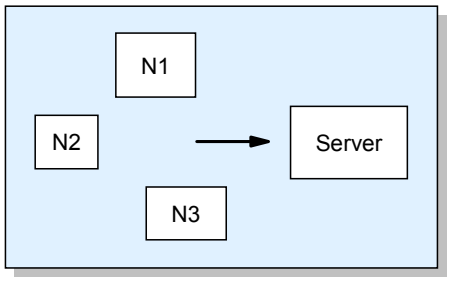


Figure 1-3 Physical consolidation

Physical consolidation in Figure 1-3 can be implemented at the departmental level, at the site level, or at the enterprise level. The applications that run on the old servers should be able to run on new servers without any changes. However, there are exceptions where applications that can only run under legacy operating systems, such as Microsoft Windows NT® 4.0, may not run on newer hardware.

Advantages of physical consolidation are:

- Reliable servers
- Consistent platform
- Standardization
- Less power and space usage
- Easy and uniform administration
- More secure
- Fewer peripherals
- Higher service levels

Virtual machines

The virtual machine (VM) is a concept pioneered by IBM for use initially with mainframes. In this concept, a virtualization layer sits on the hardware and a number of independent isolated virtual machines are created on top of this layer.

Resources such as CPU capacity, memory, disk space, and so on, can be allocated to VMs. Operating systems can be installed on the virtual machines to use and manage the resources allocated to it. Each VM is totally isolated from other instances of operating systems running on different VMs on the same hardware, as in Figure 1-4.

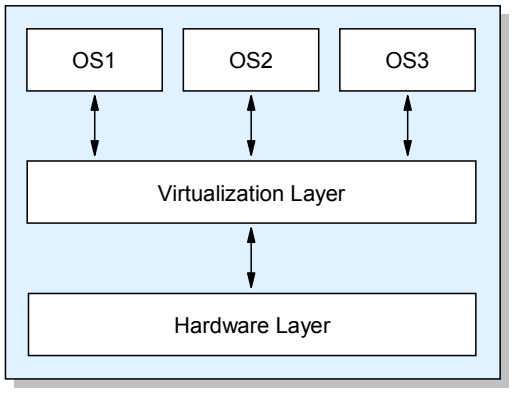


Figure 1-4 Virtual machines

This concept fits well in physical consolidation, especially for organizations that have applications with such constraints as one application for each operating system instance, or one server for each department.

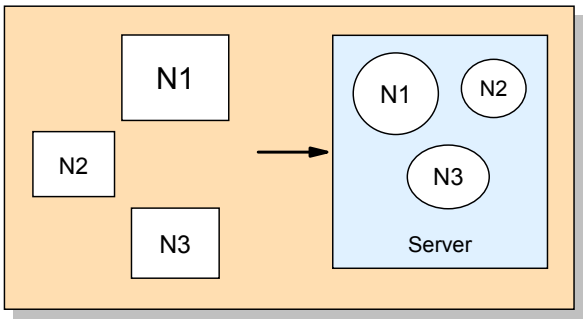


Figure 1-5 Physical consolidation with virtual machines

In Figure 1-5, all the applications and the operating system instances can be moved to virtual machines running on a server. The operating system instances and the applications are well-isolated from each other. An application or system crash in one VM does not affect the others.

With this form of consolidation, you can support legacy operating systems and applications while moving to newer hardware technology.

Another big advantage of virtualization is that many different operating system environments can coexist on and share the resources of a single system. Migration of applications running in heterogeneous operating system environments and on much under-utilized hardware servers becomes simple. One must be aware of virtualization overhead to accurately configure the virtualized solution.

Table 1-1 shows the possible benefits of virtual infrastructure.

Table 1-1 Benefits of virtual infrastructure

Strategic	Economic	Operational
IT responds faster to business demands	Costs are lower and easier to manage	It is the most flexible way to build IT
► Integrate more easily with partners and customers ► Expand faster ► Deploy and move resources more quickly ► Pay as you grow ► Create a foundation for utility computing	► Do not pay for what you do not need ► Get more out of your IT investment ► Reduce TCO by up to 60% ► Return on investment (ROI) in six months	► Leverage technologies you already own ► Increase quality and consistency ► Obtain lower-cost platform options ► Minimize technology risk ► Gain instant provisioning and deployment ► Increase capacity without more hardware

1.4.3 Data integration

Data integration is the process of taking information from several disparate sources and merging it into a single repository and a common format. Each data element is made to follow the same business logic, or, by deploying the shared storage subsystem, to manage disk requirements in a heterogeneous environment. See Figure 1-6 on page 10.

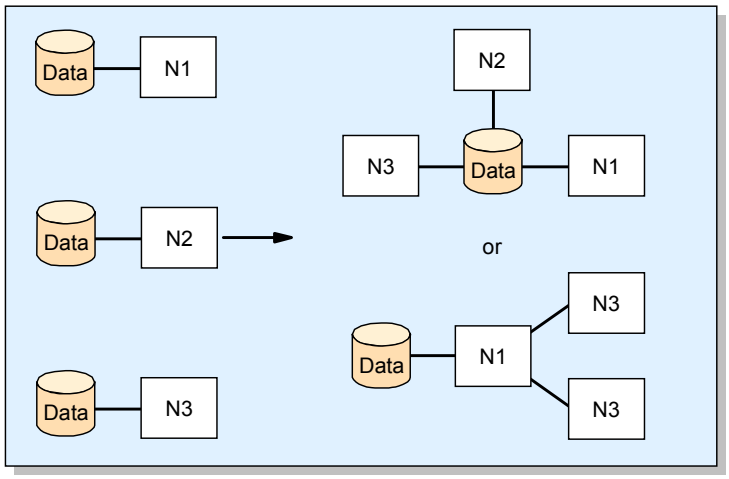


Figure 1-6 Data integration

Enterprise servers are a logical place for data that is now scattered throughout the enterprise. When all corporate data resides on the same robust system, the efficiencies can deliver immediate payback to users. Data sharing throughout the enterprise is vastly simplified.

Consolidation allows high levels of security and data integrity that are nearly impossible to achieve in a distributed environment.

There are two kinds of data integration:

- Data from several servers is consolidated into a unique repository
- Data from several repositories in one server is consolidated into a unique repository

Depending on the type of application integration selected, data integration can be performed separately or together with application integration.

1.4.4 Application integration

Application integration is the combining of multiple, similar applications, such as Web servers, onto one consolidated server.

It is also the process of migrating an application to a larger system in order to support data integration, for example migrating four Microsoft Exchange servers that support 100 users to a single server which supports 500 users. See Figure 1-7 on page 11.

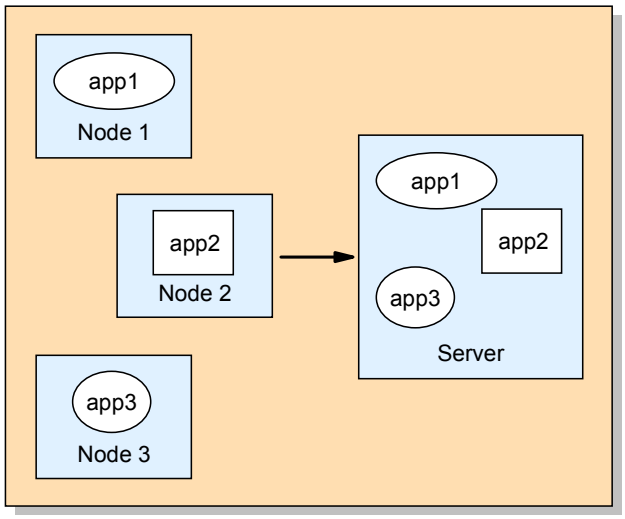


Figure 1-7 Application integration

Application integration can also be used to collocate the data and application. The level of difficulty depends on the amount of work needed to run the application in the new server:

- The application runs as is
- The application needs to be recompiled
- The application needs to be ported
- The application needs to be redesigned

1.5 Consolidation methodologies

IBM is a leader in server consolidation with its proven methodologies, patented tools and breadth of offerings, skills, and services. IBM leadership in this market helps clients optimize and simplify their existing IT infrastructure and provide a foundation for new solution investment and implementation that supports the client's business goals and objectives.

IBM's solution offerings distinguish themselves by focusing on how-to skills and the experience to implement these projects successfully. IBM offers an end-to-end solution that incorporates server, storage and network components, uses the appropriate offerings from IBM Global Services, while IBM Global Financing ensures the whole package is affordable, and leverages IBM Business Partner skills.

1.5.1 SCON SWAT team

To help the successful implementation of xSeries server consolidation solutions, IBM's Integrated Technology Services (ITS) group has a server consolidation (SCON) SWAT team. The SCON SWAT team consists of a team of experts that can quickly analyze the client environment and develop a SCON solution proposal.

The SWAT team's primary focus is on five application types:

- Collaboration
- Database
- File/print
- Thin client/Citrix
- Application servers

The majority of solutions involve the use of VMware virtualization products such as VMware ESX Server.

The SWAT team also makes use of the *consolidation discovery and analysis tool* (CDAT) to gather data, often the most difficult part of a server consolidation process. CDAT connects to the client's intranet and identifies all active Intel servers. It then reports, among other things, processor generation and server size, and processor and memory utilization rates (average and peak).

The SWAT team uses one of several tools to collate and analyze the server, business, and requirements data. These include the Align tools, Mini-Align, and VISIAN.

The SWAT process is both quick and very cost effective. A consolidation proposal can be provided in three weeks. The IBM internal use home page for the SCON SWAT team is:

`http://w3.beaverton.ibm.com/groups/SCON_SWAT/`

Business Partner support and forums can be found at the following URL:

`http://www.ibmsconbpsupport.com`

VISIAN is a comprehensive tool to help with sizing a VMware consolidation effort. It is available for IBM internal use at the following URL:

`http://ehngsa.ibm.com/home/a/g/aguilues/web/`

1.5.2 Scorpion

Scorpion is an IBM method that analyzes the IT infrastructure, combining techniques of server infrastructure technical analysis with financial analysis, capital budgeting principles and tools, and new technology solutions.

Today, IT organization has become a critical piece for improving shareholder value within corporations. It has become more and more central to the job of the IT organization to contribute to profit growth, not only by cost reduction but also by enabling new revenue opportunities through new channels such as the Internet, and new services such as ERP.

All of this activity is aimed typically at delivering a new competitive advantage to the organization quickly and effectively. As the market demands a new age of responsiveness, Scorpion outlines solutions with improved time to market abilities.

These changing realities put huge pressures on the standard ways of understanding and solving business problems. A Scorpion study helps you address these changing requirements.

Project methodology

A Scorpion study is delivered typically as a comprehensive six to eight week project conducted by experienced IT infrastructure consultants from the IBM @server organization. This team helps link IT value with IT infrastructure by examining the current and potential alternative IT server end states for an organization and applying a realistic financial analysis to build the investment case and to move to these future end states.

An *end state* is a Scorpion study infrastructure model of a possible future scenario.

In order to analyze the current server state, a number of techniques are employed using current metrics, such as people efficiency, server efficiency, quality of service, and total cost of service delivery. Incremental and future cost analysis provides an insight into future state costs, and emphasizes the importance of the utilization and server performance metrics.

The biggest breakthrough is in getting to credible outline solution designs for tactical and strategic projects that can be used as the baseline for business cases that compare multiple target server scenarios.

The Scorpion method uses two approaches:

- A full study is appropriate for analyzing the whole server infrastructure.
- A quick assessment is very effective where a focused analysis is required of a particular subset of servers.

Generally, a quick assessment is a precursor to the full study. The findings of these projects enable IT executives to balance alternative technical options and make sustainable server platform and software component decisions.

Scope and steps

As mentioned previously, the overall objective of a Scorpion project can be summarized as:

- Defining the current state of an IT server infrastructure
- Describing alternative realistic future states
- Showing cost projections of the alternatives
- Recommending a better future state alternative

Figure 1-8 gives a summary of the Scorpion methodology and architecture.

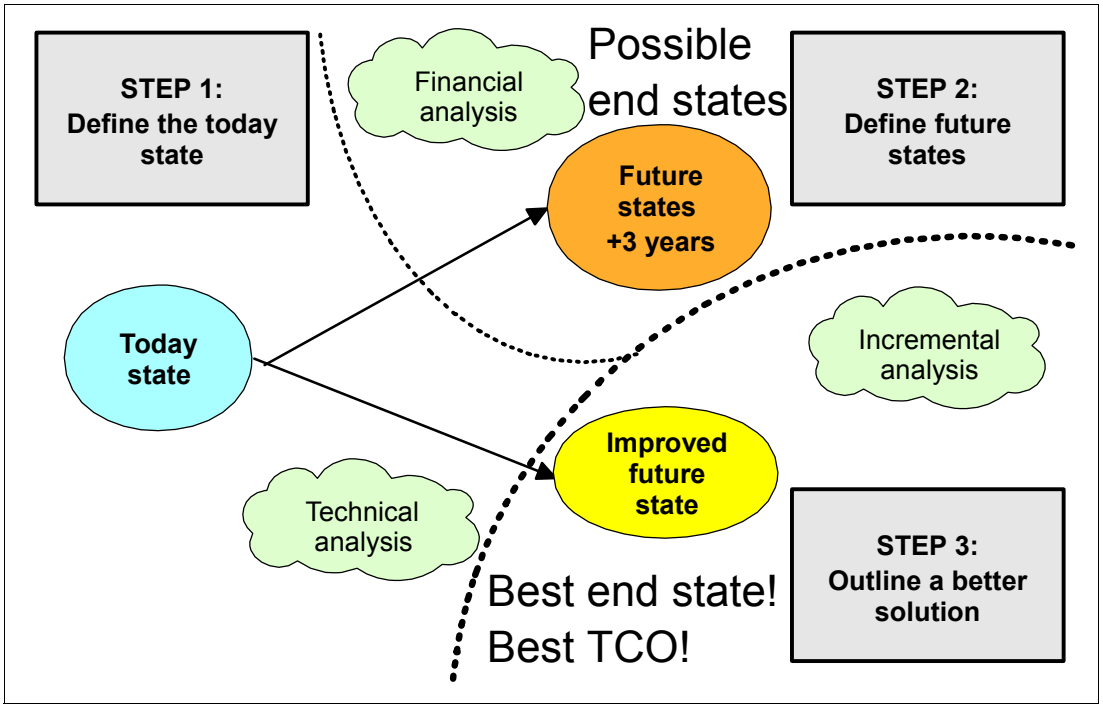


Figure 1-8 Summary of Scorpion methodology and architecture

The tasks performed during a typical project vary, but can be summarized as follows:

1. Define the study scope with the IT executive sponsor.
2. Issue data collection tools and templates for collecting server inventory, infrastructure budget, and personnel cost data.
3. Conduct comprehensive interviews with first and second-line IT management and technical staff to validate and complete the server template.

4. Analyze the IT budget and personnel costs to identify the current state of service delivery costs by server platform and user. These costs typically include hardware, software, staff, and facilities costs allocated by the main technology areas, such as servers, data network, bulk print, desktop, and so on.
5. Analyze the server and application data to identify groups of similar servers. This requires an understanding of current server utilization factors, service availability, outage impacts, and contingency requirements.
6. Run a short efficiency and service quality health check for each of the major server families to develop outline server scorecards. These give useful insights into people efficiency, server utilization levels, quality of service delivery, and actual current costs.
7. Identify outline solution areas and a credible migration path to get to a future simpler, more cost effective IT infrastructure. This involves the identification of specific technical solution areas and tactical projects.
8. Build the investment case for the future state based on a three-year cost projection for each of the selected technology areas that are candidates for optimization.
9. Deliver final recommendations to the IT executive sponsor, typically in the form of a two-hour, interactive presentation.

Data collection

The basic input data required for this type of project falls into the following categories:

- Infrastructure budget and costs
  - Provide initial templates and assistance to personnel who collect infrastructure cost data for each major technology platform and by cost category
  - Usually include hardware leases, depreciation, software amortization, consulting, services costs, and so on
- IT service support personnel
  - Normally collected using HR records, organization charts, and telephone directory
  - Includes employee, contractors, and any external services provided
  - Provides invaluable insights into often-hidden personnel costs that should be apportioned to the specific platforms in the company's service delivery cost model.
- Server demographics
  - Spreadsheet of all UNIX® and Intel servers
  - Information includes server name, vendor, model, CPU clock, location, function (for example, database, application, web server), status (for example, production, and development), operating system version, and so on
  - Collected by the client, and can be the longest duration task in the project
- Server utilization and availability
  - Samples of utilization data, 24-hour period processor utilization every 10 to 15 minutes, usually considered representative of usage
  - Availability data, for example, service hours, hours of scheduled outages in the last quarter, number of server incidents in the last quarter, and hours of unscheduled outage in the last quarter

Data analysis

Once the data collection is complete, the project team starts analyzing the data from a different perspective.

IT cost analysis

Before starting any discussion about financial analysis, it is vital to identify the core IT infrastructure costs. This is often not quite as simple as it seems.

Consider an organization with an IT budget of approximately $300 million every year. See Figure 1-9. There are many important items in this budget that are not included in the core IT infrastructure cost. For example, $50 million on a major voice network and call center operations and $50 million on bulk printing, mail operations, and postage are not counted as core items.

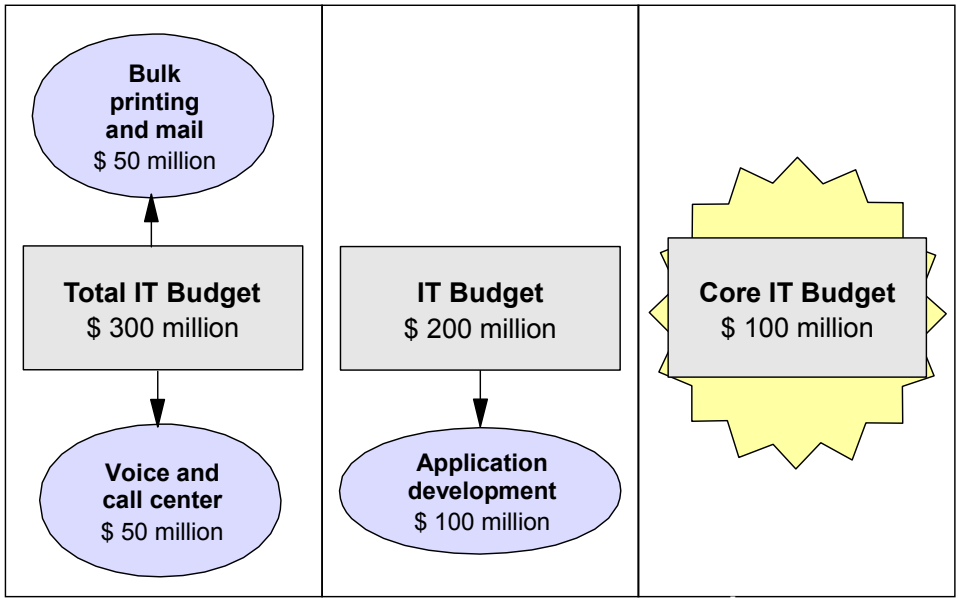


Figure 1-9 Example of IT budget spending

In this case, only $200 million is spent on central IT functions. Of this $200 million, about $100 million is spent on in-house application development, application maintenance, and external development services.

This leaves about $100 million as the core IT infrastructure expense. This typically includes all workstations, data network, and server hardware, software, services, and people associated with core IT service delivery to the user community.

Financial analysis

Once the IT infrastructure budget data collection is complete, the project team builds an end-to-end standardized model of the IT expenditure, allocating the major costs to each technology platform, application, or service.

This model enables a comparative analysis of the major cost components of each major application group and an informal benchmark of a client's costs compared with information from industry sources and other studies.

This analysis provides a very useful objective verification of the assumed cost of ownership of each major platform or application group, often uncovering important costs that are not currently allocated to these technology and application areas.

At this point, it is usually appropriate to do a comparative incremental cost analysis to illustrate the real costs of deploying future applications on various server platforms.

Financial performance and efficiency ratio analyses of overall IT support operations are also developed at this point, as well as what if scenarios and other pro-forma analyses.

This baseline financial analysis forms the baseline for business cases, cost savings, and other projections developed in the later solutions and recommendations phases of the project.

Technical analysis

There are typically hundreds of UNIX servers and thousands of Intel servers in any major enterprise. One of the first steps is to place these servers into a number of broad major categories and further subcategories. See Table 1-2. This is a very important step towards developing a server solution strategy for any enterprise.

Each of these categories has very different technical solutions and cost saving profiles. Practical experience in many projects of this type point to using simple categories of servers at this initial stage in the analysis.

Table 1-2 Categories of servers and their definitions

Category of server	Definition
Large enterprise-wide applications	These servers are often distributed across hundred of locations, for example, servers in retail bank branches and retail stores.
IT Infrastructure	Servers associated with a variety of non-application functions in the enterprise, such as file and print serving, network servers, and so on.
Medium applications	These servers are usually business-critical, applications that are typically centralized and have less than 20 production servers, such as ERP, CRM, and data warehousing.
Small applications	This category includes all applications with five or fewer production servers. This is typically an application server and a database server, or a single combo server.
Other/unknown	In practice, there are always unknown servers. This category should always be less than 10%, and ideally less than 5% of the total server count.

Outline solution design

The next major step is to identify alternative future end states for two or three solution areas, such as Web serving, e-mail, or a specific group of application servers.

These solution areas are selected from current areas, which may show poor service quality or higher than average costs of service delivery. Alternatively these areas can form the foundation for a longer term strategic shift in the IT business or technical model.

There are a variety of technical solutions that play well in this environment, and they break naturally into virtualization, encapsulation, server technology, and system redesign, as shown in Figure 1-10 on page 17.

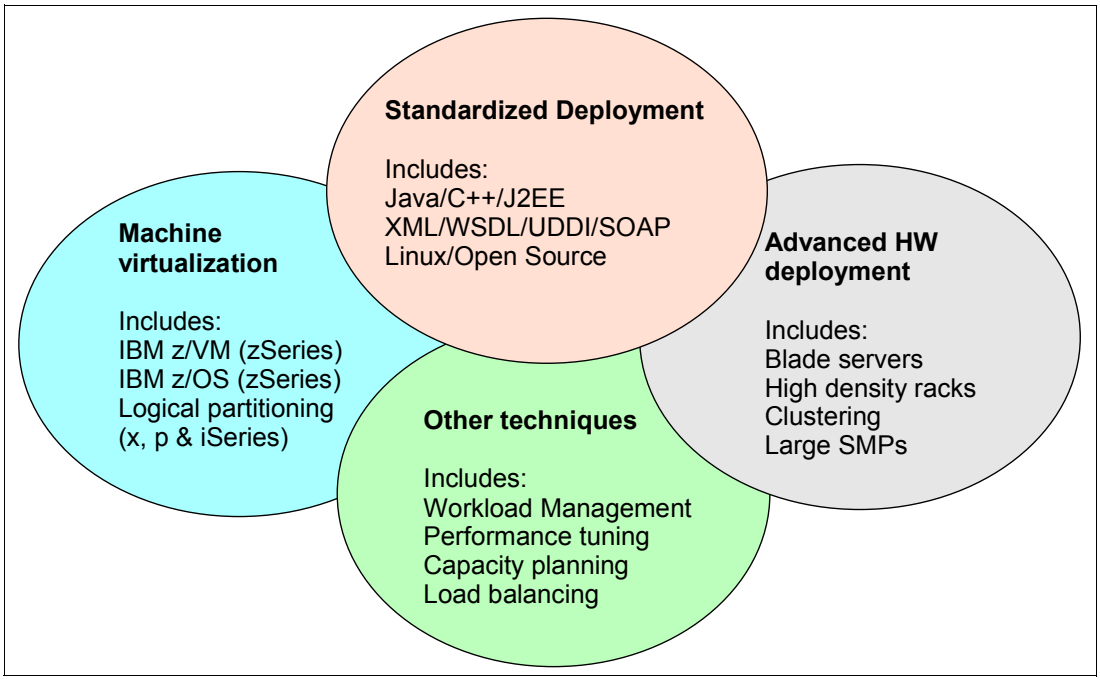


Figure 1-10 IT Server optimization tools and techniques

Real alternative configurations and scenarios are developed based on the technical options, and hardware, software, and people cost data. The realities of the specific situation are factored in, including actual server utilization and actual staff efficiency.

This solution baseline forms the baseline for actual business cases analysis. This uses the classic cost to benefit analysis to compare the alternative future state with the current continue as usual state.

These business cases are usually supplemented with three-year what-if analyses that explore the business case impact of, for example, different utilization levels, different people efficiency ratios, or different hardware configurations.

For more information and a sample case study, a Scorpion white paper is available at:

`http://www.ibm.com/servers/library/pdf/scorpion.pdf`

1.5.3 IT Systems Rationalization Study

The IBM IT Systems Rationalization Study represents the latest breakthrough method of analyzing and optimizing the Corporate IT Infrastructure by providing a quality outline assessment and credible business cases with unmatched speed of delivery.

Offering description and value

The overall approach involves working one-on-one with clients to develop a unique and balanced plan to successfully optimize IT resources and helps reach business goals with less risk, reduced TCO, and greater efficiency.

Using a mix of assumed and client supplied data, the IBM IT Systems Rationalization Study combines best practices across IBM, including IT technical and financial analysis tools, capital budgeting principles, intellectual capital, and emerging technology solutions to deliver a more rapid, affordable and effective IT infrastructure consolidation assessment.

This insight is then used to make initial recommendations on how you can optimize your IT infrastructure, outlining the tactical steps you should take to help achieve your immediate goals and setting the stage for e-business on demand™.

The project

A rapid time-boxed engagement, encompassing analysis, validation, solutions, and business case development is scheduled normally to complete in a time frame of two to three weeks, which may include up to five days of onsite collaboration between IBM and the client.

This assessment methodology is equally effective in analyzing single or multi-platform environments across all major system architectures and platforms, comprised of hundreds to thousands of central and remote Microsoft® Windows®, UNIX® and mainframe servers and storage systems, as well as IT operating expense budgets ranging from tens to hundreds of millions of dollars.

When properly executed and received, the IBM IT Systems Rationalization Study is designed to help IT executives balance conflicting technical options and make sustainable and prudent server platform, storage and software component decisions.

Typical project tasks

Typical project tasks include:

1. Project scope, resources, and core dates agreed upon
2. Baseline server and cost data collection/estimation
3. Onsite working sessions and interviews
4. Outline technical solution areas and business cases
5. Develop recommendations and final presentation
6. Deliver final recommendations during presentation

Project prerequisites

These engagements require a client CIO or CTO executive as a sponsor, and a client project driver to initiate and facilitate the process. In order to help achieve a rapid and meaningful result, a pre-study discussion with the sponsor is essential to defining clear scope and objectives, as well as ensuring that adequate server inventory, utilization and cost data is available.

From start to finish

From strategic thinking to implementation, IBM offers a full range of IT consolidation-related life cycle products and services including assessment, planning, design, migration, implementation and integration services around hardware, software, networking, safety and security, business continuity and recovery, and performance management initiatives.

Combining world-class industry, and business process insight with leading technology expertise, IBM provides superior strategic change and solutions deployment with onsite, outsourced, or on demand delivery. Armed with extensive global and cross-industry experience with proven technology solutions and accountability for results, IBM is a different kind of partner delivering the end-to-end solutions that will help to reduce costs, increase efficiencies, enhance resiliency, and position IT for the emerging on demand business environment.

For more information

To obtain more about the IBM IT Systems Rationalization Study and other consolidation related products and services, contact your local IBM sales representative or visit:

http://www.ibm.com/servers/scon/itsrs/

You can see the IBM internal site for IT Optimization sales kit here:

http://w3.ibm.com/sales/systems/portal/_s.155/254?navID=f220s240&geoID=All&prodID=IBM%20eServer%20And%20TotalStorage%20Products&docID=itopt.skit&docType=SalesKit&skCat=DocumentType

1.5.4 Accelerator for Server and Storage Consolidation

The Accelerator for Server and Storage Consolidation from IBM can help you reduce IT infrastructure operational cost, reduce systems management complexity, increase physical and data security, optimize capacity utilization on server and storage hardware and improve systems reliability, availability and scalability. To do this, IBM provides access to industry best practice in defining a future state server and storage environment based on IBM's significant experience and intellectual capital. Before the Accelerator for Server and Storage Consolidation can start, IBM performs an up-front evaluation workshop with you to enable us to define an initial scope of work and establish a corresponding initial fixed price. During the course of the project we will verify the scope in detail and agree with you about any necessary changes to the scope, price and schedule.

Unlocking consolidation benefits

Whether it is a result of mergers and acquisitions or business model design, IT departments and organizations are often overwhelmed by complex infrastructures, which may be made up of multiple platforms with heterogeneous configurations across various business functions in many locations.

In an effort to reduce total cost of ownership (TCO) and simplify the existing IT infrastructure while continuing to provide optimal client service, many information technology (IT) departments are considering server and storage consolidation.

A cost-effective end-to-end solution

The Accelerator for Server and Storage Consolidation from IBM can help address these challenges by providing a cost-effective, end-to-end consolidation solution from evaluation workshop to final implementation. By using IBM's unparalleled project method and unique tools, we are able to discuss the advantages and provide you with an initial fixed-price proposal to be verified during that project.

From evaluation to strategy

The one-day evaluation workshop is typically the starting point for a successful consolidation service. Its one purpose is to educate you on the guiding principles of server and storage consolidation and the roles and responsibilities in the various stages of the project. The evaluation workshop is also used to derive from you all the information which is necessary to discuss the benefits a consolidation may have in your particular situation and a rapid initial proposal for the engagement. For example, you should be able to describe your servers classified by operating system, management effort, application usage groups and utilization. If you are not able to provide exact numbers, we can work with assumptions and rules of thumb. IBM's long-term experience in server and storage consolidations, incorporated into unique tools, can provide high-level business case input, including ROI.

From assessment to plan

The engagement starts with a detailed assessment and analysis of your server and storage environment. If appropriate, we can deploy auto-discovery tools that automate the assessment to a high degree.

From plan to design

During the plan phase the results of the assessment are analyzed and valued based on business case considerations. We present you with a proposed architecture for your future consolidated environment. This includes, but is not limited to: server platform, storage platform, network, systems management, security and resilience. At the end of this phase we deliver to you a business case for consolidation.

The design phase concludes with a Validation CheckPoint. This Validation CheckPoint is a moment in time when IBM and you make a comparison between the assumptions on which we based the fixed-price of the engagement and the actual situation. We also compare your business case satisfaction criteria with the actual business case. At this point it is possible that contract adjustments can be made, or a termination of the contract can occur with payment of costs incurred to date if the business case does not satisfy your criteria.

Implementing the new landscape

When the detailed design of the consolidated infrastructure is completed, a detailed plan is made to implement the consolidation. Upon separate purchase by you of appropriate hardware, the implementation phase comprises build, test, pilot and deploy sub-phases. Each phase is completed with agreed upon deliverables.

For more information

To obtain more about this IBM Global Services offering contact your IBM sales representative or visit the IBM Web site for your country:

- IBM UK

  http://www.ibm.com/services/uk

- IBM Germany

  http://www.ibm.com/services/de

- IBM France

  http://www.ibm.com/services/fr

- IBM Italy

  http://www.ibm.com/services/it

1.5.5 Total cost of ownership

CIOview is an independent TCO tool vendor, so their TCOnow! tools are based on independent data from IT experts, and can help you understand the full financial impact of your IT purchase decisions over a three, four, or five year period.

These simple to use tools guide you through a structured interview process covering virtually hundreds of cost data elements that span hardware, software, staff, facilities, services, maintenance and even costs associated with downtime.

Start with the provided defaults, then customize those factors most critical to you to ensure the most relevant and accurate business case possible. All default industry and competitor data is automatically kept current through scheduled refreshes from CIOview Corp.

Each TCOnow! tool is tailored to a specific solution. So that tool will provide differentiating detail that shows an understanding of the precise intricacies and inter-relationships IBM customers must be aware of as they implement their particular solution.

The TCOnow! tool has the following features:

- Provides a variety of built-in server consolidation and application scenarios
- Includes server utilization rates to judge the effects of increases of decreases in usage
- Each tool includes built-in system configurations, as well as capacity planning, risk and financial modelling
- Context-sensitive hints provide guidance on every data field, explaining what the field represents as well as industry norms and competitive ranges for each
- All required formulas, including discounted cash flows, depreciation and taxes, are built right into the tool, and all values are easily customized to your specific needs.

TCOnow! tools perform what-if analysis, and generate a detailed, 50+-page business case report describing all assumptions used and the full results of your analysis.

These clearly-written reports include an executive summary and dozens of helpful tables, charts, and graphs, making it easy to tell your TCO story. The report supports every detail of its cost analysis in full, and is ready to present to your CFO.

More information and contact details are available at the following web address:

http://www.ibm.com/servers/solutions/serverconsolidation/tco/

1.6 Candidates for consolidation

Organizations that match the following profile make good candidates for server consolidation:

- Have a directive to reduce IT costs
- Have no control over their IT costs
- Find that they need to integrate their distributed systems
- Find it hard to retain system administration personnel
- Have server management issues
- Need to reduce staffing costs
- Have backup and recovery problems
- Have had mergers or acquisitions
- Have vital data scattered and unsecured

1.7 Decision criteria

Server consolidation is a good solution for server sprawl, but it is not a one-size-fits-all solution. There might be a situation where an organization would not gain much by consolidation.

When an enterprise decides to consolidate servers, it has to analyze what the expected benefits of the operation are and what success factors are involved. The factors that influence the choice are:

- Benefits and business value
- Costs
- Direction of business
- Time to implement
- Package availability
- Risks and contingencies

These are discussed in detail in the following sections.

1.7.1 Benefits and business value

Management and all affected departments have to clearly understand consolidation benefits as described in 1.3, “The benefits of consolidation” on page 4 in the context of their infrastructure and the business value it provides. With this information, they can decide whether consolidation is worth the time and effort for them.

1.7.2 Costs

The main reason for doing server consolidation is that it reduces costs. However, it does incur cost from the time of inception to the time of deployment. These costs include:

- Hardware costs

  Consolidation means the need for new servers, new storage and new devices to do backups. The infrastructure might need an upgrade to provide the stable environment.

- Software costs

  These costs depend on the type of consolidation. Depending on the solution chosen (for example, application integration), software costs may decrease. If the solution involves using new software such as virtualization software or workload management, those costs need to be included.

  Consolidation might involve upgrading some of the software, which would involve additional costs.

  Software license costs for applications and operating systems are different for different systems and different capabilities. These should be understood properly. For example, Microsoft requires a license for each instance of Windows installed under a virtual server.

- Prototype cost

  Depending on the size of the consolidation, there might be a need to set up and test a prototype solution to make sure the solution provides the benefits and business value expected of it. This cost needs to be included.

- Consultants and IT staff

  Consolidation involves a complete analysis of the existing infrastructure to determine the best solution, which means there might be a need to hire a consultant to do this job.

  During the process of consolidation, support staff is needed to analyze, test and implement the solution. Because the IT staff of the organization will be engaged in their current jobs, there is a need for additional IT staff.

- Disruption costs

  Some business may be lost due to disruption of service during consolidation. The question, “how long will the business be impacted and to what extent?” needs to be answered and the implications understood.

  Consolidation involves many changes: new technology, possibly new applications, new infrastructure, integrating existing applications, moving data, changing data formats and changing existing applications. Because these changes have costs attached, they should be kept to a minimum.

  Disruption costs depend on the consolidation solution:

  – With physical centralization, hardware is not available during the transfer.

  – With physical server consolidation, business processes can continue to run on the source platform, and workload is transferred when testing on the consolidated platform is completed and it becomes deployable. The disruption cost is proportional to the migration time.

- With data integration, the disruption cost is proportional to the data migration time.
- With application integration, the best plan is to continue to work on the old platforms until the new solution is implemented and fully tested. In this case, disruption costs are proportional to the applications' migration time.

One way to reduce disruption costs is to implement the solution and test it properly before deploying it.

- Retention costs

  It is best to retain the current environment for a few months as a backup to support the business-critical applications in the event anything goes wrong with the consolidation. This involves some costs, which need to be factored in. If the solution is tested properly before deployment, this cost may be reduced.

1.7.3 Direction of business

As businesses grow, the need for IT resources rises. When planning consolidation, the future growth of the organization should be an important consideration. A consolidation plan should include the ability to scale the infrastructure, not replace it entirely.

1.7.4 Time to implement

The time it takes to implement a server consolidation solution will depend on the kind of consolidation put in place.

The following are some time-to-implement versus disruption-time scenarios:

- With physical hardware centralization, the implementation time is the time taken to transfer the hardware and get it up and running again.
- With physical server consolidation, it is the time it takes for setting up the infrastructure, servers, installing software, testing the applications, and migrating the applications from the source servers.
- With data integration, the implementation time is the time taken to deploy the solution and migrate the data.
- With application integration, it is the time taken to install the new software in a server, test it, and migrate the business processes to the new application.

Although implementation time is important, disruption time is equally important.

1.7.5 Package availability

Another deciding factor is the availability of ready to run packages for the chosen type of consolidation. This refers to packages that do not require significant tailoring to support the consolidation and could grow with the organization as its needs grow in the future. These packages must be readily installable on the new servers.

1.7.6 Risks and contingencies

Server consolidation projects are huge undertakings. As with any project, there are risks associated with it. The risks should be well understood and avoided when possible. You should prepare contingency plans.

The following are a few of the risks:

- Single point of failure

  Consolidation is moving workloads from multiple servers to a single server. A failure in the consolidated server means all the applications running on the server fail. Although new technology has many features that reduce hardware failure, care must be taken to avoid any human errors that might result in the server failure. All the features provided by the hardware to provide high availability must be used to avoid such failures. If required, a high-availability solution should be deployed.

- Failures immediately after deployment

  Due to human errors and maturity of the solution, it is possible that there will be failures in the first few months of deployment of a consolidation solution. Care should be taken to avoid such failures. If required, the current environment must be retained as a backup to keep the business running while the failures are addressed.

- Time for backup

  As many servers are consolidated to a single server, more data needs to be backed up. If it is not taken care of appropriately during deployment, backup time might overlap with the regular operational time of the server, which dramatically affects performance.

1.7.7 Success factors

Almost universally, customers embarking upon a server consolidation have been quite pleased with the outcome. Nonetheless, some comments are worth noting regarding suggestions to make the transition smoother.

- Autonomous departments' fear of losing their independence and becoming subject to perceived burdensome central control has stalled some consolidation projects.
- While vendor analysis may well point out the legitimate benefits of consolidation, centralization versus decentralization remains an emotionally charged argument, since it reflects on organizational structure, company politics, and even the culture of the firm.
- Changes impacting the organizational power structure are not well received when driven from outside the organization itself. Thus, it is very important to have an in-house client champion or sponsor who is sensitive to internal organizational power structures and can overcome reluctance to a fundamental change in computing control.

1.8 Scaling up versus scaling out

The concept of scaling up versus scaling out is used to describe the difference between fewer high-end expensive servers and many low-end low-cost servers. In many cases, an application's architecture determines which approach is possible. In some cases there is a choice. For the purpose of this document, when discussing scale up and scale out, it is assumed that a single application or service is being delivered by the hardware platform.

Scaling up a solution will require the addition of resources such as processor, memory, disk, or I/O to an existing server. In cases where the server has exhausted its upgrade capabilities a later model replacement server will be required. It is also possible that the operating system or application is not able to effectively utilize additional hardware resources. Scale up is typical of many database implementations.

Scaling out a solution requires additional servers. This approach allows resources to be added in a more granular fashion. A scale-out architecture can be complex and requires the application to be capable of functioning as an entity across two or more servers. In many

cases additional software or hardware, and sometimes both, is required to provide the clustering or server farm functionality. Some examples of scale out solutions include Citrix Metaframe and high performance computing.

Table 1-3 compares scale up with scale out.

Table 1-3 Scale up compared with scale out

Scale up	Scale Out
Less hardware to manage	Potentially more hardware to manage
Higher hardware cost	Lower hardware cost
May require a cluster for fault tolerance. This may double the server hardware cost	Using an N+1 (or more) approach, fault tolerance is usually a positive side-effect of scale out
Scheduled down time required for upgrades or maintenance of hardware	Might require a cluster for fault tolerance, possibly doubling the server hardware cost
Easy to administer with basic tools	More difficult to administer, but can be simplified with the use of advanced tools
Scaling limited by hardware, operating system, or application capability	Scaling limited by clustering, application, storage or network capability
Processor upgrades require the same family and speed of processor	Additional nodes can have processors of differing types and speeds

While it is true in many cases that two processor servers may provide better performance per processor than servers with four or more processors, there are resource overheads in managing the communication between nodes in a scale out solution.

Assuming that the performance characteristics of a high-end server and several low-end servers is the same, the cost for the low-end servers is lower. Hardware cost is a significant factor in choosing a scale out approach, but it is not the only one. The difference in cost is largely attributable to the processor required. Two processor servers typically use the Intel Xeon processor, while servers with four or more processors typically use the Intel Xeon MP processor. The commoditization of servers using the Xeon processor has driven the price of these systems down, while the Xeon MP processor requires more complex engineering and is not considered a commodity item, resulting in higher prices.

It is possible to simply categorize servers with four or more processors as scale up and servers with one or two processors as scale out. This might be appropriate in some cases. However, it can be viable to adopt a scale out approach for servers with four or more processors in the following situations:

- When the application scales well beyond two processors
- When a goal is to reduce the number of servers
- To possibly reduce administration
- To take advantage of Xeon MP processor or IBM's second generation Enterprise X-Architecture™

For more information about scale up versus scale out considerations, see *VMware ESX Server: Scale Up or Scale Out?*, REDP-3953.

1.9 Pitfalls of server consolidation

During server consolidation, everyone's concentration is on the big project. It is possible for the team to overlook some of the not-so-obvious problems. The following are a few pitfalls:

- Network bandwidth has a huge effect on centralization. The consolidation team has to make sure the infrastructure can provide the necessary bandwidth for the consolidated servers.
- Network downtime and slow response have a huge effect on centralization. The effect should be clearly understood and expectations set appropriately.
- Operating system rationalization affects application centralization most. Many environments consist of several operating systems and versions. An application that is supported on one operating system and version might not be supported on another.

  For example, an application that is supported on Microsoft Windows NT 4.0 might not be supported on Windows 2000 or Windows 2003. This can also be true of various editions within the same version. An application that is supported on Windows 2000 Server might not be supported on Windows 2000 Server Datacenter Edition.
- Software compatibilities and support affects application centralization most. Applications that are migrated from multiple servers to a single server must be compatible, and behave the same way after integration as they behaved before. It is a good idea to talk with the different software vendors about their compatibilities and whether such a configuration is supported. If it is not supported directly, it is important to understand the kind of support available.

  For example, a software vendor may support their application with a downlevel remote database management system (RDBMS) such as Microsoft SQL 7.0, but may not support the same application with a later release such as Microsoft SQL 2000, or a different RDBMS such as DB2 or Oracle. This problem is compounded if the software vendor does not have a later version that does support new RDBMSs or if the business is unable to migrate to a newer version.
- Operational overhead of having systems remotely has an effect on centralization. Because the servers are moved to a remote location, the support structure is located in the remote location. Users who are accostumed to knocking on the administrator's office door when they have a problem now have to talk to a remote person. Expectations about the new service model should be communicated.
- Support for operational overhead for collocating and cohosting applications becomes an issue when centralizing applications. This is due to application dependencies on common libraries of differing versions and poorly-written applications potentially causing the server to become unstable.

  To avoid these issues, a suitable development and testing environment should be instituted. Rigorous change control should then be implemented to ensure that any changes to application code are regression tested. These processes may not be required in an environment that consists of a 1:1 relationship between servers and applications.
- Effect of time zones on IT staff support

  This has an effect on centralization. Moving servers to a location in a different time zone from that of the users affects the hours of service they get during their work day. Care should be taken to minimize this operational overhead for collocating and cohosting applications.

2

VMware ESX Server

This chapter provides an overview of VMware ESX Server V2.x, where x is any release within version 2. The sections are divided as follows:

VMware provides documentation on many aspects of VMware ESX Server. The following links provide lists of documentation available for download.

- VMware ESX Server documentation

  http://www.vmware.com/support/pubs/esx_pubs.html

- VMware ESX Server compatibility guides, white papers and tech notes

  http://www.vmware.com/support/resources/esx_resources.html

2.1 VMware ESX Server overview

VMware ESX Server is virtual machine software used for consolidating and partitioning servers in high-performance environments. It lets you transform physical computers into a pool of logical computing resources. You can partition physical servers into secure virtual machine servers. You isolate your operating systems and applications in these multiple virtual machine servers that reside on a single piece of hardware. You can then distribute these resources to any operating system or application as needed, when needed.

It is a cost-effective, highly scalable virtual machine platform with advanced resource management capabilities. Ideally suited for corporate IT and service provider data centers, VMware ESX Server allows you to:

- Implement server consolidation

  You can consolidate applications and infrastructure services onto fewer highly scalable, highly reliable enterprise-class servers.

- Deliver high availability and provide for disaster recovery

  With a stable, uniform platform, you can deliver more services and deploy new solutions faster and more efficiently. You can store critical data in secure, isolated virtual servers to protect against the vulnerabilities of physical servers.

- Guarantee service levels

  Like an internal service provider, your IT department can deliver guaranteed server resources of CPU, memory, disk bandwidth and network bandwidth at optimum performance levels, improving service to customers.

- Streamline development and testing

  Your software developers and quality assurance engineers can work effectively with multiple machine environments and build more realistic tests in less time with less hardware.

VMware ESX Server runs directly on your system hardware to provide a secure, uniform platform for deploying, managing, and remotely controlling multiple operating systems. With VMware ESX Server, you can:

- Move applications running on dedicated systems into separate virtual machines on a single, more reliable and scalable system
- Remotely manage servers from any location, simplifying server maintenance
- Guarantee service levels with advanced resource management controls
- Script common monitoring and management tasks
- Increase capacity without adding new physical systems

VMware ESX Server V2.x consists of the base product and optional add-on components. For an overview on the IBM offering, see Appendix A., “VMware products from IBM” on page 127.

2.1.1 How does VMware ESX Server work?

VMware ESX Server transforms physical systems into a pool of logical computing resources. Operating systems and applications are isolated in multiple VMs that reside on a single piece of hardware. System resources are allocated dynamically to any operating system based on need, providing mainframe-class capacity utilization and control of server resources.

VMware ESX Server simplifies server infrastructure by partitioning and isolating server resources in secure and portable virtual machines. ESX Server enables these server

resources to be remotely managed, automatically provisioned, and standardized on a uniform platform. Advanced resource management controls allow IT administrators to guarantee service levels across the enterprise. ESX Server runs directly on the system hardware to provide a secure, uniform platform for deploying, managing, and remotely controlling multiple VMs.

With the VMware ESX Server virtual infrastructure we can:

- Manage servers remotely from any location, simplifying server maintenance
- Perform zero-downtime maintenance on live systems without interrupting service
- Dynamically remap resources to software to improve efficiency and respond better to business demands
- Deploy new virtual machines using standardized templates and increase capacity without adding hardware
- Guarantee service levels with advanced resource management controls
- Script common monitoring and management tasks

2.2 VMware ESX Server architecture

The design of the VMware ESX Server core architecture implements the abstractions that allow hardware resources to be allocated to multiple workloads in fully isolated environments. VMware ESX Server incorporates a resource manager and a service console that provide bootstrapping, management and other services.

The key elements of the system's design are:

- The VMware virtualization layer, providing the idealized hardware environment and virtualization of underlying physical resources
- The resource manager, enabling the partitioning and guaranteed delivery of CPU, memory, network bandwidth and disk bandwidth to each VM
- The hardware interface components, with device drivers enabling hardware-specific service delivery while hiding hardware differences from other parts of the system

Figure 2-1 is a high level diagram of the Intel system architecture with VMware ESX Server.

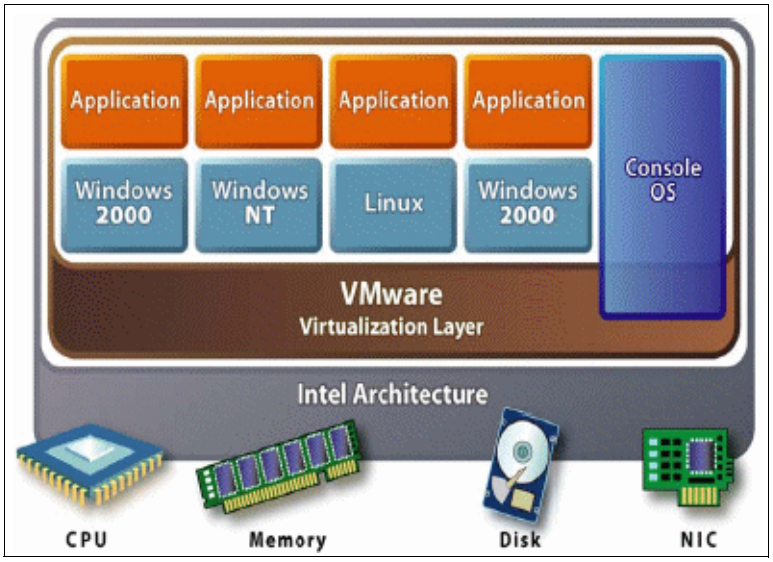


Figure 2-1 Intel system architecture with VMware ESX Server

The server is capable of simultaneously hosting heterogeneous operating systems and applications. The Console OS is provided as an interface to support the VMware ESX Server kernel.

2.2.1 Virtualization at a glance

VMware ESX Server virtualizes the resources of the physical system for use by the virtual machines.

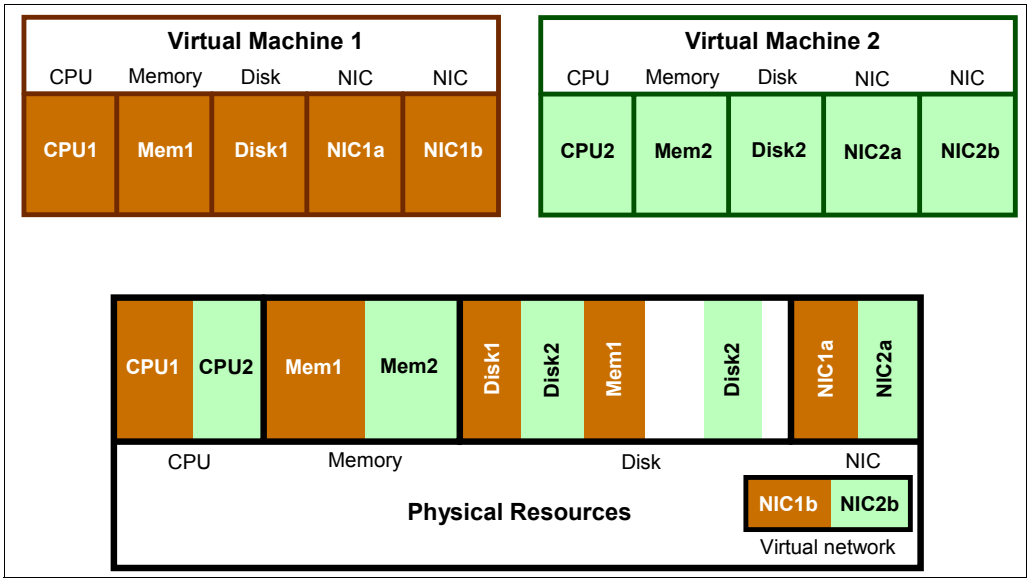


Figure 2-2 Virtualization at a glance

In Figure 2-2 on page 30, each VM is configured with one CPU, an allocation of memory and disk, and two virtual Ethernet adapters. In reality, they share the same physical CPU and access noncontiguous pages of memory. Part of the memory of one of the VMs is currently swapped to disk. Their virtual disks are files on a common file system. Each has a network interface bound to and sharing a single physical network adapter. The second network interface in each VM is bound to a virtual network interface within the ESX Server system.

2.2.2 Virtualization in detail

The VMware virtualization layer brings hardware virtualization, pioneered on IBM VM/370 and other mainframe environments, to the standard Intel server platform. The virtualization layer is common among VMware desktop and server products, providing a consistent platform for development, testing, delivery and support of application workloads from the developer desktop to the workgroup and to the data center.

As with mainframe virtualization, the VMware virtual machine offers complete hardware virtualization. The guest operating system and applications inside a VM can never directly know which physical resources they are accessing, such as which CPU they are running on in a multiprocessor system or which physical memory is mapped to their pages. The virtualization of the CPU incorporates direct execution. The hardware CPU executes non-privileged instructions without overheads introduced by emulation.

The virtualization layer provides an idealized physical machine that is isolated from other VMs on the system. It provides the virtual devices that map to shares of specific physical devices. These devices include virtualized CPU, memory, I/O buses, network interfaces, storage adapters and devices, human interface devices, BIOS and others.

Each VM runs its own operating system and applications. They cannot talk to each other or leak data, other than through networking mechanisms similar to those used to connect separate physical machines. This isolation leads many users of VMware software to build internal firewalls or other network isolation environments, allowing some VMs to connect to the outside while others are connected only to virtual networks through other VMs.

A comparison between the traditional Intel architecture and the VMware ESX Server installed on Intel architecture is shown in Figure 2-3. This diagram shows that VMware presents individual virtual hardware to each VM. In this way, the operating systems and applications installed in VMs are not aware that the hardware is virtual.

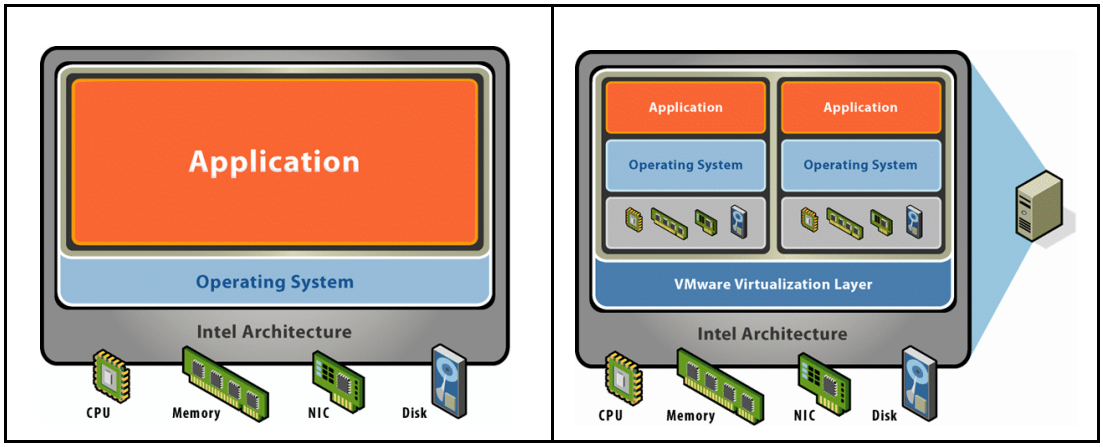


Figure 2-3 A comparison between a traditional Intel architecture and VMware ESX Server

CPU virtualization

Each virtual machine appears to run on its own CPU, or set of CPUs, fully isolated from other VMs, with its own registers, translation lookaside buffer, and other control structures. Most instructions are directly executed on the physical CPU, allowing compute-intensive workloads to run at near-native speed. Privileged instructions are performed safely by the patented and patent-pending technology in the virtualization layer.

Memory virtualization

While a contiguous memory space is visible to each VM, the physical memory allocated may not be contiguous. Instead, noncontiguous physical pages are remapped efficiently and presented to each virtual machine. Some of the physical memory of a virtual machine may in fact be mapped to shared pages, or to pages that are unmapped or swapped out. This virtual memory management is performed by ESX Server without the knowledge of the guest operating system and without interfering with its memory management subsystem.

Disk virtualization

Support of disk devices in VMware ESX Server is an example of the product's hardware independence. Each virtual disk is presented as a SCSI drive connected to a SCSI adapter. This device is the only disk storage controller used by the guest operating system, despite the wide variety of SCSI, RAID and Fibre Channel adapters that might be used in the system.

This abstraction makes VMs at once more robust and more transportable. There is no need to worry about the variety of potentially destabilizing drivers that may need to be installed on guest operating systems, and the file that encapsulates a virtual disk is identical no matter what underlying controller or disk drive is used.

VMware ESX Server can be used effectively with storage area networks (SANs). ESX Server supports QLogic and Emulex host bus adapters, which allow an ESX Server computer to be connected to a SAN and to see the disk arrays on the SAN.

Network virtualization

You may define up to four virtual network cards within each VM. Each virtual network card has its own MAC address and may have its own IP address or multiple addresses as well. It might be mapped to a dedicated network interface on the physical server, known as the Outbound Adapter, or virtual network interfaces from multiple virtual machines may be connected to a single network card. VMware ESX Server manages both the allocation of resources and the secure isolation of traffic meant for different virtual machines even when they are connected to the same physical network card.

Before virtual machines defined to ESX Server can be configured to access a network, at least one virtual ethernet switch must be created within ESX Server. When two or more virtual machines are connected to the same virtual ethernet switch, network traffic between the virtual machines will be routed locally. If an outbound adapter is attached to the virtual ethernet switch, each virtual machine will also be able to access the external network through the outbound adapter. See Figure 2-4.

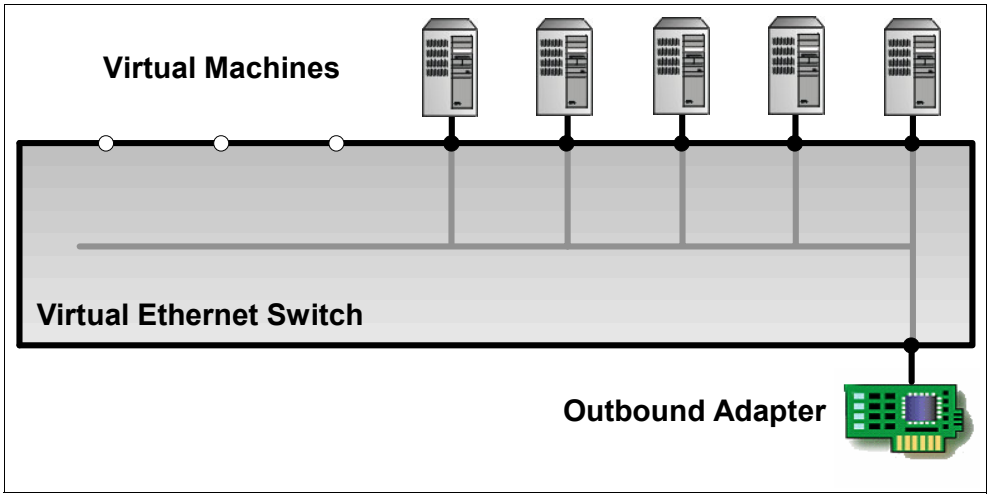


Figure 2-4 Virtual Ethernet Switch

Virtual LAN (VLAN) and port groups

An additional network option involves patching a virtual network interface to a Virtual LAN (VLAN). A VLAN is made up of a port group created within ESX Server. Each port group is assigned a unique VLAN ID. VLANS can be used to isolate virtual machine network traffic on the ethernet virtual switch and on the external network. See Figure 2-5.

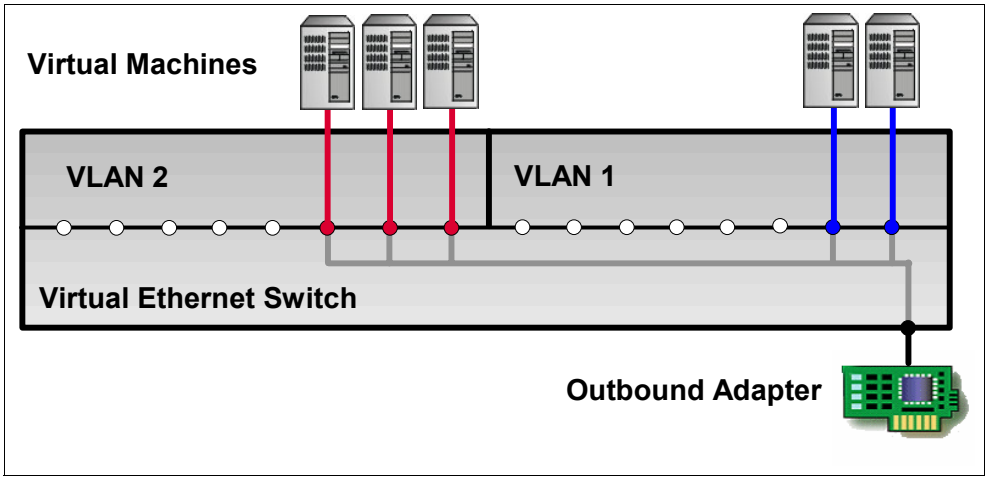


Figure 2-5 Virtual LANs

VLAN connections may be used for high-speed networking between VMs, allowing private, cost-effective connections between the them. The isolation inherent in their design makes them especially useful for supporting network topologies that normally depend on the use of additional hardware to provide security and isolation.

Figure 2-6 on page 34 shows an example of how VLANs can be used to isolate virtual machines from the physical network.

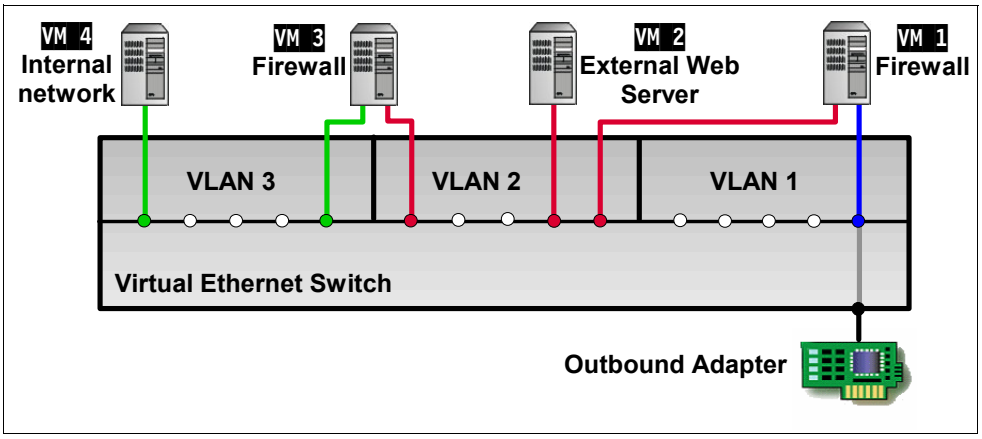


Figure 2-6 Using VLANs to create a DMZ

For example, an effective firewall can be constructed by configuring one virtual machine on an ESX Server system with two virtual Ethernet adapters, one bound to a VLAN connected to an outbound adapter, giving it a connection to a physical network. The other is bound to a different VLAN ID.

Other virtual machines would be connected only to the internal VLANs. By running filtering software in the dual-homed virtual machine, a user can construct an effective firewall without the need for additional hardware and with high-performance virtual networking between the virtual machines.

A similar approach can be used with multitier applications, with the Web or application servers reachable from other systems but with the database server connected only to the other tiers.

2.2.3 Software compatibility

In the VMware ESX Server architecture, guest operating systems interact only with the standard x86-compatible virtual hardware presented by the virtualization layer. This provides the capability for VMware to support any x86-compatible operating system. In practice, however, VMware supports a subset of x86-compatible operating systems that are tested throughout the product development cycle. VMware documents the installation and operation of these guest operating systems and trains its technical personnel in their support.

Because applications interact only with their guest operating system, and not the underlying virtual hardware, once operating system compatibility with the virtual hardware is established, application compatibility is not an issue.

2.2.4 Service console

The service console is a modified Red Hat 7.2 Linux operating system which could be considered a privileged virtual machine with access to some hardware.

A popular misconception is that the VMware ESX Server kernel is also based on a modified Red Hat kernel. This is not the case. VMware ESX Server is a highly efficient hypervisor/kernel built for the sole purpose of managing VM. VMware uses the Console OS's file system to:

- Store its kernel files and drivers
- House CD images, VM configuration files, log files and suspend files

VMware provides all drivers for ESX Server. Since VMware is its own OS, Linux drivers are incompatible and will *not* work with VMware ESX Server. The complete architecture is shown in Figure 2-7.

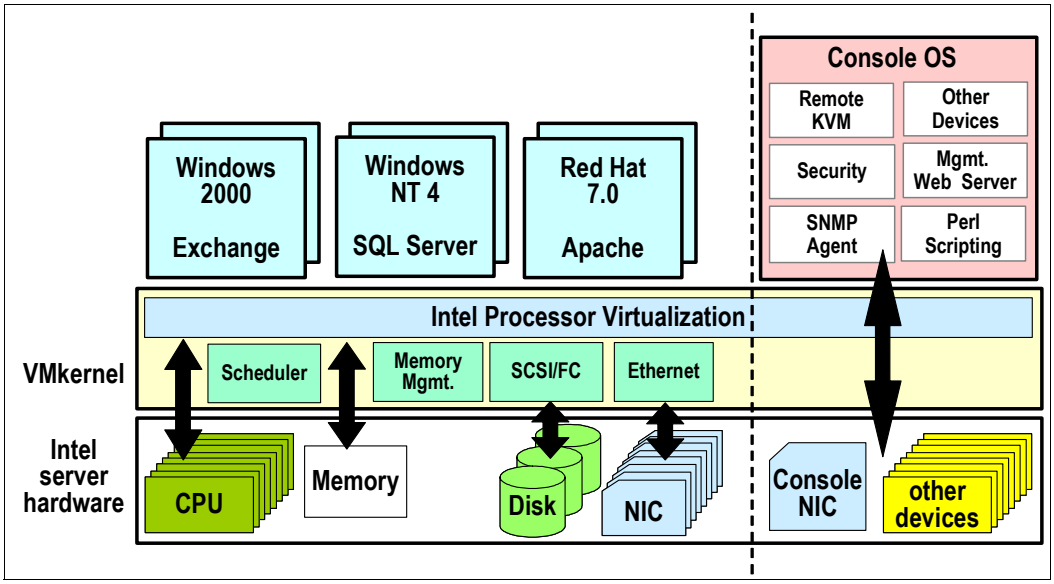


Figure 2-7 VMware ESX Server architecture with Console OS

Service console functions

The VMware ESX Server system management functions and interfaces are implemented in the service console. These include the HTTP, SNMP and API interfaces described above, as well as other support functions such as authentication and low-performance device access. The service console is also installed as a first component and is used to bootstrap the ESX Server installation and configuration, as well as to boot the system and initiate execution of the virtualization layer and resource manager. In VMware ESX Server, the service console is implemented using a modified Linux distribution.

2.3 VMware ESX Server virtual machine specifications

VMware ESX Server virtual machines present the following hardware under the supported guest operating systems:

- Processor
  - Virtual Intel® IA-32 based uniprocessor system
  - Virtual Intel® IA-32 based 2-way system (with VMware Virtual SMP add-on module)
- Memory
  - Up to 3.6GB of usable memory per VM
- IDE Drives
  - IDE-CD-ROM
- SCSI Devices
  - Up to four virtual SCSI adapters and up to 15 SCSI disks or pass-through devices per adapter
  - Support for SCSI devices, including DAT and DLT SCSI tape and SCSI CD-R/RW drives

- BIOS
  - PhoenixBIOS™ 4.0 Release 6-based BIOS
- Networking
  - Up to four virtual Ethernet NICs
  - Each virtual NIC may be high-performance VMware virtual NIC or AMD® PCnet™-PCI II compatible virtual NIC
  - Supports any protocol that guest OS supports over Ethernet
  - Multiple high-performance Ethernet-compatible virtual networks
  - Pre-boot Extension Environment (PXE) support for guests allows installation of operating systems in virtual machines over the network
  - VLAN (802.1q) support enhances security and leverages your existing network security infrastructure
- Guest operating systems
  - Microsoft® Windows® Server 2003 (Standard, Enterprise and Web)
  - Microsoft® Windows® XP Professional
  - Microsoft® Windows® 2000: Server, Advanced Server, Terminal Services Edition
  - Microsoft Windows NT®: 4.0 Server, Service Pack 6a
  - Red Hat Enterprise Linux (AS) 2.1 and 3.0; Red Hat 7.2, 7.3, 8.0, and 9.0
  - SuSE Linux 8.2 and SuSE Linux Enterprise Server 8.0
  - Novell NetWare 6.5, 6.0 (Support Pack 3), and 5.1 (Support Pack 6)

2.4 Support for Microsoft software in VMware virtual machines

VMware addresses questions of support for Microsoft software in VMware virtual machines and offers resolution paths.

2.4.1 Support options

Microsoft officially does not support issues that occur in Microsoft operating systems or programs that run in a VMware virtual machine until it is determined that the same issue can be reproduced outside the virtual machine environment. A virtual machine can offer a more stable platform than the hardware equivalent, where the corresponding devices can change on a daily basis. As a result, VMware has provided significant productivity benefits to Microsoft customers since the adoption of VMware's virtual infrastructure technology began seven years ago.

Other measures customers can take to ensure their production environments are fully supported include:

- IBM can provide end-to-end support packages that include a Microsoft operating system and VMware software. In fact, IBM provides Windows and Linux support within virtual machines when VMware ESX Server support is purchased with a bundled VMware solution. IBM support of VMware ordered through IBM is mandatory and is separately available from IBM Technical Support Services. See "IBM support for VMware software" on page 130.
- Review the documented procedures for reproducing a virtual environment on a physical system, as described on the VMware Web site at:

  http://www.vmware.com/support/v2p/
- VMware can also work with customers to reproduce issues on native hardware. VMware and EMC will continue to work on this issue to help Microsoft do what is right for clients.

- Affected clients are encouraged to voice their opinions to Microsoft and convince the company to adopt an official policy to broaden the support for clients that choose to take advantage of VM technology.

2.4.2 Summary

As VMware software adoption in the Enterprise accelerates, the issue of support for Microsoft software running in VMware virtual machines is raised frequently. The reality is that Microsoft has responded to support calls from customers, running applications in virtual machines.

- IBM, Dell, and HP all have support offerings for Windows running in a VMware virtual machine.
- There are tools and procedures available to replicate virtual machine workloads on physical platforms.

In recent months, questions have been raised regarding Microsoft support for Windows operating systems and Microsoft applications running inside VMware virtual machines. Many of these questions are fueled by the rapidly-growing deployments of VMware virtual infrastructure in production and mission-critical environments.

2.5 VMware ESX Server usage scenarios

The following usage scenarios provide examples of how VMware ESX Server can be used to reduce costs and improve business processes.

2.5.1 Server consolidation

Server consolidation is one of the primary IT infrastructure issues addressed by VMware ESX Server. For an overview of server consolidation issues and considerations, refer to Chapter 1, “Server Consolidation” on page 1.

Reduced total cost of ownership

With VMware ESX Server, one physical server hosts multiple instances of different operating systems in the virtual machines. Total cost of ownership (TCO) is reduced by:

- Increasing efficiency of resources
- Increasing efficiency in operations
- Reducing costs in technical support and training
- Reducing floor space and energy requirements
- Faster deployment of new services to support changing business needs
- Building cost-effective and consistent development and test environments
- Reducing disaster recovery time

According to VMware, using their solution reduces hardware costs by 28-53% and operations costs by 72-79%, reducing the overall total cost of ownership by 29-64%.

Productivity

In a consolidated environment, it is rare for all VMs to be actively using the resources allocated to them (CPU, memory and I/O) simultaneously. ESX Server capitalizes on this fact by actively monitoring the usage of resources and allocating more resources to active VMs than idle VMs. Thus by efficiently allocating the resources, ESX Server provides VMs more virtual resources when available in the physical server, improving the productivity of the server.

Guaranteed service levels

Servers with different workloads can be consolidated to VMs on a single hardware platform. One server may need four times the CPU capacity, twice the memory and 10 times the I/O bandwidth than that of another server, to satisfy the applications it is running. If, under the consolidated environment, such a service level were not guaranteed, the performance of the first server could be poor and may not be as productive as it was before consolidation.

ESX Server works on a proportional-share based policy that guarantees service level. All resources are divided into number of shares and each VM is allocated a number of shares by the administrator, based on their priority. ESX Server guarantees that a VM with 2000 shares of a particular resource gets twice the amount of that resource than a VM that has 1000 shares when it needs it.

With this feature, the productivity of a server with more workload need not be compromised by moving to a consolidation solution.

CPU resources

CPU resources are allocated to virtual machines on a proportional share-based processor scheduling. Each VM is allocated a specified number of shares. The amount of time the VM gets depends on the proportions of its share against the total number of available shares and the current usage of CPU resources in the systems.

Shares represent a guarantee of service. If two VMs are running and one has twice the CPU shares as the other, it is allocated twice as many CPU cycles as the other, provided both VMs are active.

Shares are not hard reservations, so under-utilized allocations are not left idle. Instead, they are allocated to active VMs that can benefit from them. For example, in a single CPU system, it is possible for a VM to use more than 90% of the CPU even its share is just 10% of the total number of shares.

Most instructions executed by the guest operating systems are run directly on the processor, allowing compute-intense workloads to run at near-native speed. Only privileged instructions are performed in the virtualization layer.

In an SMP server, a VM can be restricted to a single CPU or a set of CPUs. By using this feature, you can improve the productivity of VMs that interact heavily by restricting them to two isolated sets of CPUs.

Memory resources

ESX Server provides an oversubscription feature that uses virtual memory techniques to allow virtual machines to access more memory than available in the hardware, thereby increasing the productivity of the operating systems running on them. VMs can be created with memory that is three times more than the physical memory available in the system.

Security

Virtual machines created under ESX Server are totally isolated. An operating system or application running on one VM cannot access memory of other VMs nor can they access ESX Server directly. Invalid instructions executed on a VM will not affect any other VMs or ESX Server in the same physical machine.

To increase performance, most instructions executed by the VMs are run directly in the processor. For safety reasons, privileged instructions are executed in the virtualization layer of the ESX server, thereby securing the system from runaway applications or operating system.

2.5.2 Quality assurance

Developers and quality assurance (QA) departments spend much of their time developing for and testing new builds of their software across multiple operating environments.

Some problems that are encountered in this environment include having to ensure that a given application can work with different versions of browsers and operating systems, and the setup and maintenance of a duplicated and isolated environment.

With VMware ESX Server, multiple virtual machines can be created on a single physical server. Side-by-side comparisons can then be done to monitor how the software under development works in various operating environments. Additionally, VMs can be networked to allow for the development and testing in a virtual networked environment. See “Network virtualization” on page 32.

2.5.3 Development and testing

In addition to the problems associated with software compatibility, developers and testers operate in a work in progress state. If they write code that is not effective, they risk damaging the current build of the software.

With VMware ESX Server, each VM can be completely isolated from another. Not only does this provides a safe box from which to work, but it is also an easy way to recover from system crashes due to unstable code.

VMware products offer flexibility in three disk modes: Persistent, non-persistent and undoable. In either non-persistent or undoable mode, developers and testers can literally trash the VM, reboot it and, within seconds, return to the their start state. This is particularly useful when trying to debug a new program or DLL.

2.5.4 Supporting several environments

Engineers may have a requirement to support a wide variety of computing environments. They may have several machines on their desk. Even with this many machines, there are limitations:

- Time consuming to configure each machine to replicate the development environment, then reconfigure for the next challenge
- Testing on the production machine is dangerous and can crash your system
- Software incompatibilities

This problem can be addressed by creating multiple VMs that mirror the various environments being supported. The engineer can start the appropriate VM and troubleshoot the problem. When the engineer is finished with the VM, it can simply be shut down, returned it to its original state, making it ready for the next time.

2.5.5 IT department and operations

IT departments, in addition to making sure that a company’s infrastructure is running smoothly, is responsible for migrating the company to new operating systems and applications. This includes the need to evaluate software and network configurations prior to introducing them to the entire company. This is a costly and time consuming process, requiring several test machines and network configurations.

Using VMware ESX Server running multiple VMs networked together, an IT professional can rehearse administration tasks, ensure the integrity and reliability of a new corporate operating environment, and test all applications for compatibility from one machine saving precious time, money, and hardware costs.

2.5.6 Provisioning

This is one of the greatest and sometimes under-estimated benefits of VMware ESX Server.

In a standard IT department there is usually a process to justify and procure hardware for new projects or for ad hoc tasks. The process to generate an order can be quite lengthy, and once ordered there may be a delay in the delivery of all the requested components.

Redeploying hardware internally can also be a time-consuming task, while the engineer adds or removes components, and searches for the right maintenance software or drivers.

It is typical for many IT departments to have a variety of hardware platforms running with different BIOS and firmware levels, thus providing inconsistencies in installations and problem determination.

With a VMware ESX Server environment, the consistency of the virtual hardware creates an ideal environment for developing and maintaining fewer standard or common operating environments (SOE or COE). With these standard environments, the IT department can simply rubber stamp virtual machines as required.

If a project is starting up and they require a development or proof of concept server, it will only take minutes to create a VM based on the standard image. Providing there are available resources, there is no need to acquire new hardware. Processor, memory, disk and network resources can be allocated to the project in a controlled fashion, which is excellent for IT departments wanting to adopt a charge-back system to the business.

For IT departments, the flexibility of adding VMs will save the time of its support staff and allow them to quickly duplicate environments for troubleshooting, education or evaluation.

2.5.7 High availability

High availability presents a number of challenges, in both technical and business areas. Having a single point of failure is a key concern. However, most failures are either due to software failure or human error, not hardware failure. Virtual Machines isolate the impact of these common errors.

Traditional high availability involves buying two servers, or a cluster of servers, for each major application, which is obviously very costly. For that reason customers have tended to shy away from high availability and have left key applications unprotected from failure.

With VMware ESX Server it is possible to cluster virtual machines, providing the guest operating system natively supports this function, in the following ways:

- *One box clustering* means clustering from one VM to another on the same physical ESX Server machine.
- *Fully virtualized* means clustering from one VM to another on separate ESX Server machines, which requires shared storage.
- *Mixed environments* means clustering from a physical server to a VM, which requires shared storage.

VMware technology can greatly benefit high availability. Because VMs are completely isolated, application failures will not take down an entire server. Only one partition can be affected.

Standard Microsoft Windows or Linux clustering technologies can be used to provide high availability between VMs. One VM can fail over to another in the event of a software problem. This is an N+1 failover solution on a single system.

Clustering between VMs is far more cost-effective than purchasing duplicate servers. Cluster failures are usually due to subtle differences in hardware or software. This can be prevented because clusters installed in VMs are running on the same virtual hardware and software. Regardless of the underlying hardware, VMware ESX Server presents virtual machines with identical virtual hardware.

Note: VMotion is not supported on clustered virtual machines as they must be stored on a separate private Virtual Machine File System (VMFS) volume.

One box clustering

Using VMware to cluster can make clustering more affordable for organizations and may offer additional resiliency where previously there was none. See Figure 2-8. While on the surface it may seem as though there is no need to cluster two virtual machines in the same physical server, consider a situation where the availability of an application is affected by problems with the operating system, the application itself, or operator errors. In many cases (approximately 80-90%) downtime is not caused by a hardware fault.

Having a standby cluster node on the same physical ESX Server will use some memory resources, but will not use many CPU resources.

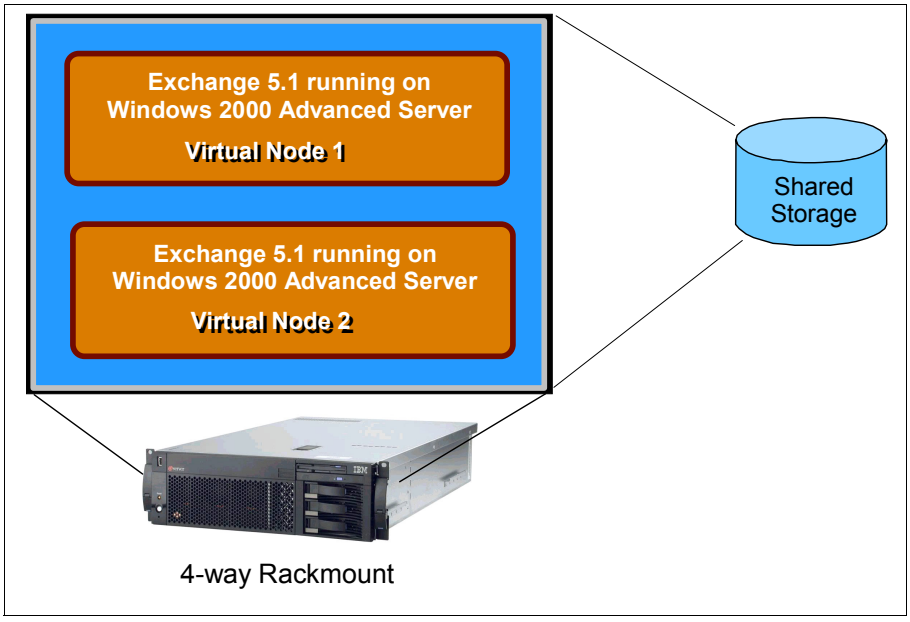


Figure 2-8 The benefits of a one box cluster

The benefits to the client of one-box clustering are:

- Protects against all OS and application faults
- Lower cost makes clustering more viable for organizations that cannot otherwise afford it
- Creates an inbox failover solution for applications that are not cluster aware

Fully virtualized

Fully virtualized clustering adds VMware based hardware failover protection to the one box cluster solution. This solution requires shared storage as the cluster nodes are installed on separate servers. The advantage of this approach over conventional clustering is that many VM clusters can be implemented using only two physical servers. See Figure 2-9. For example, think of consolidating four clusters of two machines each to two physical machines with four VMs each.

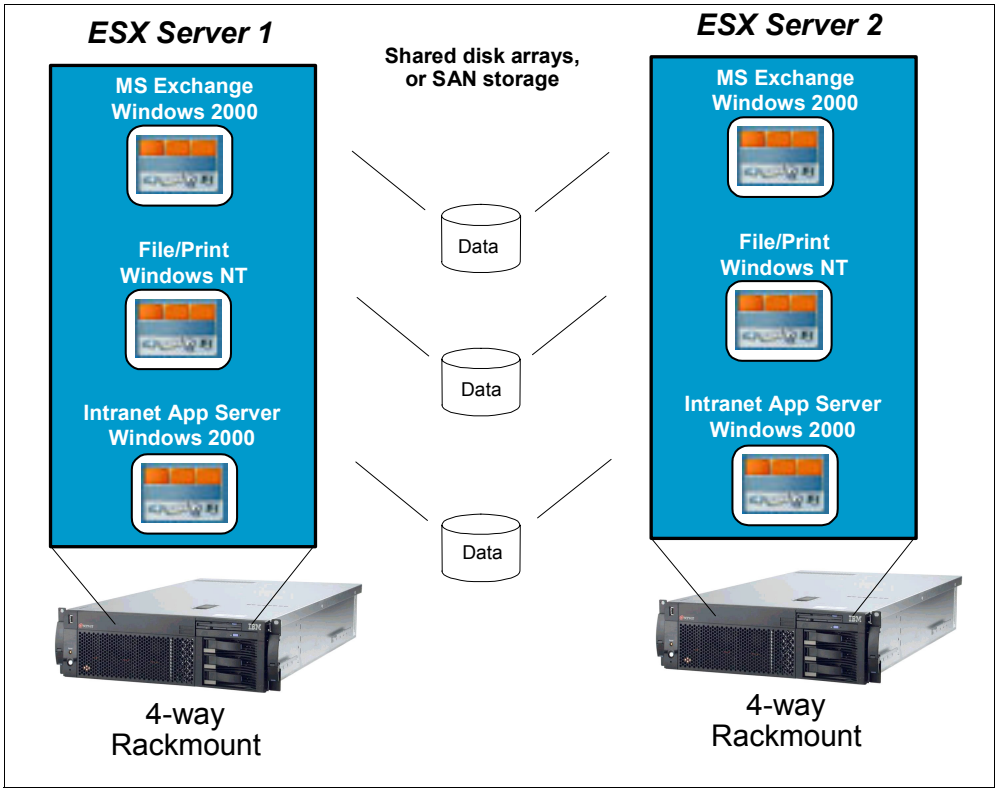


Figure 2-9 Fully virtualized clustering with VMware based hardware failover

The benefits to the client of fully virtualized clustering are:

- Cluster is still made out of VMs
- Allows high availability with less hardware (machines, network cards, and so on)
- Full support for NLB, MSCS, and VCS in VMs in ESX Server with no modifications.
- Ability to deal with crash of a physical machine
- Protection from both hardware and software failures

Mixed environments

Using mixed environments clustering is a good compromise for organizations wanting to implement low-cost clustering, but are not ready to run their production mission-critical systems on VMware ESX Server. Using active and passive clustering, the active node can operate on a standard server. However the passive, or standby, node can be implemented as a VM. In doing so, it is possible to cluster many physical servers to a single VMware ESX Server. See Figure 2-10 on page 43.

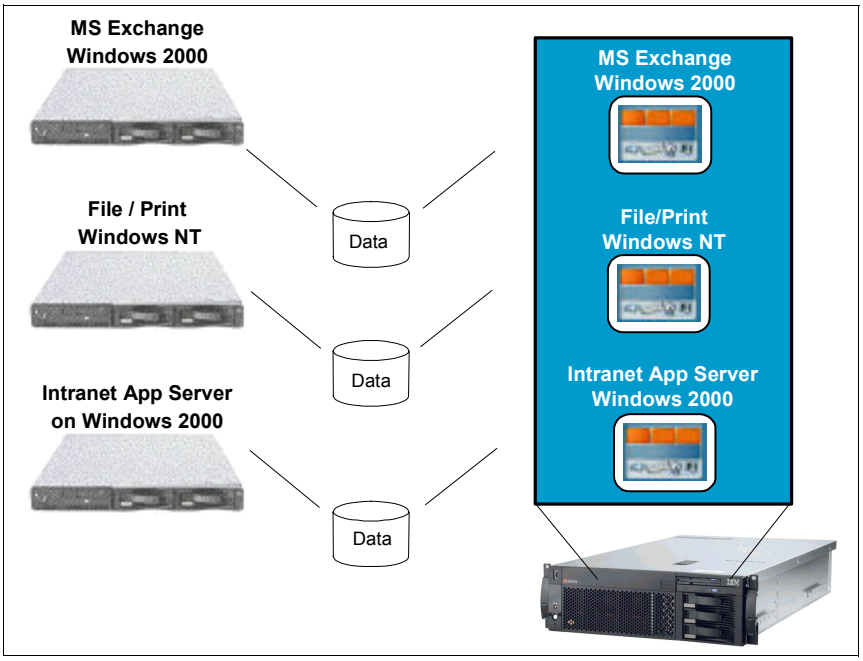


Figure 2-10 Mixed environments clustering

The benefits of mixed environments clustering to the client are:

- It offers a highly flexible and cost-effective solution.
- One failover server is purchased, instead of the several that would be required for traditional clustering of several servers.
- It retains the benefit of having the failover server on a separate piece of hardware. Should a production server experience hardware or software failure, they are still protected.

2.5.8 Backup

There are several backup options available for VMs. It is possible to attach a physical tape drive to a VMware ESX Server. However, without proper management there is a risk of device contention.

For many organizations the simplest way to back up their VMs is to adopt the same procedure that is used for their physical servers. Install a backup agent inside each VM and backup over the network, preferably a dedicated segment.

Another way to backup is to attach a tape unit to the VMware ESX Server and configure a VM to be the backup server with exclusive access to the tape unit. All the VMs are then installed with an agent that backs up to the VM backup server utilizing one of the available virtual nets. The files are moved to the backup server at memory transfer speed.

2.6 VMware ESX Server technical requirements

For information about supported hardware, download the VMware ESX Server *Systems Compatibility Guide* from the VMware Web site:

`http://www.vmware.com/products/server/esx_specs.html`

The technical requirements that follow are valid for VMware ESX Server V2.1.

2.6.1 Minimum server requirements

You need the following hardware and system resources to install and use ESX Server:

- At least two, and up to sixteen, processors:
  - Intel® 700MHz Pentium® III Xeon and above for ESX Server
  - Intel® 900MHz Pentium® III Xeon and above for Virtual SMP
- 512MB RAM minimum
- Two or more Ethernet controllers

  Supported controllers include:

  - Broadcom® NetXtreme 570x Gigabit controllers
  - Intel PRO/100 adapters
  - Intel PRO/1000 adapters
  - 3Com® 9xx based adapters

  Note: For best performance and security, use separate Ethernet controllers for the service console and the VMs.

- A SCSI adapter, Fibre Channel adapter or internal RAID controller

  The basic SCSI adapters supported are:

  - Adaptec®
  - LSI Logic
  - most NCR™/ Symbios™ SCSI adapters

  The RAID adapters supported are:

  - HP® Smart Array
  - Dell® PercRAID (Adaptec RAID and LSI MegaRAID)
  - ServeRAID™
  - Mylex® RAID

  The Fibre Channel adapters that are supported are:

  - Emulex™
  - QLogic™

  The supported SCSI controllers are:

  - Adaptec® Ultra-160 and Ultra-320
  - LSI Logic Fusion-MPT
  - most NCR/Symbios™ SCSI controllers.

  The supported RAID controllers are:

  - HP® Smart Array
  - Dell® PercRAID (Adaptec RAID and LSI MegaRAID)
  - IBM® (Adaptec) ServeRAID
  - Mylex RAID

  The supported Fibre Channel adapters are:

  - Emulex™
  - QLogic™ host-bus adapters (HBAs).

- A SCSI disk, Fibre Channel LUN or RAID LUN with unpartitioned space

  In a minimum configuration, this disk or RAID is shared between the service console and the VMs.

2.6.2 Recommended for enhanced performance

The following items are recommended for enhanced performance:

- A second disk controller with one or more drives, dedicated to the VMs
- Sufficient RAM for each VM and the service console
- Dedicated Ethernet cards for network-sensitive VMs

The lists above outline a basic configuration. In practice, you may use multiple physical disks, which can be SCSI disks, Fibre Channel LUNs or RAID LUNs. For best performance, all of the data used by the VMs should be on the physical disks allocated to VMs. Therefore, these physical disks should be large enough to hold disk images that will be used by all the VMs.

Similarly, you should provide enough RAM for all of the VMs plus the service console. For background on the service console and how to calculate the amount of RAM you need, refer to the *VMware ESX Server Administration Guide*.

Note: To ensure the best possible I/O performance and workload management, VMware ESX Server provides its own drivers for supported devices. Be sure that the devices you plan to use in your server are supported. For additional detail on I/O device compatibility, download the VMware ESX Server *I/O Adapter Compatibility Guide* from the VMware Web site at:

`http://www.vmware.com/products/server/esx_specs.html`

VMware ESX Server VMs can share a SCSI disk with the service console. For enhanced disk performance, you can configure the VMs to use a SCSI adapter and disk separate from those used by the service console. You should make sure enough free disk space is available to install the guest operating system and applications for each VM on the disk that they will use.

2.6.3 Maximum physical machine specifications

The following specifications detail the maximum hardware capabilities of VMware ESX Server.

- Storage
  - 16 host bus adapters per ESX Server system
  - 128 logical unit numbers (LUNs) per storage array
  - 128 LUNs per ESX Server system
- VMware Virtual Machine File System (VMFS)
  - 128 VMFS volumes per ESX Server system
  - Maximum physical extents per VMFS volume
    - VMFS-2 volumes: 32 physical extents
    - VMFS-1 volumes: 1 physical extent
    - 2TB per physical extent
  - Maximum size for a VMFS volume
    - VMFS-2 volumes: approximately 64TB, with a maximum of 2TB for each physical extent
    - VMFS-1 volumes: approximately 2TB
- CPU:
  - 16 physical processors per system
  - 80 virtual CPUs in all virtual machines per ESX Server system

    Any VM configured for virtual SMP uses two virtual CPUs.
- Memory:
  - 64GB of RAM for each ESX Server system

  - Up to 8 swap files, with a maximum file size of 64 GB for each swap file
- Adapters:
  - 64 adapters of all types, including storage and network adapters, per system
  - Up to 8 Gigabit Ethernet or 16 10/100 Ethernet ports per system
  - Up to 32 VMs per virtual network device

2.7 What is new in VMware ESX Server V2.5

Following are a few of the new features in version 2.5 of VMware ESX:

- Boot from SAN
  - Version 2.5 will allow you to install ESX Server in a diskless operation mode where the ESX Server boot partition is hosted on a SAN. This greatly enhances support for common blade and rack mount configurations.
- Support for additional storage and server hardware
- Improved support for raw LUNs as virtual disks using Raw Disk Maps (RDMs) and pass-through RDMs
  - Allows raw disks to be managed as easily as virtual disk files
  - Pass-through RDMs enable SAN replication software to run inside of virtual machines
  - Pass-through RDMs enable backup offloading solutions so that ESX Servers can be backed up faster without any load placed on the virtual machines or on the Service Console
  - Allows REDO logs to be added to raw disks
- Improved SSH security
  - Default Secure Login Client is now using SSH Protocol version 2.
- Support for LSI 1020/1030 controllers in shared mode, with mirrored disks
  - This fixes the known current issue where the vmkernel does not load when the mirror is synchronizing. Version 2.5 supports using a LSI 1020/1030 SCSI controller in shared mode with mirrored disks.
- Enhanced support for scripted installations
  - Allows third party systems management products to remotely install and configure ESX Servers
- Improved Support for Clustered virtual machines
  - MSCS in virtual machines using shared disk more reliable
- Additional support for DR backups or virtual disks
  - Virtual disk snapshot scripts added to the Service Console to enable crash consistent backups of entire virtual machines
- Support for additional guest operating systems:
  - FreeBSD 4.9 (Uniprocessor mode only)
  - Windows XP Pro (Uniprocessor mode only)
  - Windows 2003 Small Business Server

2.8 IBM ServerProven®

The IBM ServerProven program was first announced in 1996 to ensure that leading vendor products were compatible with the then Netfinity® servers and today, the xSeries range. When a product enters the ServerProven program, it must pass a rigorous testing and certification process. Once this process has completed successfully, the vendor will achieve the IBM ServerProven status for their product. The IBM ServerProven program includes IBM and non-IBM hardware, software and middleware.

When you see the ServerProven logo, you can be assured that the product displaying this emblem has been tested and certified to run reliably on the IBM xSeries Servers. This allows IBM customers to purchase a complete, verified and tested solution. While ServerProven reduces installation problems and the need for post-installation support, ultimately warranty and software obligations are the responsibility of the vendor, not IBM.

For additional information about IBM ServerProven program participants, refer to the following Web site:

http://www.pc.ibm.com/us/compat/serverproven

For additional information about the IBM ServerProven program, refer to following Web site:

http://www.pc.ibm.com/ww/eserver/xseries/serverproven

3

IBM @server xSeries 445 with VMware ESX Server

In this chapter, we focus on the x445-VMware ESX Server relationship. The following topics are discussed:

3.1 IBM xSeries 445 overview and key features

The next-generation scalable enterprise server, the xSeries 445, powered by Enterprise X-Architecture, leads the market for 8-way and 16-way processing with XpandOnDemand processor and I/O scalability, high performance, and OnForever™ availability for mission-critical database, ERP and CRM solutions. See Figure 3-1.

Figure 3-1 The xSeries 445 server

The key features of the x445 are:

- XA-32 second generation chipset
- Models with the Intel Xeon MP processors, up to 3.0 GHz and 4 MB L3 cache, upgradable to 4-way, 8-way, 16-way, and 32-way
- Entry-level models with Intel Xeon DP, up to 3.0 GHz, upgradable to 4-way
- Active Memory with Memory ProteXion, memory mirroring, hot-swap and hot-add memory
- XceL4 Server Accelerator Cache, now 64 MB per 4-way (128 MB per 8-way system)
- 2 GB DDR memory standard, expandable to 64 GB per 8-way system
- Six 64-bit Active PCI-X slots per chassis
- Integrated LSI Logic 1030 dual-channel Ultra320 SCSI with integrated RAID-1 support
- Integrated dual Gigabit Ethernet
- Connectivity to an RXE-100 external PCI-X Remote eXpansion Enclosure for an additional 12 PCI-X slots which can optionally be shared between two x445s
- Remote Supervisor Adapter standard

  Future plans include making the x445 upgradable to the Remote Supervisor Adapter II.

3.1.1 X-Architecture

By building robust capabilities into xSeries servers and working on new technologies for the future, IBM X-Architecture technology addresses these requirements and helps prepare you for the next wave of e-business. By modifying and using proven server technologies from the other successful IBM product lines, the xSeries servers are able to leverage these through IBM X-Architecture technology.

As introduced in 1998, IBM X-Architecture technology outlined capabilities for performance leadership, technology innovation, increased reliability and business-critical service and support capabilities. We continue to build IBM X-Architecture technology into our systems and set the bar for Intel processor-based servers across the industry.

The IBM xSeries technology agenda is straightforward. We leverage proven IBM technologies to bring enterprise capabilities to the Intel processor-based server platform. The X-Architecture technologies can be grouped into the following categories.

- Core logic of advanced technologies that optimize system performance and scalability
- Enterprise storage solutions designing subsystems and I/O bandwidth to match increasing storage and network requirements
- Availability of system, storage and clustering technologies to increase availability
- Systems management technologies for comprehensive local and remote server and storage management within heterogeneous environments
- Technology-enabled service and support tying advanced system and management technologies directly to service and support infrastructures

In short, IBM X-Architecture technology is the IBM design blueprint for the Intel processor-based xSeries servers. We accelerated the advancement of xSeries servers to meet client demands, as well as those of e-businesses. This has introduced the next step forward, Enterprise X-Architecture.

More information about X-Architecture technology can be found at:

http://www.pc.ibm.com/us/eserver/xseries/xarchitecture/principals.html

3.1.2 Enterprise X-Architecture

IBM continues to build on the X-Architecture blueprint with Enterprise X-Architecture.

Enterprise X-Architecture technologies yield revolutionary advances in the I/O, memory and performance of xSeries servers. This server design creates a flexible pay as you grow approach to buying high-end 32-bit and 64-bit xSeries systems that can be scaled quickly, easily and inexpensively. Enterprise X-Architecture technology is designed to achieve unparalleled levels of availability, scalability and performance in industry-standard enterprise computing.

Enterprise X-Architecture technology enables the following capabilities:

- XpandOnDemand scalability
- System partitioning
- PCI-X I/O subsystem
- Active PCI-X
- Remote I/O
- Active Memory
- High-speed (DDR) memory
- Memory ProteXion
- Chipkill™ memory
- Memory mirroring
- Hot-add/hot-swap memory
- XceL4 server accelerator cache

These features deliver application flexibility and new tools for managing e-business. They bring to industry-standard servers the kinds of capabilities formerly available only to users of mainframes and other high-end systems. Combined with existing X-Architecture technologies, these innovations result in unmatched economies of scale and new levels of server availability and performance.

Much of the Enterprise X-Architecture offering is delivered through IBM-developed core logic. IBM has more proven product technology and expertise in designing core logic than anyone

else in the industry. The IBM XA-32 and XA-64 families of chip sets for 32-bit and 64-bit industry-standard servers contain advanced core logic, which is the heart of a computer system. Core logic determines how the various parts of a system, processors, system cache, main memory, I/O, and so on, interact. These new copyists bring to the next generation of industry-standard servers key advantages, including modular system nodes, system partitioning, high-performance clustering and high-speed remote PCI-X I/O support.

The Enterprise X-Architecture paradigm sets IBM xSeries servers apart in the industry while maintaining the advantages of compliance with industry standards for processors, memory, I/O, storage and software. Enterprise X-Architecture also establishes a revolutionary new economic model for servers through its flexible modular design. The x445 Enterprise X-Architecture-based server allows customers to XpandOnDemand, the ability to grow when you need to, using the existing infrastructure.

More Information on Enterprise X-Architecture technology can be found at:

http://www.pc.ibm.com/us/eserver/xseries/xarchitecture/enterprise

3.1.3 Symmetric multiprocessing (SMP)

Symmetric multiprocessing (SMP) designs are currently being challenged by the demand for additional Intel processors within a single system. In SMP, CPUs have equal access to a single, shared memory controller and memory space. Because processor speed has dramatically increased to be much faster than memory, for most CPU intensive workloads the single front-side bus and single memory controller are often a bottleneck, resulting in excessive queuing delays and long memory latencies when under heavy load.

Processor instruction and data caches have been introduced to help reduce the likelihood of a memory controller bottleneck. A performance boost is often obtained from using large caches because of two reasons:

- Caches have lower latency and faster access time than a memory access. Because a cache can be accessed quicker than memory, the processor waits less time before receiving a response from a request.
- Caches also improve performance because they reduce queuing time of accesses which miss the caches and require a physical memory access.

  For most commercial applications, process cache hit rates are usually greater than about 70%. In this case, the cache greatly reduces memory latency because most processor memory requests are serviced by the faster cache. The caches act as filters and reduce the load on the memory controller, which results in lower queuing delays, waiting in line, at the memory controller, thereby speeding up the average memory access time.

Another bottleneck in many SMP systems is the front-side bus. The front-side bus connects the processors to the shared memory controller. Process to memory requests travel across the front-side bus. It can become overloaded when three or more high speed CPUs are added to the same bus. This, in turn, leads to a performance bottleneck and lower system scalability. Large processor caches also help improve performance because they assist in filtering many of the requests that must travel over the front-side bus. A processor cache-hit does not require a front-side bus for memory transaction.

However, even with a large L3 cache, the number of memory transactions which miss the cache is still so great that it often causes the memory controller to bottleneck; this happens when more than three or four processors are installed in the same system.

3.1.4 NUMA architecture

Non-uniform Memory Access (NUMA) is an architecture designed to improve performance and solve latency problems inherent in large SMP systems with more than four processors. The x445 implements a NUMA-based architecture and can scale up to 32 processors using multiple *SMP Expansion Modules*.

The SMP Expansion Modules each contain up to four CPUs and sixteen memory DIMMs. Each module also contains a dedicated local memory controller and 64 MB XceL4 Level 4 cache. The additional fourth level of cache greatly improves performance for the four processors in the SMP Expansion Module because it is able to respond to a majority of processor-to-memory requests, thereby reducing the load on the memory controller and speeding up average memory access times.

As shown in Figure 3-2, each SMP Expansion Module can be connected to other SMP Expansion Modules using three independent 3.2 GBps scalability links shown in red. These scalability links mirror front-side bus operations to all other SMP Expansion Modules in the system and are key to building large multiprocessing systems.

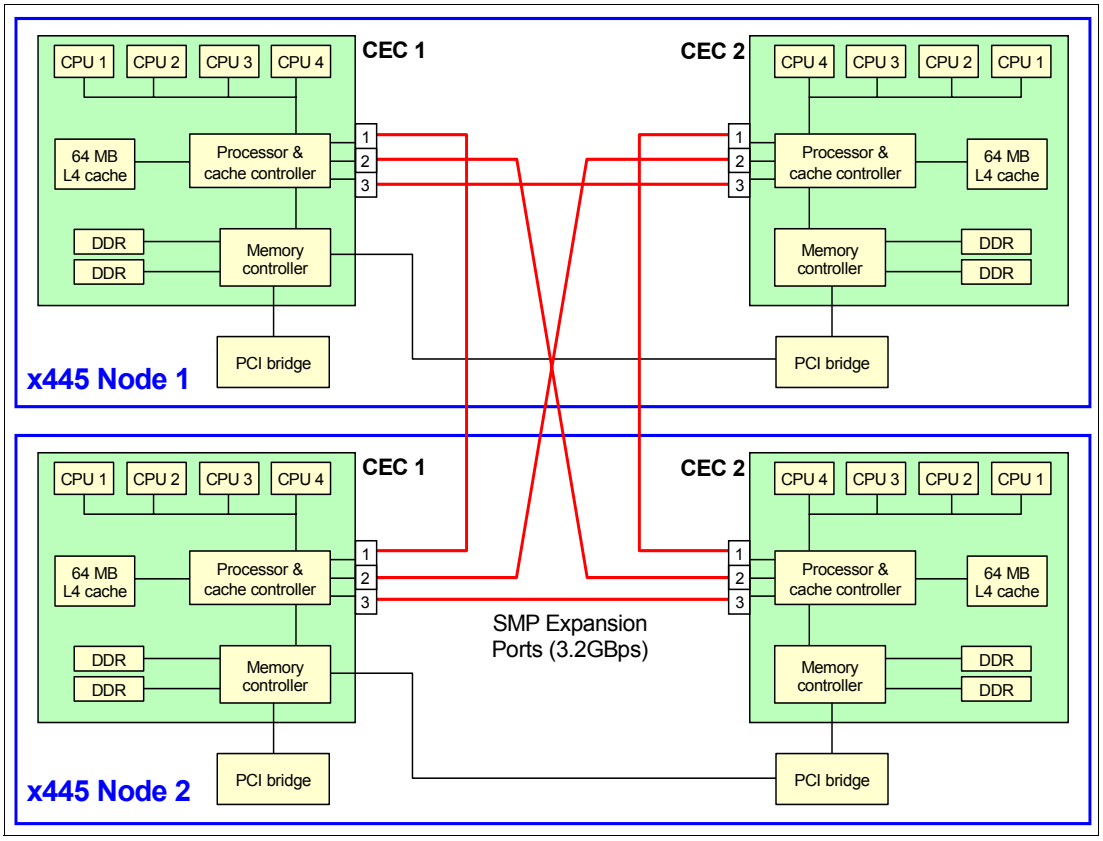


Figure 3-2 Two 8-way x445 servers connected as a 16-way configuration

By mirroring transactions on the front-side bus across the scalability links to other processors, the x445 is able to run standard SMP software. All SMP systems must perform processor-to-processor communication, known as *snooping*, to ensure all processors receive the most recent copy of requested data. Because any processor can store data in a local cache and modify that data at any time, all processor data requests must first be sent to every other process in the system so that each processor can determine if a more recent copy of the requested data is in that processor cache.

Snooping traffic is an important factor affecting performance and scaling for all SMP systems. The overhead of this communication becomes greater with an increase in the number of processors in a system. Also, faster processors result in a greater percentage of time spent performing snooping because the speed of the communications does not improve as the processor clock speed increases, because latency is largely determined by the speed of the front-side bus.

It is easy to see that increasing the number of processors and using faster processors results in greater communication overhead and memory controller bottlenecks. But unlike traditional SMP designs, which send every request from every processor to all other processors, greatly increasing snooping traffic, the x445 has a more optimal design. The XceL4 cache in the x445 improves performance because it filters most snooping operations.

The IBM XceL4 cache improves scalability with more than four processors because it also caches remote data addresses. Before any processor request is sent across the scalability link to a remote processor, the memory controller and cache controller determine whether the request should be sent at all. To do this, each cache controller keeps a directory of all the addresses of all data stored in all remote processor caches. By checking this directory first, the cache controller can determine if a data request must be sent to a remote processor and only send the request to that specific SMP Expansion Module where the processor caching the requested data is located.

The majority of data requests are not found in the XceL4 cache and can be sent directly to the memory controller to perform memory look-up without any remote processor-to-processor communication overhead. The directory-based coherency protocol used by the XceL4 cache greatly improves scalability of the x445 because processor-to-processor traffic is greatly reduced.

Performance is also improved in an 8-way configuration. With two SMP Expansion Modules, the x445 has two EXA memory controllers and two 64MB XceL4 caches. These two memory controllers share the memory request load from all eight processors and greatly reduce memory contention typically found in single-memory-controller SMP designs. For example, think of walking into a very busy airport to get your ticket. Would it not be more efficient to see multiple ticket agents available to process the large number of travelers waiting ahead of you? Also, if there were an increase in the number of travelers, would you not want to see more ticket agents to absorb the load? Similarly, the x445 employs a new ticket agent, memory controller, for every four processors added to the system. Therefore, a 16-way configuration has four EXA memory controllers and a total of four 64MB XceL4 caches to absorb the additional processor to memory request load.

By using this cache and providing fast local memory access to all processors, typical bottlenecks that are associated with large SMP systems using a single memory controller design are eliminated. Operating systems such as Windows 2000 perform well on this type of architecture, but dramatic performance gains can be achieved when the operating system is NUMA-aware, as with VMware ESX Server.

The NUMA architecture of the x445 server and its support by VMware ESX Server is further discussed in 3.2.3, “VMware NUMA integration” on page 59.

3.1.5 XpandOnDemand

One of the challenges of optimizing server performance is to provide memory and I/O subsystems that allow the new processor architectures to operate fully optimized. Traditional industry-standard server designs begin to encounter performance bottlenecks beyond 4-way scaling, due to memory controller and front-side bus congestion. The EXA design provides

support for advanced I/O and memory architectures and includes a high-speed, shared-cache architecture.

New levels of scalability for industry-standard servers are achieved with the EXA platform using enhanced, high-performance SMP building blocks to allow the effective scalability beyond 4-way SMP. These technologies provide scalability from 2-way to 4-way, 8-way and 16-way systems, using the x445 server.

You can start out with one 4-way SMP node with six PCI-X slots. If this is all the computing power you need for that server, you have saved yourself expense by not buying a more expensive 8-way or larger system. If you need more processing power, you only have to add a second SMP Expansion Module to the existing x445 server to create an 8-way system.

To grow to 16-way without having to deploy a brand new server with operating system and applications, you can connect a second x445 8-way server to the original one to form a single-image 16-way system with 12 PCI-X slots. Do you need more PCI-X slots? Simply add the RXE-100 external I/O enclosure. This total configuration is shown in Figure 3-3.

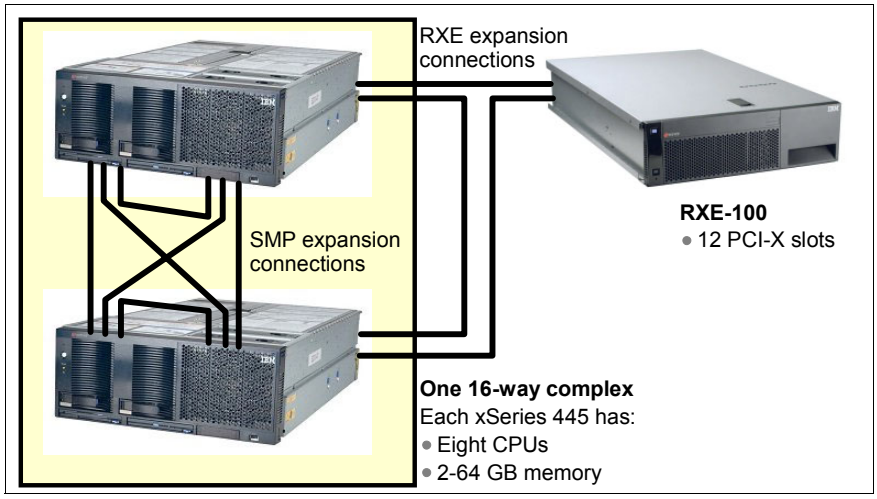


Figure 3-3 x445 XpandOnDemand to 16-way, 128 GB of RAM, and 18 PCI-X slots

With the EXA approach, adding a second SMP Expansion Module doubles the number of processors and the amount of system memory, but it does so without the bottlenecks. With the EXA design, you double the number of processors, copyists, memory, memory controllers, and front-side buses, which helps spread the burden of the data traffic. This occurs with two SMP Expansion Modules. Connecting two x445 servers together translates into more system bandwidth and more system capacity. The contention and resource problems increase in a 16-way traditional SMP system but these are lessened with the x445 EXA design.

XpandOnDemand scalability represents revolutionary advances for Intel-based servers. You will be able to scale up server performance without any downtime while your initial purchase is secured.

3.1.6 SMP Expansion Module

The SMP Expansion Module, as shown in Figure 3-4, contains many of the major components for the x445 server:

- CPUs (two or four Xeon MP or two Xeon DP processors)
- XceL4 level four cache

- Memory controller
- System RAM (16 sockets)

The SMP Expansion Module is also referred to as the Central Electronics Complex (CEC). These terms are interchangeable.

All x445 servers ship with at least two processors and 2 GB RAM. The SMP Expansion Modules provide the NUMA-based architecture as discussed in 3.1.4, “NUMA architecture” on page 53. When configured as a two-way or 4-way Xeon MP server, the x445 will contain one SMP Expansion Module. When configured as an 8-way server, the x445 is configured with two SMP Expansion Modules. The two SMP Expansion Modules are installed one above the other in the server.

Note: For Xeon DP models, when configuring the x445 using the Xeon DP processors, a two-way Xeon DP x445 server will contain one SMP Expansion Module. When configured using four Xeon DP processors, the x445 contains two SMP Expansion Modules.

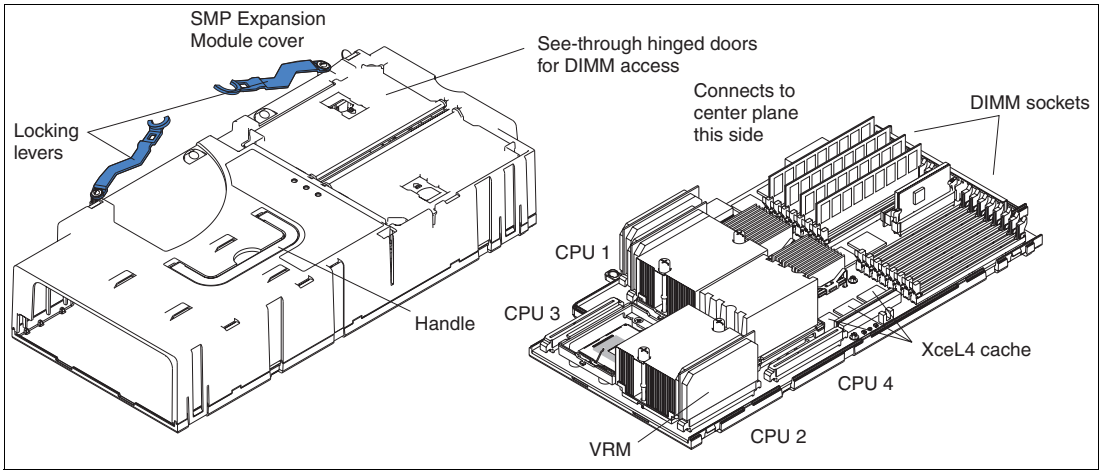


Figure 3-4 xSeries 445 SMP Expansion Module

Each SMP Expansion Module is also equipped with LEDs for light path diagnostics. LEDs are available for each memory DIMM and for each processor to aid in the quick identification of failed components. LEDs are also located on the top of the SMP Expansion Module to identify the operational components of Active Memory.

3.1.7 Memory implementation

The x445 implements NUMA by placing the DIMM sockets in the SMP Expansion Modules, meaning that in a 16-way configuration, there are four local regions of memory.

The sixteen sockets in each SMP Expansion Module are divided into four banks and two ports. A bank is the pair of DIMMs required for two-way interleaving. The ports and banks are shown in Figure 3-5 on page 57.

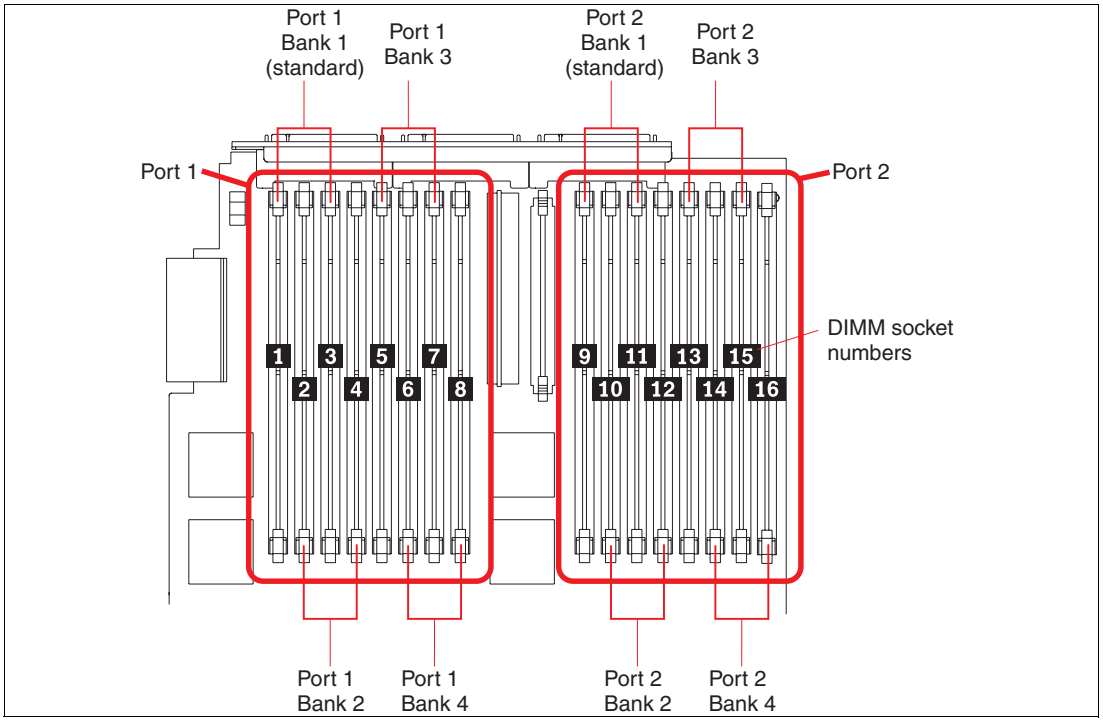


Figure 3-5 Memory DIMM sockets on the x445 SMP Expansion Module

There are a number of advanced features implemented in the x445 memory subsystem, collectively known as *Active Memory*:

- Memory ProteXion

  Memory ProteXion, also known as *redundant bit steering*, is the technology behind using redundant bits in a data packet to provide backup in the event of a DIMM failure.

  In the event that a chip failure on the DIMM is detected, the memory controller can re-route data around that failed chip through the spare bits. This is similar to the hot-spare drive of a RAID array. It can do this automatically without issuing a Predictive Failure Analysis® (PFA) or light path diagnostics alert to the administrator. After the second DIMM failure, PFA and light path diagnostics alerts would occur on that DIMM as normal.

- Memory scrubbing

  Memory scrubbing is an automatic daily test of all the system memory that detects and reports memory errors that might be developing before they cause a server outage.

  When a bit error is detected, memory scrubbing determines whether or not the error is recoverable. If it is recoverable, Memory ProteXion is enabled and the data that was stored in the damaged location is rewritten to a new location. The error is then reported so that preventative maintenance can be performed. As long as there are enough good locations to allow the proper operation of the server, no further action is taken other than recording the error in the error logs.

  If the error is not recoverable, then memory scrubbing sends an error message to the light path diagnostics, which then turns on the proper lights and LEDs to guide you to the damaged DIMM. If memory mirroring is enabled, then the mirrored copy of the data in the damaged DIMM is used until the system is powered down and the DIMM is replaced. If hot-add is enabled in the BIOS then no rebooting is required and the new DIMM will be enabled immediately.

- Memory mirroring

  Memory mirroring is roughly equivalent to RAID-1 in disk arrays, in that memory is divided in two ports and one port is mirrored to the other half. If 8 GB is installed, then the operating system sees 4 GB once memory mirroring is enabled. It is disabled in the BIOS by default. Since all mirroring activities are handled by the hardware, memory mirroring is operating system independent.

- Chipkill memory

  Chipkill is integrated into the XA-32 second-generation chipset and does not require special Chipkill DIMMs. Chipkill corrects multiple single-bit errors to keep a DIMM from failing. When combining Chipkill with Memory ProteXion and Active Memory, the x445 provides very high reliability in the memory subsystem. Chipkill memory is approximately 100 times more effective than ECC technology, providing correction for up to four bits per DIMM (eight bits per memory controller), whether on a single chip or multiple chips.

  If a memory chip error does occur, Chipkill is designed to automatically take the inoperative memory chip offline while the server keeps running. The memory controller provides memory protection similar in concept to disk array striping with parity, writing the memory bits across multiple memory chips on the DIMM. The controller is able to reconstruct the missing bit from the failed chip and continue working as usual.

  Chipkill support is provided in the memory controller and implemented using standard ECC DIMMs, so it is transparent to the OS.

- Hot-swap and hot-add memory

  Two new features have been added to the x445: hot-swap and hot-add memory. These technologies are mutually exclusive. Enable one or the other through the system BIOS followed by a reboot.

  - Hot-add memory

    The hot-add memory feature enables you to add DIMMs without turning off the server. This feature is supported only in those servers using Windows Server 2003 Enterprise or Datacenter Editions. The current implementation is restricted such that one of the two ports must be completely empty before you can add DIMMs. DIMMs must be added in pairs, and you can also only add memory to the upper SMP Expansion Module.

  - Hot-swap memory

    The hot-swap memory feature allows you to replace DIMMs of the same type, size, and clock speed as the failed DIMM without turning off the server. Hot-swap memory is operating system independent. Before you can use the hot-swap feature, you must enable memory mirroring. You can also only replace DIMMs that are installed in the upper SMP Expansion Module.

3.1.8 IBM ServerProven

IBM technicians have worked side by side with their VMware counterparts to ensure that the IBM x445 server and its most popular hardware and software configurations have been ServerProven with the VMware ESX Server operating system. This is essential to ensure that the IBM VMware Program has industry-leading hardware and software solutions to provide the highest level of availability for today's mission-critical environments.

For a list of supported hardware, refer to the VMware xSeries ServerProven xSeries Compatibility Web page:

http://www.pc.ibm.com/us/compat/nos/vmware.html

3.2 x445 features beneficial to VMware ESX Server

This section discusses the features of the x445 when combined with VMware ESX Server.

3.2.1 Enterprise X-Architecture

Based on the features described in 3.1.2, “Enterprise X-Architecture” on page 51, the x445 provides superior performance and scalability with the flexibility of pay as you grow. These factors provide organizations with an excellent ROI, and when coupled with VMware ESX Server will help to reduce the TCO.

3.2.2 Remote eXpansion Enclosure (RXE)

The x445 8-way server provides six PCI-X slots and the x445 16-way server provides twelve PCI-X slots. In most situations, this is sufficient for VMware ESX Server, however, if there is need for more slots to cater for additional network interfaces or host bus adapters, the remote expansion enclosure described in 3.1.5, “XpandOnDemand” on page 54 will provide twelve additional PCI-X slots.

This feature allows an organization to expand the capability of the server on demand, and without needing to implement additional servers.

3.2.3 VMware NUMA integration

IBM and VMware have worked closely together to ensure that ESX Server is highly optimized for Enterprise X-Architecture platforms. IBM is a founding member of the VMware Preferred Hardware Provider program. ESX Server has supported the eServer x440 and x445 in configurations with up to 16 processors since the release of VMware ESX Server 1.5.1. This redpaper, however, focuses on enhancements added to improve NUMA performance in ESX Server V2.x. Customers using older releases on Enterprise X-Architecture hardware should consider upgrading to version 2 in order to take advantage of these new features.

The intelligent, adaptive NUMA scheduling and memory placement policies in VMware ESX Server V2.x can manage all VMs transparently, so that administrators do not need to deal with the complexity of balancing VMs between nodes by hand. However, manual override controls are also available and administrators with advanced skills can still optimize their systems as they see fit.

It is important to note that these optimizations work seamlessly regardless of guest OS type. ESX Server will provide powerful, transparent NUMA support even to guests that do not support NUMA hardware, such as Windows NT 4.0. This unique feature of VMware ESX Server allows clients to take advantage of cutting-edge new hardware, even when tied to legacy operating systems.

VMware NUMA Optimization Algorithms

The following information is provided as a technical insight into how VMware implements NUMA support on the x445.

Home nodes and initial placement

VMware ESX Server V2.x assigns each VM a home node when the VM begins running. A VM will only run on processors within its home node. Newly-allocated memory comes from the home node as well. Thus, if a VM’s home node does not change, the VM uses only local memory, avoiding the performance penalties associated with remote memory accesses to other NUMA nodes. Furthermore, because the VM only runs on a single node, it can take

maximum advantage of the high speed memory of the per-node XceL4 cache. New VMs are assigned to home nodes in a round robin fashion. The first VM goes to the first node, the second VM to the second node, and so on. This policy ensures that memory will be evenly used throughout all nodes of the system.

Several commodity operating systems, such as Windows 2003 Server, provide this level of NUMA support, which is known as *initial placement*. It might be sufficient for systems that only run a single workload, such as a benchmarking configuration, which does not change over the course of the system's uptime. However, initial placement is not sophisticated enough to guarantee good performance and fairness for a datacenter-class system that is expected to support changing workloads with an uptime measured in months or years.

To understand the weaknesses of an initial-placement-only system, consider the following example: An administrator starts four VMs. The system places two of them on the first node and two on the second node. Now consider what happens if both VMs on the second node are stopped, or if they simply become idle. The system is then completely imbalanced, with the entire load placed on the first node. Even if the system allows one of the remaining VMs to run remotely on the second node, it will suffer a serious performance penalty because all of its memory will remain on its original node.

Dynamic load balancing and page migration

To overcome these weaknesses, VMware ESX Server V2.x combines the traditional initial placement approach with a dynamic rebalancing algorithm. Periodically, every two seconds by default, the system examines the loads of the various nodes and determines whether it should rebalance the load by moving a virtual machine from one node to another. This calculation takes into account the relative priority of each virtual machine to guarantee that performance is not compromised for the sake of fairness.

The rebalancer selects an appropriate VM and changes its home node to the least-loaded node. When possible, the rebalancer attempts to move a VM that already has some memory located on the destination node. From that point on, the VM allocates memory on its new home node, unless it is moved again. It only runs on processors within the new home node.

Rebalancing is an effective solution to maintain fairness and ensure that all nodes are fully utilized. However, the rebalancer may need to move a VM to a node on which it has allocated little or no memory. In this case, the VM will incur a performance penalty associated with a large number of remote memory accesses. ESX Server V2.x can eliminate this penalty by transparently migrating memory from the virtual machine's original node to its new home node. The system selects a page, 4 KB of contiguous memory, on the original node and copies its data to a page in the destination node. The system uses the VM monitor layer and the processor's memory management hardware to seamlessly remap the VM's view of memory, so that it uses the page on the destination node for all further references, eliminating the penalty of remote memory access.

When a VM moves to a new node, ESX Server V2.x will immediately begin to migrate its memory in this fashion, usually at a rate of approximately 100 KB, 25 pages, per second. It adaptively manages this rate to avoid overtaxing the system, particularly when the VM has very little remote memory remaining or when the destination node has little free memory available. The memory migration algorithm also ensures that it will not move memory needlessly if a VM is moved to a new node for only a short period of time.

When all these techniques of initial placement, dynamic rebalancing, and intelligent memory migration work in conjunction, they ensure good memory performance on NUMA systems, even in the presence of changing workloads. When a major workload change occurs, for instance when new VMs are started, the system takes time to readjust, migrating VMs and

memory to new, optimal locations. After a short period of time, the system completes its readjustments and reaches a steady state.

Manual NUMA controls

Some administrators with advanced skills might prefer to control the memory placement and processor utilization by hand. This can be useful, for example, if a VM runs a memory-intensive workload, such as an in-memory database or a scientific computing application with a large dataset. Such an application can have performance improvements if 100% of its memory is allocated locally, while VMs managed by the automatic NUMA optimizations often have a small percentage (5-15%) of their memory located remotely. An administrator might also wish to optimize NUMA placements manually if the system workload is known to be simple and unchanging. For example, an eight-processor system running eight VMs with similar workloads would be easy to optimize by hand.

VMware ESX Server V2.x provides two sets of controls for NUMA placement, so that administrators can control both memory and processor placement of a VM. The ESX Server web-based Management User Interface allows the you to indicate that a VM should only use the processors on a given node through the Only Use Processors option and that it should only allocate memory on the desired node through the Memory Affinity option. If both of these are set before a VM starts, it only runs on the desired node and all of its memory is allocated locally. An administrator can also manually move a VM to another node after the VM has started running. In this case, the page migration rate of the VM should also be set manually, so that memory from the VM's previous node can be moved to its new node. The ESX Server V2.x documentation contains a full description of how to set these options.

Note that manual NUMA placement may interfere with the ESX Server resource management algorithms, which attempt to give each VM a fair share of the system's processor resources. For example, if ten VMs with processor-intensive workloads are manually placed on one node, and only two VMs are manually placed on another node, then it is impossible for the system to give all twelve VMs equal shares of the system's resources. You should take these issues into account when using manual placement.

Conclusion

VMware ESX Server V2.x's rich manual and automatic NUMA optimizations allow you to fully exploit the advanced scalability features of the Enterprise X-Architecture platform. By providing a dynamic, self-optimizing NUMA load balancer in conjunction with patented memory migration techniques, ESX Server can maintain excellent memory performance even in the face of changing workloads. If you need more information about the technical details of NUMA system configuration, consult the ESX Server documentation or the numa(8) man page on an ESX Server system.

3.3 Scaling up VMware ESX Server with xSeries 445

This section describes the advantages and disadvantages of implementing the VMware infrastructure using fewer high-end servers with the smallest number of VMware ESX Server installations possible. For more information about scale up versus scale out considerations, see *VMware ESX Server: Scale Up or Scale Out?*, REDP-3953.

The x445 stands alone as a scale up platform for VMware ESX Server with features such as:

- 16 processor support
- Enterprise X-Architecture
- Remote eXpansion Enclosure (RXE)
- Non-Uniform Memory Access (NUMA)

Although this topic discusses scale up, there are limitations to how far an Intel platform can scale. As a result, it is conceivable that organizations that can justify the capacity of an x445 most likely will implement several of these servers. This creates a hybrid environment, using a scale up approach as far as the architecture will accommodate, then scaling out with these larger servers to both achieve greater capacity and higher resiliency.

Figure 3-6 illustrates a typical 16-way configuration using xSeries 445 modules in a single system image. This configuration requires only one installation of VMware ESX Server hosting many virtual machines. For example, if each processor on a VMware ESX Server is capable of hosting four VMs, the following applies:

- A 4-way x445 using 4U of rack space could host 16 VMs
- An 8-way x445 using 4U of rack space could host 32 VMs
- A 16-way x445 using 8U of rack space could host 64 VMs

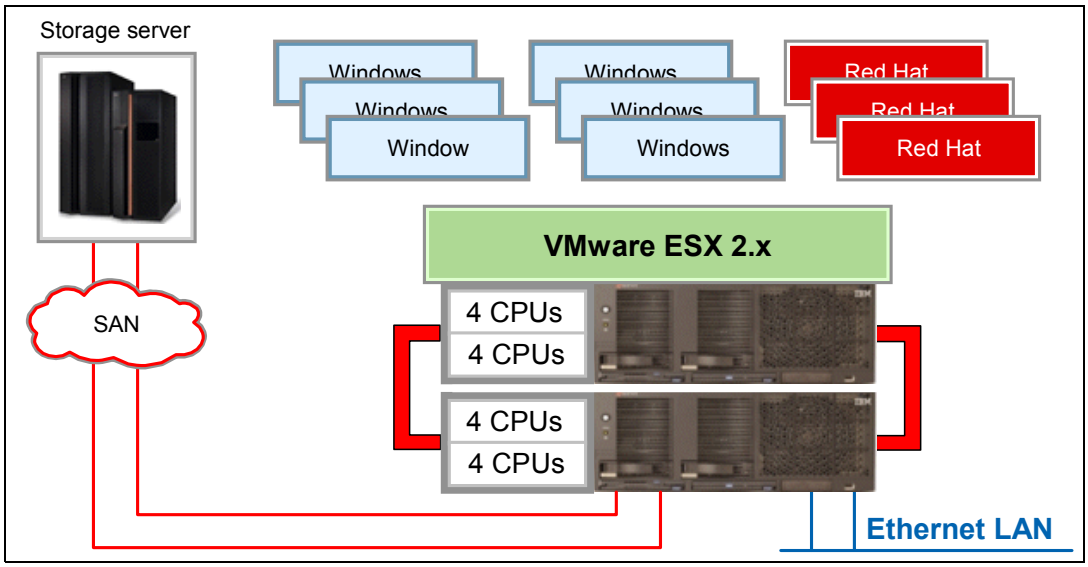


Figure 3-6 16-way x445 running a single system image

Table 3-1 lists some advantages and disadvantages of the scale up approach.

Table 3-1 Advantages / Disadvantages of scale up with VMware

Advantages	Disadvantages
► Lower Ethernet/SAN costs ► Automatic in-box resource management ► Fewer VMware ESX Server images to manage ► VMware VirtualCenter/VMotion not strictly required ► Lower software costs if not implementing VMware VirtualCenter/Vmotion ► Lower infrastructure costs (Ethernet / SAN switches)	► Higher hardware costs for servers ► Large high availability impact in the case of a node failure ► Fewer CPUs supported per rack ► Capacity added on demand in larger increments

Of course, these advantages and disadvantages are subject to different interpretations. One reader may rate the Fewer ESX Server images to manage as a very key advantage, while another reader may not consider having to deal with many ESX Server images an issue. In general, most organizations agree that ease of management is important, however the perceived value is different from one organization to another.

A typical concern with high-end configurations like these is performance and scalability. This is a real concern in most of the scenarios where a single application tries to benefit from all the resources available. In these cases, there is indeed a scaling penalty due to the fact that some applications simply cannot scale.

With VMware ESX Server implementations this is not usually a concern. In fact, we are not dealing with a single application running on the high-end system, but rather with multiple and OS images and applications running on the same high-end systems. Even though the virtualization layer is very efficient, you can add more VMs when there is available capacity. This translates to more workload for the server. Furthermore, the xSeries 445 leverages the NUMA architecture, which allows administrators to add CPU/Memory bandwidth as they add CPU's to the configuration. VMware ESX Server fully exploits the NUMA architecture of the xSeries 445. See section 3.1.4, “NUMA architecture” on page 53.

When determining the total cost of ownership for a server consolidation based on the scale up approach, you must also consider the cost of the environment. This includes floor and rack space, cooling, and power. Using the earlier example of four VMs per CPU, a fully-populated 42U rack would house ten 8-way x445 servers resulting in 320 VMs per rack. Compare this with a BladeCenter where six chassis can be housed in a 42U rack, consisting of 14 dual processor blades. Assuming the same capability of four VMs per CPU, a rack fully populated with VMware ESX Server blades could host 672 VMs.

Using x445 to implement a scale up approach to VMware ESX Server can avoid some software costs that are present in the scale out approach. When implementing many VMware ESX Servers, it is almost mandatory to utilize VMware’s VirtualCenter software to reduce the management costs associated with monitoring and deploying virtual machines. Fewer large servers can avoid the need for VMware VirtualCenter, although there are still advantages to using it in a scale up environment. See Chapter 5, “VMware VirtualCenter and VMotion” on page 77.

One of the major disadvantages of this approach is availability. While the technology that goes into the x445 makes this server very reliable, there is still a risk that the server may crash due to a hardware problem, or be disconnected due to network or power problems. It is understandable that organizations may be hesitant to use only this one solution. For this reason, appropriate planning should be done to mitigate the risk.

The following suggestions can help to reduce the risk to the organization when adopting the scale up approach:

- Implement more than one VMware ESX Server with shared storage, and consider an N+1 architecture. That is, ensure there is enough capacity across all servers to handle the workload if one server is made unavailable.
- Keep a copy of the VM configurations in a central location or on each VMware ESX Server.
- Use IBM Director, IBM Virtual Machine Manager, and VirtualCenter with Vmotion to monitor and manage the environment.
- In addition to considering mixing complementary workloads on the servers, also consider mixing workloads of varying criticality. For example, if a server is made unavailable, only some of the workload is mission critical.
- Initially consider only using VMware for development, testing, and low critical workloads.
- Consider limiting the scale up to eight processors for each ESX Server installation, rather than 16 processors.

4

IBM @server BladeCenter with VMware ESX Server

In this chapter we introduce the IBM BladeCenter and explain the benefits of using the BladeCenter to deliver a virtual server solution with VMware ESX.

Topics covered in this chapter are:

4.1 BladeCenter overview

Both VMware and BladeCenter allow for large-scale consolidation of server environments. The have been seen as somewhat competing technologies. In reality however, the BladeCenter compliments a VMware solution by further reducing the physical space required for physical servers, while increasing the management and availability of the physical server environment.

VMware is IBM ServerProven for both the HS20 dual Xeon processor blade and the HS40 quad Xeon MP processor blade, giving complete flexibility in a range from a small scale VMware implementation on a single HS20 through to an enterprise level implementation on many HS20 or HS40 blades. For further information, see:

http://www.pc.ibm.com/us/compat/nos/vmwaree.html

Also, if the power of individual two or four-processor blades is needed, then these can cohabit with blades running VMware in the same BladeCenter chassis, reducing the number of different physical server models and giving the BladeCenter benefits to all servers, whether virtualized or not.

The following sections discuss the IBM IBM @server BladeCenter.

4.1.1 Introduction to the IBM @server BladeCenter

Blade servers are a relatively new technology that has captured industry focus because of their high density, high power, and modular design, which can reduce cost. This cost reduction comes with a more efficient use of valuable floor space, reduced network and power infrastructure requirements, and simplified management.

The BladeCenter is a 7U modular chassis that takes advantage of advancements in server technology. It is capable of housing up to 14 functionally separate blade servers. The BladeCenter chassis allows individual blade servers to share resources such as power, switch, management and cooling modules.

The BladeCenter design combines:

- IBM Enterprise X-Architecture technology

  IBM Enterprise X-Architecture technology leverages proven innovative IBM technologies to build powerful, capable, reliable Intel processor-based servers.

  Enterprise X-Architecture technology includes features such as Predictive Failure Analysis (PFA) and Advanced System Management.

- Expansion capabilities

  Blades can be added to the BladeCenter unit as needed, up to a maximum of 14 blades.

  IBM blade servers have connectors for options used to add capabilities to the blade, such as an I/O expansion card to add a network interface, or a storage expansion unit to add SCSI hard disk drives. Note that companies such as Cisco, Brocade, QLogic and others have built products specifically for the IBM BladeCenters.

- Hot-swap capabilities

  The front bays on the BladeCenter unit are hot-swap blade bays. The rear bays on the BladeCenter unit are hot-swap module bays. You can add, remove, or replace blades or management, switch, power, or blower modules without removing power from the BladeCenter unit.

- Redundancy capabilities

  The redundant components in the rear of the BladeCenter unit enable continued operation if one of the components fails. Normally, the redundant power modules and blowers share the load. If one of the power modules or blowers fails, the non-failing power module or blower handles the entire load. You can then replace the failed blower or power module without shutting down the BladeCenter unit.

- Redundant network connection capabilities

  Configuring a pair of Ethernet switch modules in switch-module bays 1 and 2 identically provides support for Ethernet failover configured on blade servers. If blade server I/O expansion options can be configured for failover, configuring a pair of switch modules in switch-module bays 3 and 4 identically provides support for the failover configured on I/O expansion options.

  Other network-interface I/O expansion options, such as the IBM HS20 Fibre Channel (FC) Expansion Card, that has similar capability for redundant SAN connections.

- System-management capabilities

  The BladeCenter unit comes with a system-management processor in the management module. This system-management processor, in conjunction with the system-management firmware that is provided with the BladeCenter unit and the system-management processor in each blade server, enables you to remotely manage the BladeCenter unit, its components, and the blade servers.

  The management module also multiplexes the keyboard, mouse, and video ports and the USB port across the multiple blade servers.

  The system-management processor in each blade server provides blade server system monitoring, event recording, and alert capability.

- Network environment support

  The BladeCenter unit supports a minimum of one four-port 1 Gb Ethernet switch module, expandable to two Ethernet switch modules. Each switch module provides one internal connection to each blade server, up to 14 internal connections per switch module.

  The BladeCenter unit also supports two additional switch modules, for a total of four switch modules. The two additional switch modules support the network interface on the optional I/O expansion card installed on one or more blade servers in the BladeCenter unit.

  Note: The two additional switch modules must have the same I/O type, such as Fibre Channel, and must match the network interface on the optional I/O expansion cards in the blade servers.

  Each of these two additional switch modules provides one internal connection to the optional I/O expansion card, up to 14 internal connections per switch module.

These BladeCenter features alone can reduce the cost of deployment, reprovisioning, updating, and troubleshooting. The cost savings come from the fact that modern computing environments are often made up of hundreds of servers. With that many systems, even simple infrastructure, such as network cabling, can become very expensive. Blade-based computing reduces the amount of infrastructure required to support large numbers of servers. By integrating resources and sharing key components, costs are reduced and reliability, availability, and serviceability are increased. These features are also known as RAS.

RAS makes up three of the most important features in server design. RAS features are found in unprecedented levels in the BladeCenter. These factors help to ensure the integrity of the data stored on the blades:

- The blades are available when needed.
- Should a failure occur, diagnosis and repair of the failure happens with minimal inconvenience.

The following list describes some of the RAS features that the BladeCenter unit supports:

- Shared key components, such as power, cooling, and I/O
- All components serviced from the front or rear of the chassis
- Automatic error retry and recovery
- Automatic restart after a power failure
- Built-in monitoring for blower, power, temperature, voltage, and for module redundancy
- Remote system management through the management module
- Remote management module firmware upgrade
- Remote upgrade of blade server system-management processor microcode
- Predictive Failure Analysis (PFA) alerts
- Redundant components
  - Cooling fans (blowers) with speed-sensing capability
  - Power modules
- Hot-swap components
  - Cooling fans (blowers) with speed-sensing capability
  - Power modules
  - Management module
  - Switch modules
  - Blades
  - Media tray
- System automatic inventory at startup
- System error logging with error codes and messages

4.1.2 BladeCenter disk storage

The direct-attach storage capacity in blade-based computing solutions is limited by the very small nature of the blades themselves. This drawback has the potential to limit the applicability for blade-based computing. Fortunately, BladeCenter provides an alternative. BladeCenter blades can easily attach to a Fibre Channel SAN via the FC I/O expansion card and FC switches. This ability is critical for implementing solutions that require access to shared disk, like VMware VMotion for example.

The disk storage options available for BladeCenter blades are described in the following sections.

HS20 model 8832 blades

Any mix and match of these disk configurations can be implemented for complete flexibility, providing that the IDE and FC I/O expansion card implementation requirements are followed.

- 2 x 2.5-inch IDE disks

  These disks can be mirrored in firmware for resilience.

- 2 x U320 SCSI disks

  These disks use the blade SCSI sidecar which has two standard IBM xSeries hot-swap SCSI drive bays and an LSI U320 SCSI controller capable of implementing hardware RAID levels 0 or 1.

- 1 x Dual-Port Fibre Channel I/O expansion card

  This card, in conjunction with the BladeCenter FC switch or OPM, allows a blade to connect to a SAN. The card takes the space required by the second IDE disk, reducing the number of possible IDE disks to one.

HS20 model 8843 blades

The HS20 model 8843 has the following features:

- 2 x 2.5-inch SCSI disks

  These disks can be mirrored using the integrated Ultra320 dual channel SCSI controller for resilience.

- 1 x Dual-Port Fibre Channel I/O expansion card

  This card, in conjunction with the BladeCenter FC switch or OPM, allows a blade to connect to a SAN. Unlike the HS20 model 8832, the Fibre Channel card does not take the space required by the second SCSI disk, allowing for mirrored local disks and SAN connectivity.

HS40 blades

Any mix and match of these disk configurations can be implemented for complete flexibility, providing that the IDE and FC I/O expansion card implementation requirements are followed.

- 2 x 2.5" IDE disks

  These disks can be mirrored in firmware for resilience.

- 2 x U320 SCSI disks

  These disks use the blade SCSI sidecar which has two standard IBM xSeries hot-swap SCSI drive bays and an LSI U320 SCSI controller capable of implementing hardware RAID levels 0 or 1.

- 2 x Dual-Port Fibre Channel I/O expansion card

  This card, in conjunction with the BladeCenter FC switch or OPM, allows a blade to connect to a SAN. The cards take the space required by the each IDE disk. If two cards are used, then no IDE disks can be installed.

IBM BladeCenter Fibre Channel switch module

When the FC I/O expansion card is used, this switch module is inserted into the rear of the BladeCenter chassis to enable the blades to connect to a SAN. You can install one or two modules, depending upon the need for redundancy in the fabric. See Figure 4-1 for details.

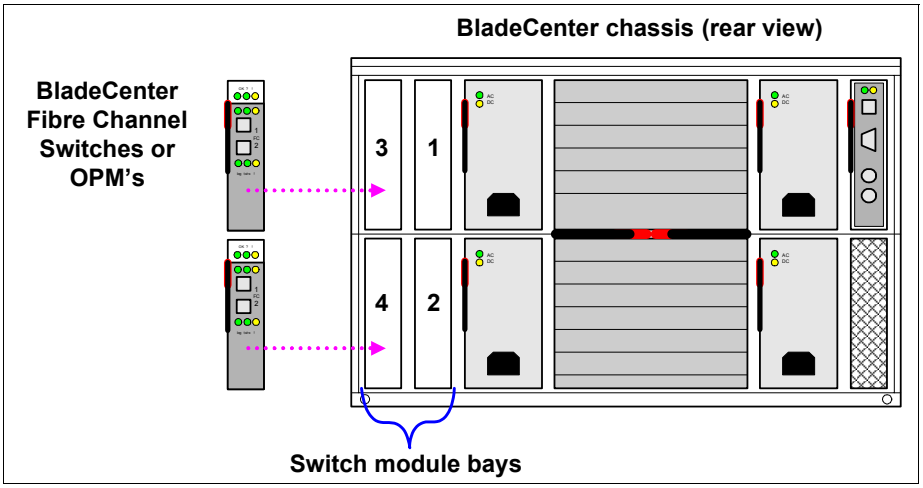


Figure 4-1 FC switch, illustrated, or OPM installation in a BladeCenter chassis

The IBM BladeCenter FC switch module features include these components:

- Ports
  - Two external ports to connect to storage devices or Storage Area Networks
  - Fourteen internal ports to connect to blade servers
- Protocols support various Fibre Channel protocols
- Scalability of a maximum 239 switches, depending on configuration
- Maximum user ports is 475,000, depending on configuration
- Media type is Small Form Pluggable (SFP) hot pluggable, optical transceivers
- Fabric port speed of 1.0625 or 2.125 Gigabits per second
- Maximum frame size of 2148 bytes with a 2112-byte payload
- Fabric point-to-point bandwidth is 212 or 424 MBs full duplex
- Fabric aggregate bandwidth is 64 MBs for a single switch
- Fibre Channel Cable Media
  - 9 micron single mode
  - 50 micron multi-mode
  - 62.5 micron multi-mod

Brocade SAN switch module for IBM BladeCenter

When the FC I/O expansion card is used, this switch module(s) is inserted into the rear of the BladeCenter chassis to enable the blades to connect to a SAN. One or two modules can be installed depending upon the need for redundancy in the fabric. See Figure 4-1 for details.

This recently announced FC SAN switch module for the BladeCenter enhances the value proposition of the BladeCenter particularly for environments where a Brocade fibre fabric already exists. Features of this switch include:

- Ports
  - Two external ports to connect to storage devices or Storage Area Networks
  - Fourteen internal ports to connect to blade servers
- Fabric switch delivering Brocade functions, including performance, manageability, scalability, and security to support demanding Storage Area Networks
- The Brocade Entry SAN Switch Module is ideal for customers that will integrate BladeCenter into small Brocade SANs while still providing the ability to grow as needed
- The Brocade Enterprise SAN Switch Module delivers full SAN fabric for large Storage Area Networks
- Fully upgradeable with Brocade's suite of Advanced Fabric Services
- Compatible with existing Brocade fabric, Fabric OS features, and Brocade SAN management tools

IBM BladeCenter Optical Passthrough Module (OPM)

When the FC I/O expansion card is used, this passthrough module is inserted into the rear of the BladeCenter chassis to enable the blades to connect to a SAN. One or two modules can be installed depending upon the need for redundancy in the fabric. See Figure 4-1 on page 69 for details.

The OPM provides the ability to transmit and receive network data traffic between all 14 blade bays. Unlike the FC switches, this module simply breaks out the blade network or I/O card ports into fourteen cables. The networking environments supported by the OPM are explained in the following sections.

Fibre Channel

To enable Fibre channel, the OPM can be inserted into switch module bays 3 and 4. In order for the OPM to function in bays 3 and 4, the Fibre Channel Expansion Card is required on the blade server.

An external FC switch with fibre ports is needed to connect the OPM and blades to the network. One port for each blade is required.

Gb Ethernet

To enable Ethernet, the OPM can be inserted into switch module bays 1, 2, 3 or 4.

If the OPM is inserted into switch module bays 1 or 2, the OPM interfaces with the integrated dual Gb Ethernet controllers on the blade server. For the OPM to function in bays 3 and 4, the Gb Ethernet Expansion Card is required on the blade server.

An external network switch with fibre ports is needed to connect the OPM and blades to the network. One port for each blade is required.

Connecting to DS4000

Using FC Switches to connect to a DS4000, for example, gives the options presented in Figure 4-2 and Figure 4-3 on page 72, depending upon whether there is an existing SAN fabric or not.

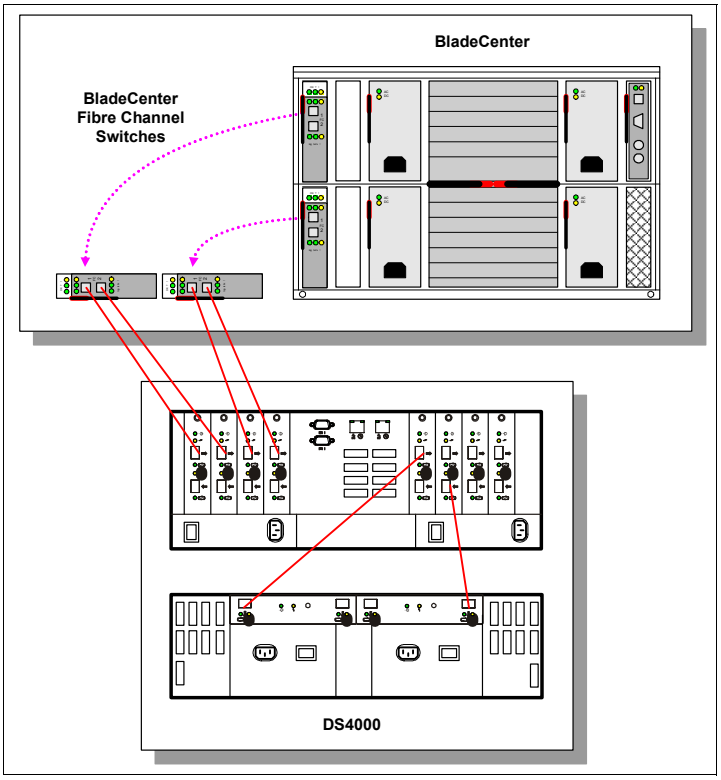


Figure 4-2 BladeCenter with FC switch modules direct to DS4000

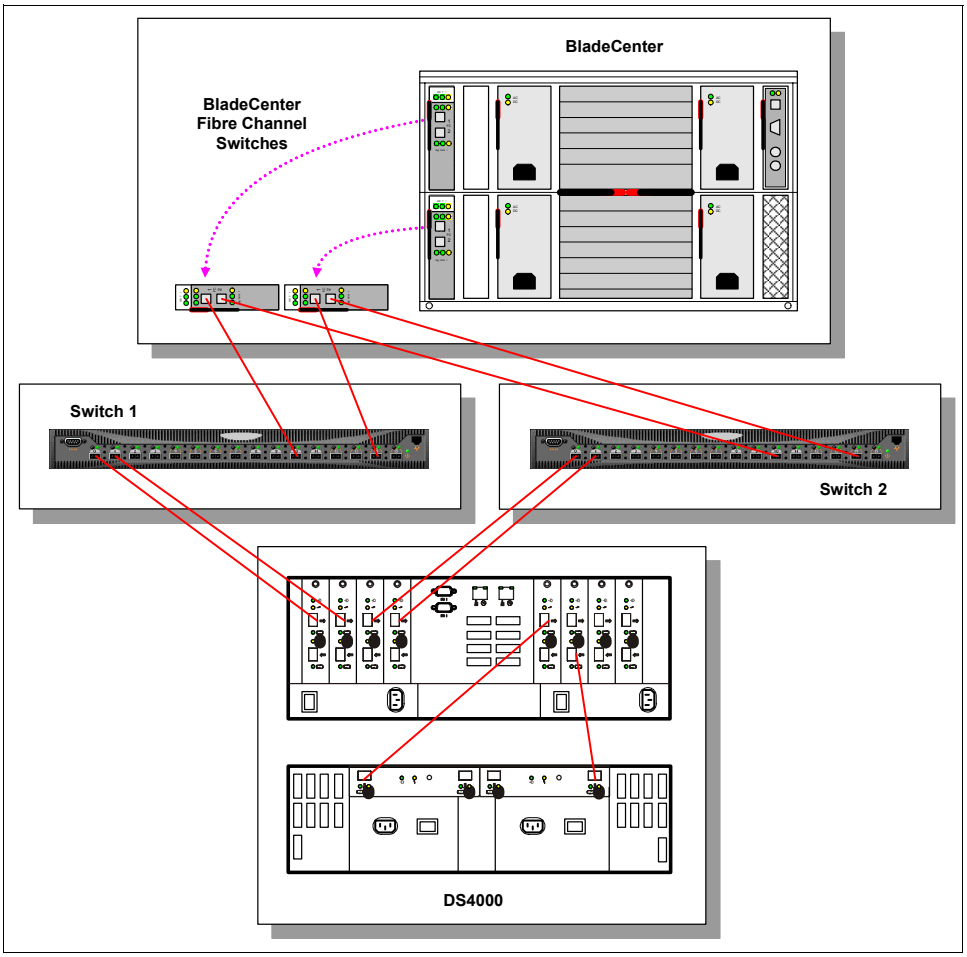


Figure 4-3 BladeCenter with FC switch modules and external FC switches (fabric) to DS4000

Because of the dramatic reduction and change in cabling when connecting a BladeCenter that may house up to fourteen blades attached to a DS4000, special consideration is needed. For example, when using IBM FC switches, the fabric must operate in interoperability mode for the BladeCenter switches to participate in the SAN.

For more details on IBM BladeCenter SAN interoperability, go to:

`http://www-1.ibm.com/servers/eserver/bladecenter/literature/index.html`

Click the Solutions tab and look for the *IBM @server BladeCenter SAN Interoperability Guide* and the *IBM @server BladeCenter SAN Solutions Guide*.

4.1.3 IBM ServerProven

IBM technicians have worked side by side with their VMware counterparts to ensure that the BladeCenter and its most popular hardware and software configurations have been ServerProven with the VMware ESX Server operating system. This is essential to ensure that the IBM VMware program has industry-leading hardware and software solutions to provide the highest level of availability for today's mission-critical environments.

For a list of support IBM blade server hardware refer to the VMware BladeCenter ServerProven Web page:

`http://www.pc.ibm.com/us/compat/nos/vmwaree.html`

4.2 Scaling out VMware ESX Server on BladeCenter

This section describes the advantages and disadvantages of implementing the VMware infrastructure using many small servers with VMware ESX Server on each of them. For more information about scale up versus scale out considerations, see *VMware ESX Server: Scale Up or Scale Out?*, REDP-3953.

Although many options exist for implementing a farm comprised of small low-end servers, IBM BladeCenter offers the most viable alternative when discussing this requirement. The IBM BladeCenter is a powerful integrated chassis that, among many other infrastructure components, contains the IBM HS20 and IBM HS40 blades. These are, respectively, the 2-way and 4-way Intel-based blades.

The following discussions on using blades for VMware could also apply to other 2-way and 4-way servers, but the use of blades is superior in many ways.

Figure 4-4 shows that a BladeCenter chassis containing 14 blade servers can run 14 instances of VMware ESX Server, each of which may have many VMs. For example, if it is assumed that each VMware ESX Server, or blade, can host eight VMs, it is possible that one BladeCenter chassis can host 112 VMs in 7U of rack space. The number of VMs that can be operational on a blade server depends on the configuration of the blade and the nature of the applications installed within the VMs.

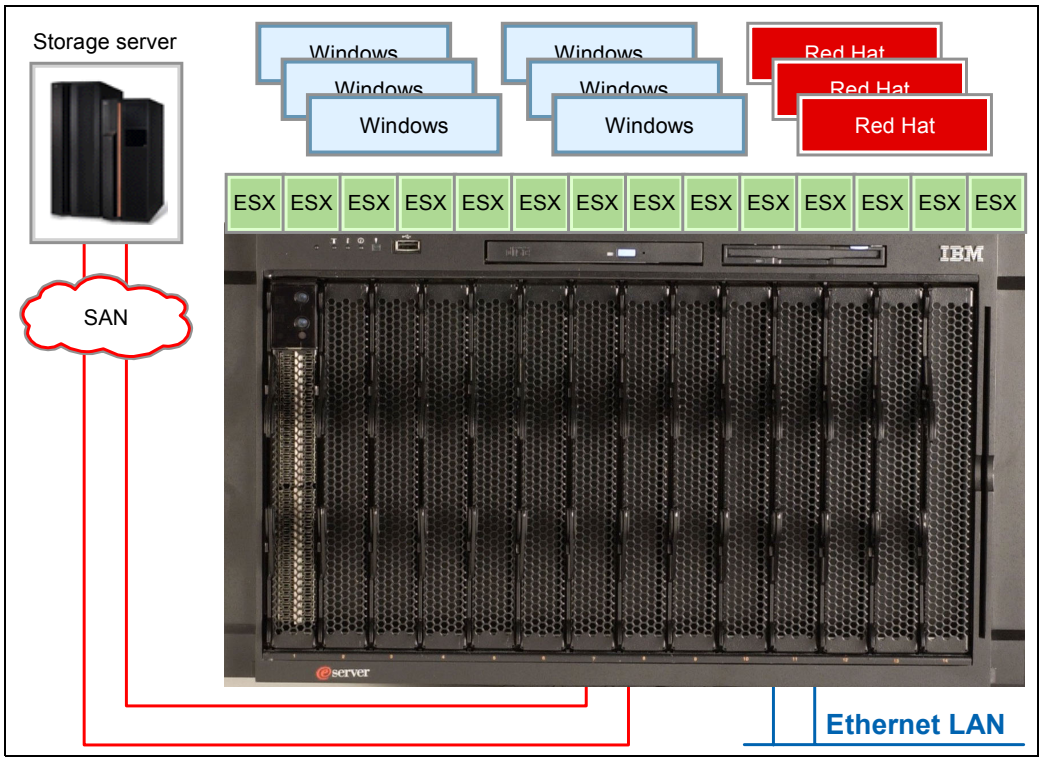


Figure 4-4 Scale out architecture for VMware ESX Server using IBM BladeCenter

Even though the xSeries 445 solution, see Figure 3-6 on page 62, and BladeCenter solution, Figure 4-4, look similar, there is one key difference. While the 16-way x445 is operated by a single VMware ESX Server instance, each blade in the BladeCenter has its own instance, resulting in 14 installations of VMware ESX Server. The software cost is nearly identical to the VMware ESX Server licenses on a per-processor basis.

Scaling out with VMware ESX Server alone creates additional administrative overhead and is not really considered scaling out, but is instead duplication. One of the aims of server consolidation is to minimize the requirement for administration.

To scale out with VMware ESX Server, and minimize administrative overhead, additional software is required. This software is available from VMware and is called VMware VirtualCenter, incorporating VMotion technology. See Chapter 5, “VMware VirtualCenter and VMotion” on page 77 for more information.

Table 4-1 Advantages / Disadvantages of scale out with VMware

Advantages	Disadvantages
► Lower hardware costs for servers ► Small high availability impact in the case of a node failure ► More CPUs supported per rack ► Can add capacity on demand without impacting running systems	► Higher ethernet and SAN costs ► Manual resource management across servers ► Many hypervisor (VMware ESX Server) images to maintain ► VMware VirtualCenter/VMotion mandatory for reasonable management (this is not a big disadvantage as VirtualCenter is a useful component to implement anyway) ► Higher software costs to implement VirtualCenter and VMotion if this component would not have otherwise been implemented ► Higher infrastructure costs for ethernet and SAN switches

As you can see in Table 4-1, this approach has a number of advantages, including the high availability and resiliency of the infrastructure and BladeCenter hardware. Should a single node fail, only a small part of the virtual infrastructure is lost as compared with the significant impact of losing a server in the scale up approach.

For example, assuming that a given solution is capable of hosting four VMs on each processor, the failure of a 2-way blade server would result in the loss of eight VMs. However, should an 8-way server fail, this would result in the loss of 32 VMs.

When VMware ESX Servers are connected to shared storage, it is possible to restart VMs on other servers with available capacity. When using VMotion, it is also possible to move VMs between physical servers, allowing for scheduled downtime of hardware without requiring you to shut down VMs. See Chapter 5, “VMware VirtualCenter and VMotion” on page 77 for more information.

Another advantage is the ability to add resources in smaller incremental amounts. Blades can be added as 2-way or 4-way, without impacting on the existing virtual infrastructure or the running state of VMs. Once the additional server is implemented and included in the virtual infrastructure or farm, stateful VMs can be migrated to the new server, providing that VMotion is licensed.

There are also disadvantages to this approach, one of which is the management of the many instances of the console operating system. Another disadvantage is, although the cost of the server hardware is lower, the complete solution needs to be considered, including:

- Additional cost of ethernet switches (BladeCenter) or ports (non-BladeCenter)
- Additional cost of fibre-channel ports on the switch (non-BladeCenter)
- Additional cost of ethernet and HBA adapters per server
- Cost and management of cabling and IP addresses

It should be noted that the additional cost of ethernet switches in the BladeCenter is offset by the reduction in required network ports. For example, 14 conventional servers with two network ports each would require a total of 28 hub and switch ports in the network. A fully populated BladeCenter with two ethernet switch modules would only require two switch ports in the network to provide the same functionality.

5

VMware VirtualCenter and VMotion

Managing a physical server becomes more important and increasingly complex when you must manage not only the physical hardware and VMware ESX, but also multiple guest operating systems and applications that run in each of them. In this chapter, we discuss the systems management components of VMware ESX Server:

- VMware VirtualCenter
- VMotion

For additional information about the topics in this chapter, refer to the *VirtualCenter User's Manual*. This can be downloaded from VMware's VirtualCenter documentation site:

http://www.vmware.com/support/pubs/vc_pubs.html

5.1 VMware VirtualCenter overview

VMware VirtualCenter is virtual infrastructure management software that centrally manages an enterprise's VMs as a single, logical pool of resources. An administrator can manage thousands of Windows NT, Windows 2000, Windows Server 2003, Linux and NetWare servers from a single point of control. VirtualCenter provides the central point of control for workload management, provisioning and availability. It simplifies IT so companies leverage their storage, network, and computing resources to control costs and respond faster to changing business demands. The VirtualCenter interface provides an overview of all the virtual resources in a data center. From the VirtualCenter interface, administrators can perform:

- Continuous workload consolidation

  Monitor and optimize the utilization of data center resources to minimize unused capacity while maintaining application service levels, by adjusting the resources dedicated to each software service.

- Instant provisioning

  Reduce server-provisioning time from weeks to tens of seconds, allowing administrators to respond immediately to requests for IT services. Using server templates, administrators can ensure that new servers are fully consistent with current build and security policies.

- Zero-downtime maintenance

  Safeguard business continuity 24/7, with no service interruptions for hardware maintenance, deployment, or migration. Use VMotion to move running operating systems and their applications from a system that needs maintenance, then transfer them back when maintenance is complete.

The VirtualCenter server collects and stores persistent data in a dedicated database that contains per-system and environmental information. The VirtualCenter server automatically executes user-specified scheduled tasks, such as powering on, or moving powered-off virtual machines. VirtualCenter with VMotion moves a virtual machine from one host to another, while the virtual machine continues operation. This form of migration with VMotion occurs without service interruption on the virtual machine.

Key management functions built in to VirtualCenter include:

- A single view, or dashboard, of all virtual machines
- Monitoring of system availability and performance
- Automated notifications with e-mail alerts
- Access control integrated with Windows authentication

VirtualCenter runs as a service on Microsoft Windows 2000, Windows XP Professional and Windows 2003. See 5.6, “VMware VirtualCenter requirements” on page 95 for more specifications and requirements.

5.2 VMware VirtualCenter terminology and structure

In this section, we discuss some of the terminology used when describing Virtual Center functions, along with a brief review of its structure. Key to understanding the structure is the concept of VirtualCenter components. The following VirtualCenter components are discussed:

- “Managed components” on page 79
- “Software components” on page 79

5.2.1 Managed components

The following physical and virtual infrastructure components can be monitored and managed by VirtualCenter:

- *Virtual machines* are virtualized x86 personal computer environments in which a guest operating system and associated application software can run. Multiple VMs can operate on the same host machine concurrently.
- *Virtualization platforms* are VMware products, such as VMware ESX Server, GSX Server, or Workstation. These are used to create the VMs in the form of a set of configuration and disk files that together perform all the functions of a physical machine. Through the virtualization platform, you run the VMs, install operating systems and run applications. You can also configure the VMs, which includes identifying a VM's resources, such as storage devices.
- *Hosts* are the physical computers on which the virtualization platform software, such as VMware ESX Server or GSX Server, are installed. It hosts the VirtualCenter VMs.

  Note: When VirtualCenter refers to a host, this means the physical machine on which the virtual machines are running. All VMs within the VirtualCenter environment are physically on ESX Server or GSX Server hosts. The term *host* used in this chapter, means the ESX Server or GSX Server host that has virtual machines on it.

- *Datastores* are the storage locations for the VM files. Datastores hide the idiosyncrasies of various storage options, such as VMFS, SAN, GSX Server and ESX Server, and so on. They provide a uniform model for managing GSX Server and ESX Server hosts and various storage products in the context of VM management.

5.2.2 Software components

VMware VirtualCenter monitors and manages all the virtual machines and hosts registered with VirtualCenter. The software components that make up VirtualCenter perform these functions. VirtualCenter has these components:

- *VirtualCenter client* is a user interface that runs locally on a Windows machine.

  The VirtualCenter client runs on a machine with network access. This can be on the same machine as the VirtualCenter server or on another machine with network access. The VirtualCenter client requires a computer monitor for access to the graphical user interface.
- *VirtualCenter server* is a service that acts as a central administrator for VMware servers connected on a network, to direct actions upon the virtual machines and the virtual machine hosts. VirtualCenter server provides the central working core of VirtualCenter.

  VirtualCenter server runs on a machine as a service. It can also, and typically is, another server machine in the network. It receives direction through the VirtualCenter client and executes actions upon the identified hosts and virtual machines.

  VirtualCenter Web service is a service that can optionally be installed with the VirtualCenter server. It is a required component for third party applications that use the VMware SDK application programmer interface (API).
- *VirtualCenter agent* is installed on each virtual machine host, it collects, communicates, and executes the actions received from the VirtualCenter server.

- *VMotion* is a feature that is activated by the VirtualCenter agent. It enables moving running virtual machines from one VMware ESX Server to another without service interruption. It requires licensing on both the source and target host. The VirtualCenter server centrally coordinates all VMotion activities.
- *VirtualCenter database* is a persistent storage area, for maintaining status of each virtual machine, host and user managed in the VirtualCenter environment. This can be local or remote to the VirtualCenter server machine.

Figure 5-1 illustrates the relationships of the VirtualCenter installed components.

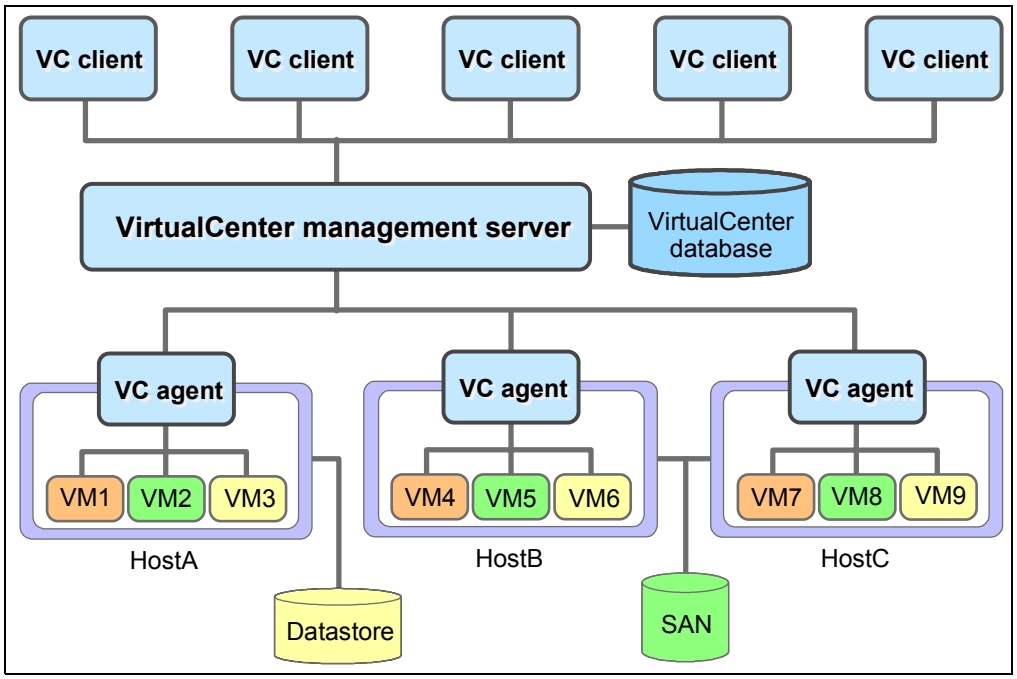


Figure 5-1 VMware VirtualCenter installed components

5.2.3 Organizational components

VirtualCenter provides the ability to define your virtual infrastructure enterprise into a hierarchy of managed organizational components. Organizational components assist in the handling of the potentially hundreds of VMs. They can be renamed to represent their organization purposes. For example, they can be named after company departments, locations or functions. The organizational components are:

- *Server Farms* are the top-level structure for the VirtualCenter server. Only one Server Farms object exists for each VirtualCenter server. Server Farms can contain multiple farm groups and farms. The term Server Farms is the default value, the actual name used can be changed.
- *Farm Groups* are an optional grouping structure. It is hierarchically contained within the Server Farms structure. The VirtualCenter server supports multiple farm groups. Farm groups can contain other farm groups and farms.
- *Farm* is the main structure under which hosts and their associated VMs are added to the VirtualCenter server. VirtualCenter server supports multiple farms.

Notes: A host can be managed by only one farm at a time. All operations between hosts and virtual machines occur within a single farm. For example, hosts and virtual machines are not migrated between farms.

- *Virtual Machine Groups* are an optional grouping structure. They are contained within a farm. VirtualCenter server supports multiple virtual machine groups. Virtual machine groups contain virtual machines and other virtual machine groups.

All the VirtualCenter components, the hosts, the virtual machines, and the organizational groupings of server farms, farms, and farm groups are contained within the VirtualCenter environment.

Figure 5-2 illustrates the hierarchy of the VirtualCenter organizational components.

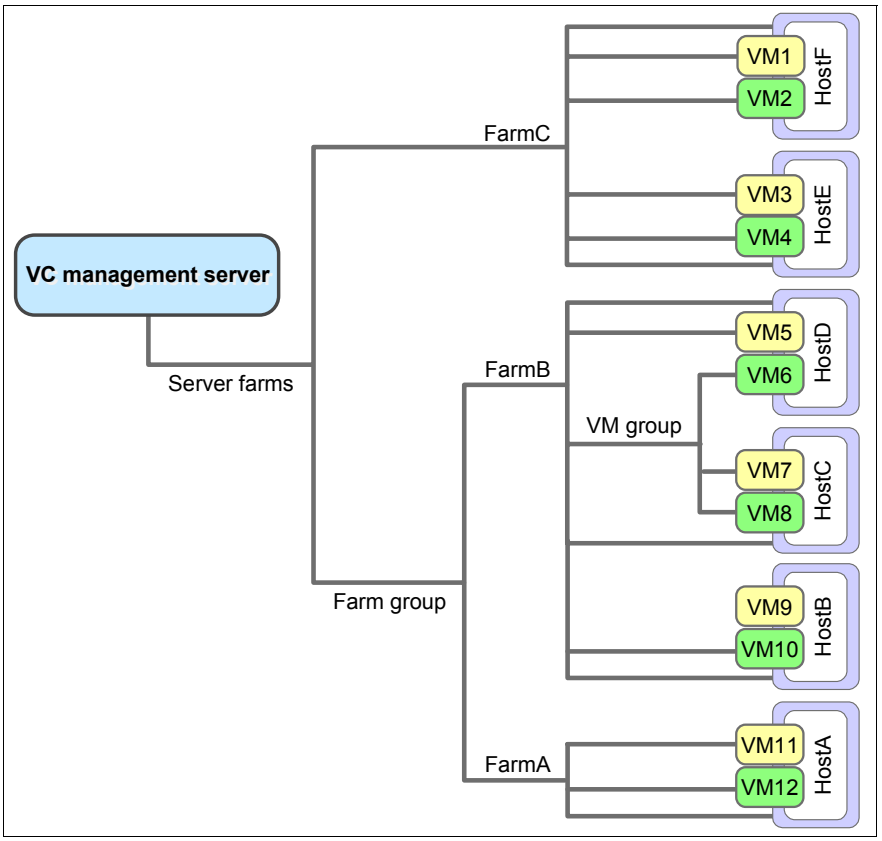


Figure 5-2 VirtualCenter organizational components

5.2.4 Functional components

The functional components group the VirtualCenter monitoring and managing tasks. The functional components are the following:

- *Inventory* provides a view of all the monitored objects in VirtualCenter. Monitored objects include Server Farms, farms, farm groups, hosts, VMs and VM groups.
- *Scheduled Tasks* is a list of activities and a means to schedule those activities.
- *Templates* provides a way to import virtual machines and store them as templates for deploying at a later time to create new virtual machines.

- *Alarms* provides a way to create and modify a set of alarms that you define. Alarms are applied to an object and contain a triggering event and a notification method. Alarms do not have a navigation toolbar option. Alarms are viewed through the Alarms tab for each object.
- *Events* provides a list of all the events that occur in the VirtualCenter environment. Use the Navigation option to display all the events. Use an object specific panel to display only events relative to that object.

Figure 5-3 illustrates the relationship of the VirtualCenter functional components.

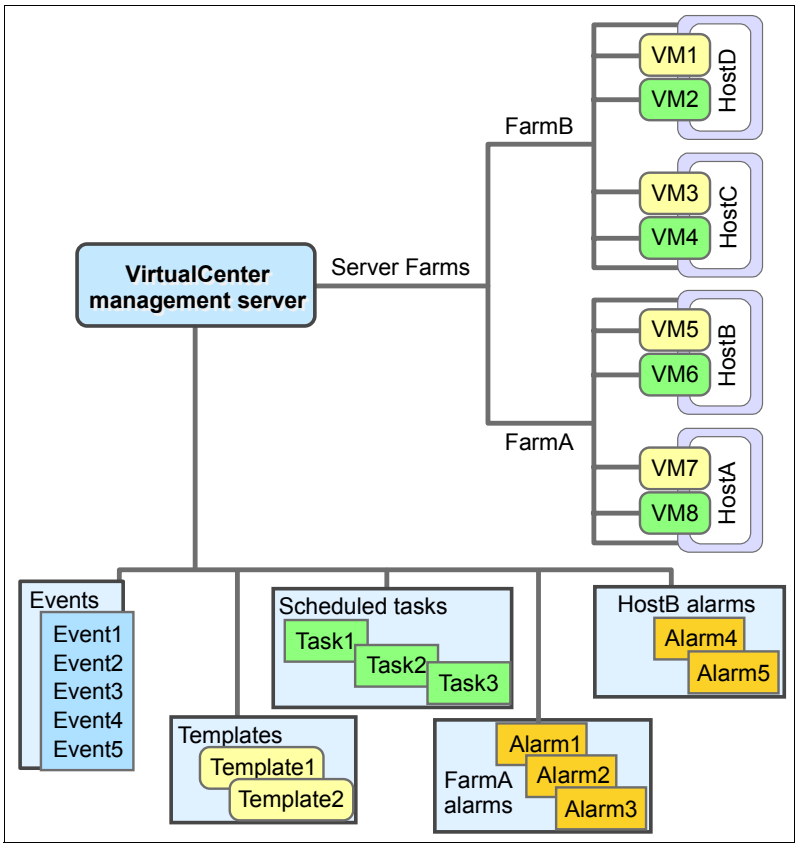


Figure 5-3 VirtualCenter functional components

5.2.5 Access privileges components

Each VirtualCenter user logs on to the VirtualCenter service through the VirtualCenter client. Each VirtualCenter user is identified as someone who has assigned rights and privileges to selected objects, such as farms and VMs, within the VirtualCenter environment. The VirtualCenter itself has full rights and privileges on all hosts and VMs within the VirtualCenter environment. VirtualCenter passes on only those actions and requests to a VirtualCenter user that the user has permissions to perform.

VirtualCenter controls access to each VirtualCenter object, farm, farm group, VM and VM group. To do this, VirtualCenter assigns a role and a user, or group, to each object.

Individual permissions are assigned through VirtualCenter by pairing a user and a role and assigning this pair to a VirtualCenter object.

- Users and Groups are created through the Windows domain or Active Directory database. VirtualCenter registers users and groups as part of the assigning privileges process.

- Roles are a VirtualCenter predefined set of access rights and privileges. There are four roles. Each subsequent role includes the privileges of the previous role. The types of roles that can be paired with a user and assigned to an object are:
  - Read Only User

    Users assigned this role for an object are allowed to view the state of virtual machines, hosts, farms, and groups. With this role, the user can view virtual machines, hosts, farms, farm groups, and virtual machine group attributes. That is, all the tab panels in VirtualCenter except the Console tab. You cannot view the remote console for a host. All actions through the menus and toolbars are disallowed. A Read Only User role can view the templates and scheduled tasks but not perform any actions with them.
  - Virtual Machine User

    Users assigned this role for an object are allowed to perform power operations on virtual machines. With this role, you can connect to a remote console and view the states of virtual machines. You cannot modify the configuration of hosts or virtual machines.
  - Virtual Machine Administrator

    Users assigned this role for an object are allowed to add, remove, or modify objects. With this role you:
    - Connect and disconnect host devices, migrate and migrate with VMotion, clone, remove and configure virtual machines
    - Create, import, and deploy templates
    - Add and remove hosts from farms
    - Create, remove, or modify farms, farm groups, and virtual machine groups and their content
  - VirtualCenter Administrator

    Users in this role are allowed to change privileges for an object. With this role, you can add, remove and set access rights and privileges for all the VirtualCenter as well as all the virtual objects in the VirtualCenter environment.

Figure 5-4 on page 84 illustrates the three users and their respective access to VirtualCenter objects.

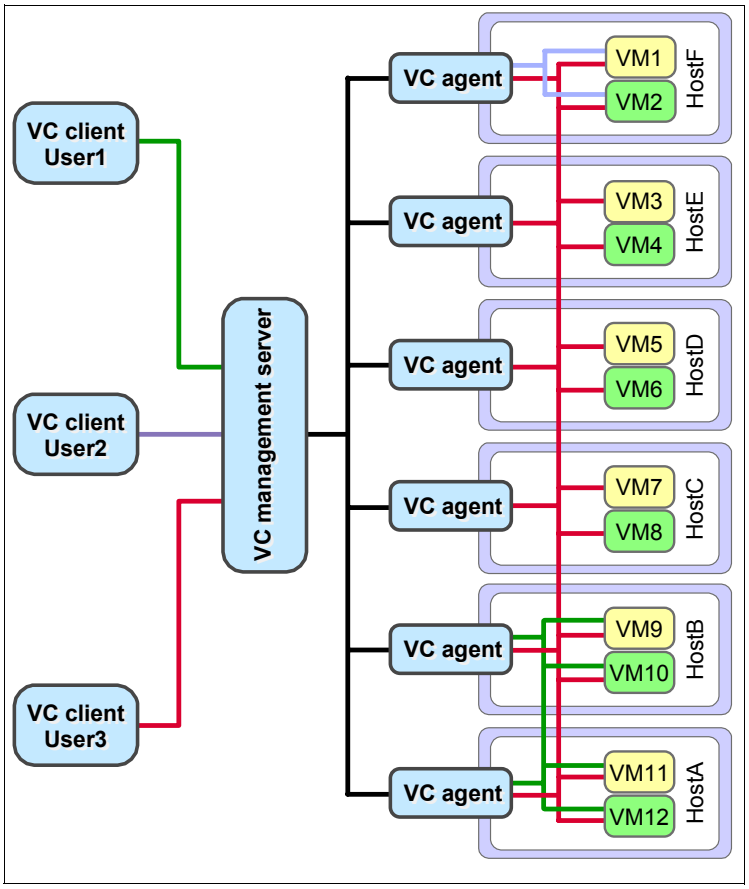


Figure 5-4 VirtualCenter access control

5.3 VMware VirtualCenter operations overview

Typical VirtualCenter operations are focused around managing virtual machines on multiple hosts.

Figure 5-5 on page 85 illustrates a VirtualCenter operation overview.

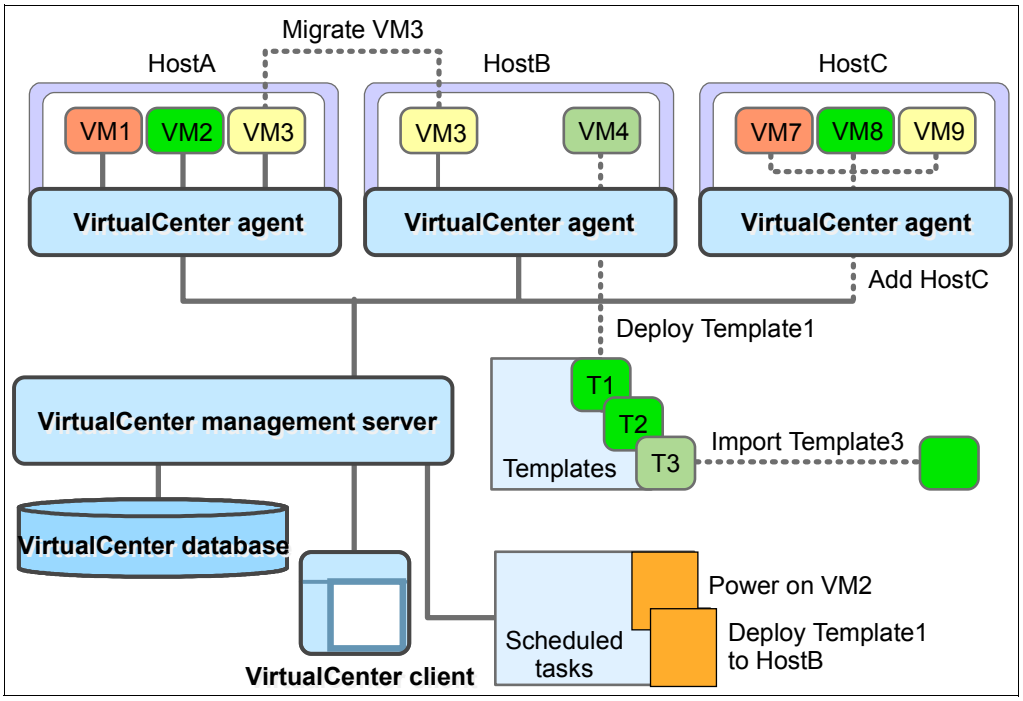


Figure 5-5 Working VirtualCenter environment

Note: For the remainder of this section, the term *VirtualCenter* refers to the VirtualCenter client, unless explicitly stated otherwise.

There are several options for adding VMs to VirtualCenter:

- Register a host with VMs.
- Create VMs.

The following list describes some of the key process options.

- A VM can be created from scratch using the New Virtual Machine wizard. After creating a VM, you must install a guest operating system.
- An existing VM can be used as source for new templates using the New Template wizard.
- An existing VM can be cloned to create new VMs using the Clone wizard.
- Templates can be deployed as new VMs using the Template Deployment wizard.
- The new VM created through the Clone or Template wizards can be customized using the Guest Customization wizard. This wizard customizes a VM that already has a guest operating system installed.
- After creating a new VM, you can edit it with the Virtual Machine Properties dialog box.

5.3.1 VirtualCenter virtual machines and hosts

Hosts are added to the VirtualCenter environment through an Add Host wizard. When VirtualCenter adds a host, it automatically discovers and registers all the virtual machines on that host. Each host is associated with a specific farm. Select a farm object to enable the Add Host wizard menu option. Figure 5-6 on page 86 illustrates the process of adding virtual machines to VirtualCenter control.

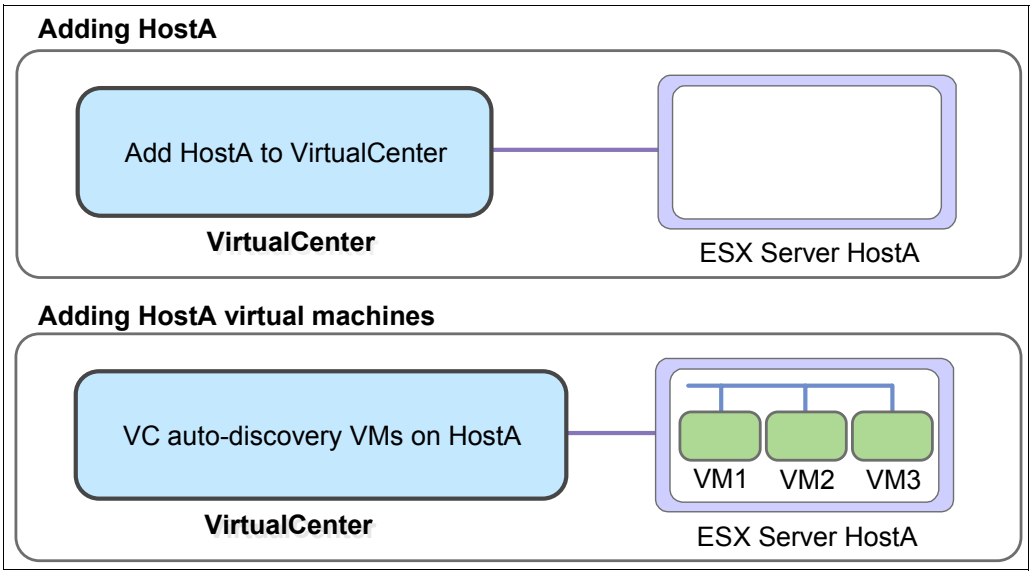


Figure 5-6 Adding a host to a farm and performing automatic discovery of VMs

5.3.2 Virtual machine templates

A template is a golden image of a VM. The template typically includes a specific operating system and configured components which provide virtual counterparts to hardware components. VirtualCenter uses templates to create new VMs.

Creating a template

You create templates from existing virtual machines. These virtual machines can be either:

- VMs located on any VirtualCenter registered host
- VMs stored on a disk local to the VirtualCenter server

  Virtual machines local to the VirtualCenter server can be VMs created by supported versions of VMware ESX Server, GSX Server, or Workstation. Refer to the *VirtualCenter User's Manual* for additional information.

In addition, you can create templates by copying, or cloning, an existing template.

When you create a template, VirtualCenter provides an option to store the template at one of three locations:

- Original datastore

  You can store the template in the original datastore location of the source VM. The original VM is the template. This option works best with a shared datastore between hosts on a farm. This option is typically used for creating templates from VMs on registered hosts.

- Template upload directory

  You can store the template in the template upload directory, located on the VirtualCenter server machine. It contains copies of the original VM virtual disks. Specify a directory local to the VirtualCenter server as the template upload directory. You use this typically when a template needs to be deployed to any host.

- Local VirtualCenter server

  You can use the local VirtualCenter server. This option is available if source VM and a VirtualCenter client are local to the VirtualCenter server.

Figure 5-7 illustrates the two methods for storing templates in VirtualCenter.

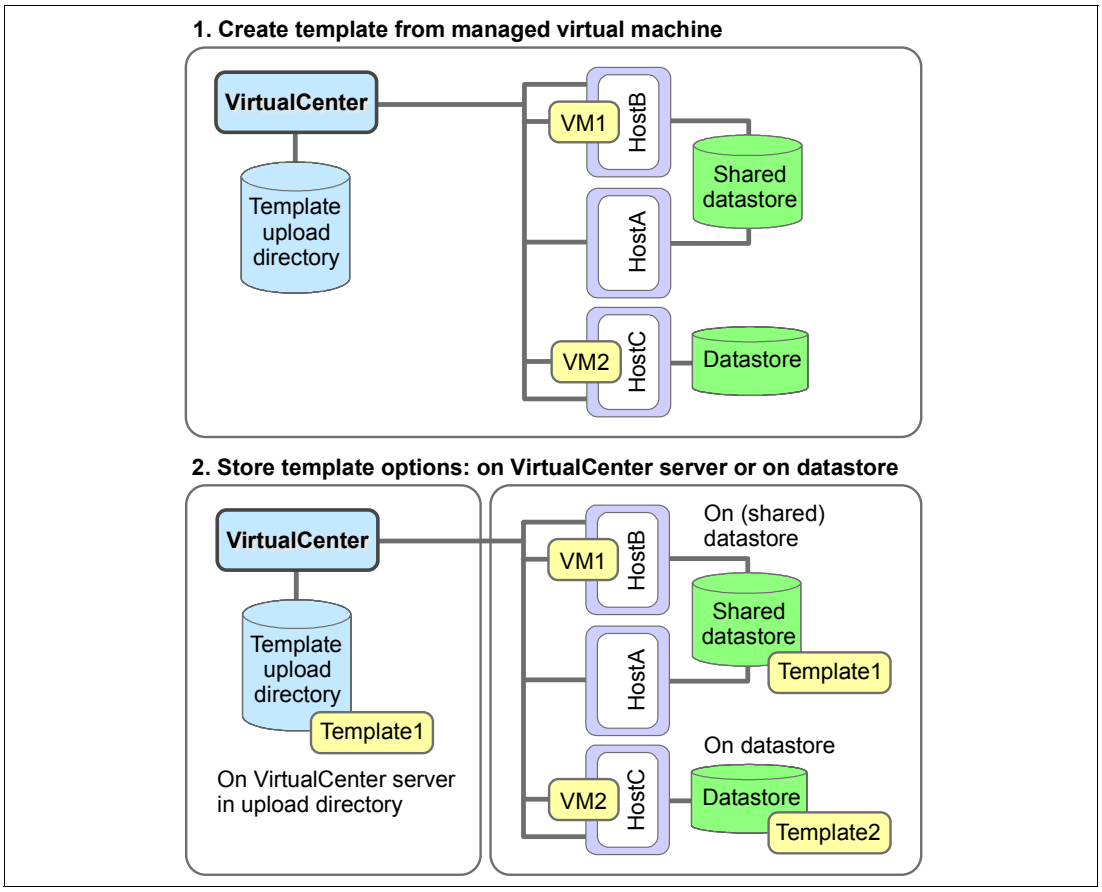


Figure 5-7 Creating a template

Deploying a template

When you deploy a template, you create a new VM from the template. If the template is in the template upload directory, place the new VM on any VirtualCenter managed host.

If the template is stored on the datastore of a host, place the new VM only on another host that has access to the datastore. Only those hosts that share the datastore are listed as Deploy a Template wizard target host choices.

Figure 5-8 on page 88 illustrates the process for deploying a template to create a new virtual machine.

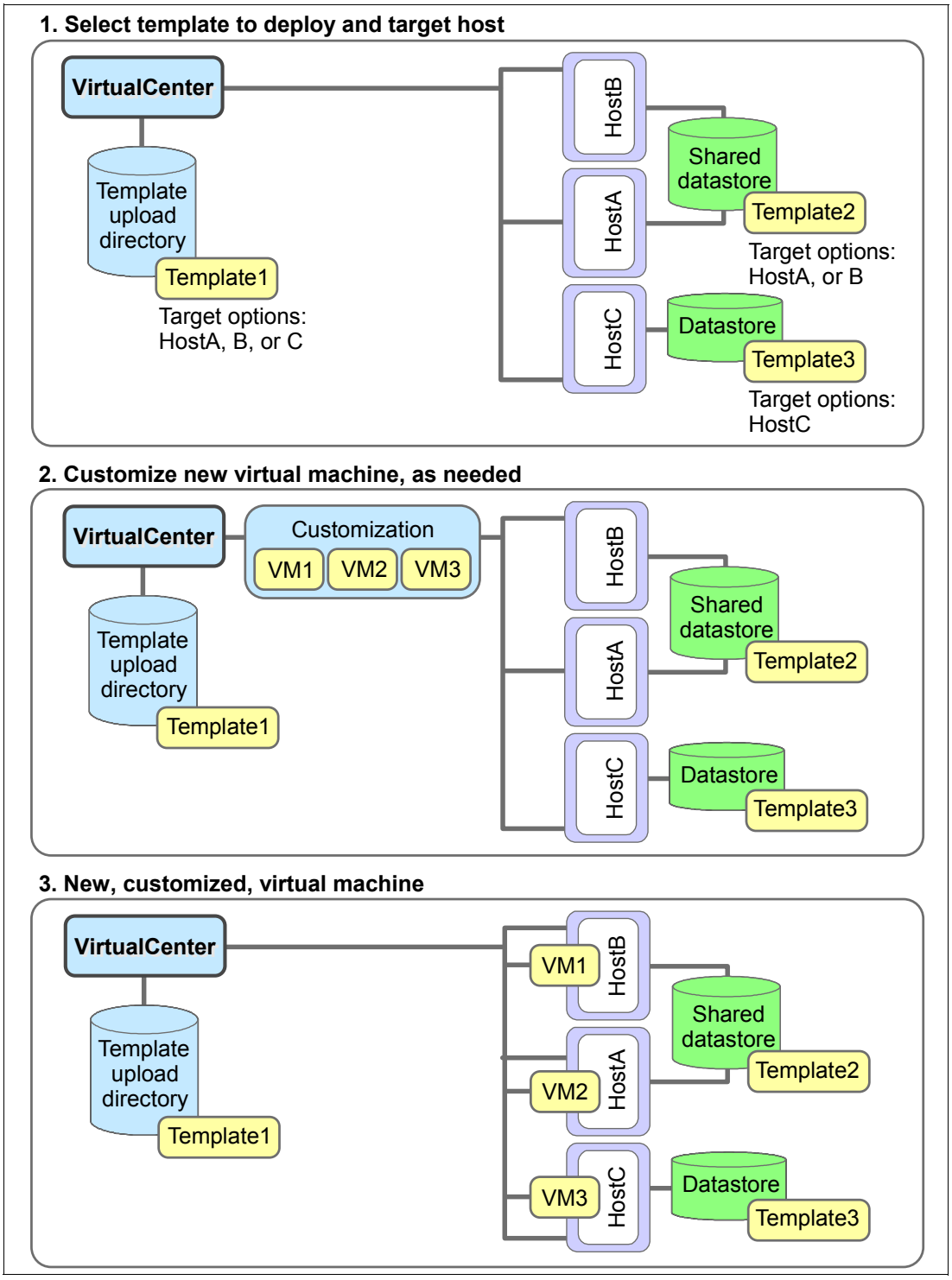


Figure 5-8 Deploying a template

5.3.3 Cloning virtual machines

A clone is a copy and a customization of a VM. When you create a clone, VirtualCenter provides an option to customize the guest operating system of that virtual machine. Store clones on any host within the same farm as the original virtual machine.

Figure 5-9 illustrates the process for cloning an existing virtual machine to create a new virtual machine.

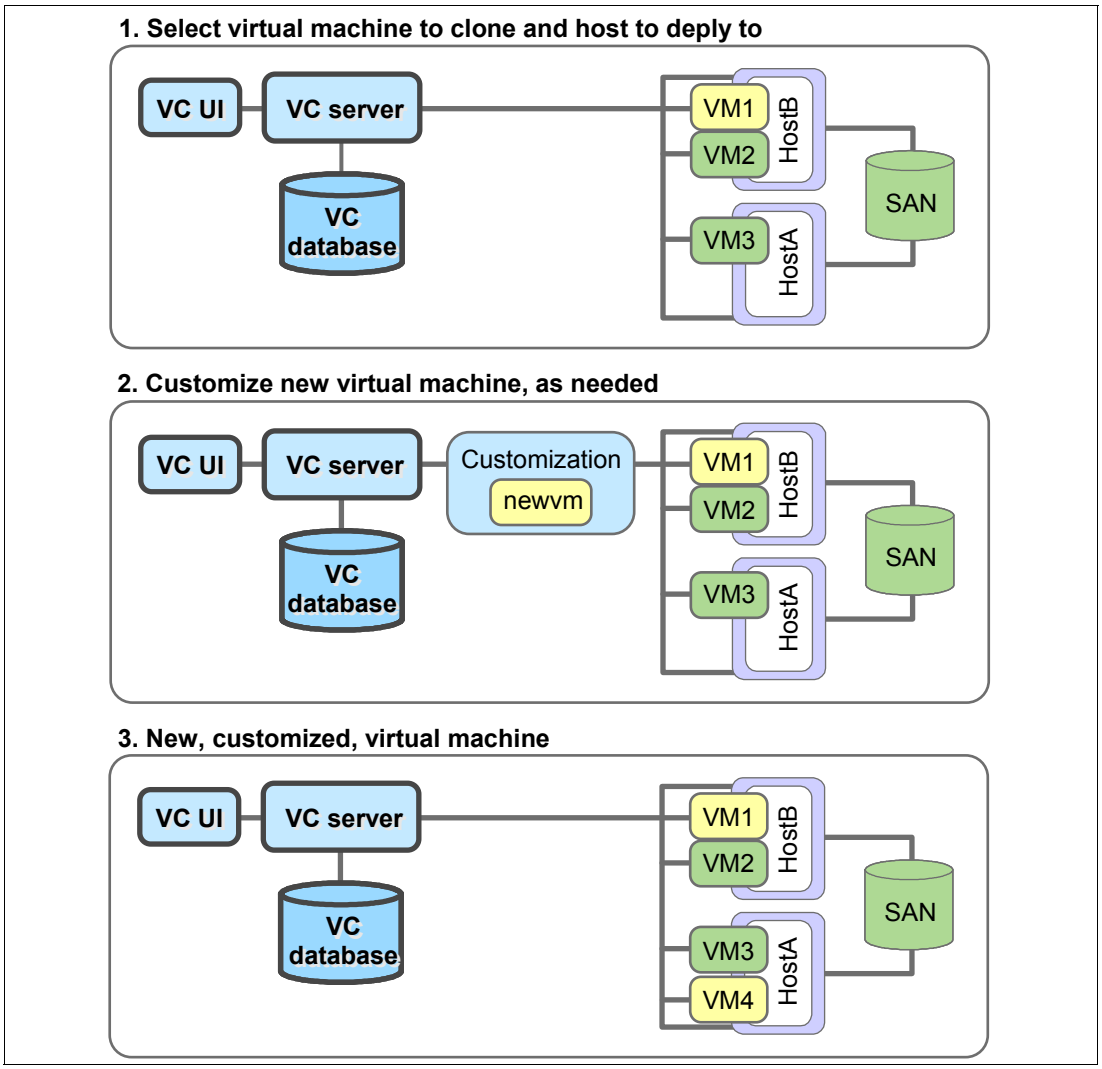


Figure 5-9 Creating a clone

5.3.4 Customizing virtual machines

When creating a virtual machine by either deploying a template or cloning an existing virtual machine, an option is provided to customized the guest operating system installed on that template or existing virtual machine.

Customizing Windows guest operating system options include:

- Join workgroups and domains
- Network interface configuration
- Domain suffix, security ID (SID) change

Customizing Linux guest operating system options include:

- Host name
- Domain suffix
- Network interface configuration

5.3.5 Migrating virtual machines

Migration is the act of moving a virtual machine from one host to another. There are two types of migration:

- Migrate an active virtual machine using VMotion
- Migrate a powered off virtual machine

Virtual machines can be moved between hosts within the same farm, but VMs cannot be moved between farms.

Migration with VMotion

VMotion allows you to move a powered on virtual machine between hosts in a farm. Moving a powered on virtual machine allows the virtual machine to continue performing transactions without interruption. This functionality requires activation of VMotion on both the source and target host. Figure 5-10 illustrates VMotion functionality.

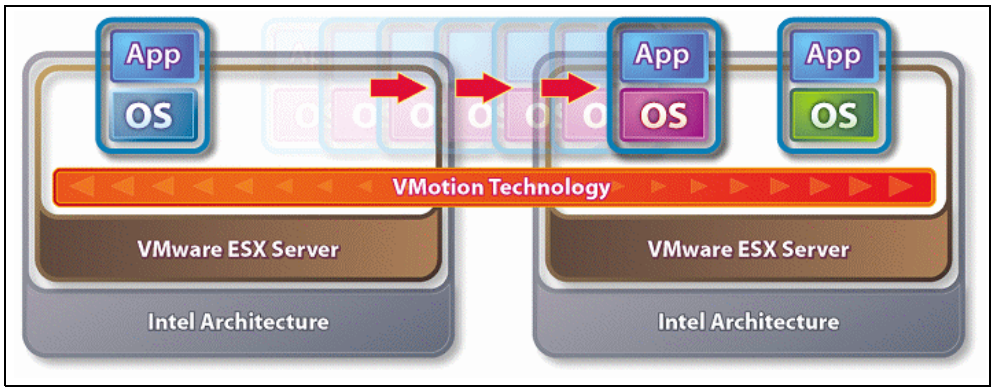


Figure 5-10 Moving an active VM to a different VMware ESX Server hosts

During a migration with VMotion, you move only the virtual machine configuration file from host to host. The associated virtual disk remains in the same location on the SAN storage that is shared between the two hosts. Once the configuration file is migrated to the alternate host, the virtual machine is then managed by the new host.

When you move the configuration file, you also move the entire state of the virtual machine to the new host, even while the data storage remains in the same location on the SAN.

The state information includes the current memory content and all the information that defines and identifies the virtual machine. The memory content includes transaction data and whatever bits of the operating system and applications are in the memory. The defining and identification information stored in the state includes all the data that maps to the virtual machine hardware elements, such as BIOS, devices, CPU, MAC addresses for the Ethernet cards, chip set states, registers, and so forth.

Migration with VMotion happens in four stages:

1. When the migration with VMotion is requested, VirtualCenter verifies that the existing virtual machine is in a stable state with its current host.
2. The migration begins when VirtualCenter makes a baseline copy of the current memory content and stores it on the target host.
3. When the baseline copy is complete, VirtualCenter copies the last changes in the current memory content and copies the virtual machine state information to the target host.
4. The virtual machine resumes its activities on the new host.

Figure 5-11 illustrates migrating a powered on virtual machine from one host to another using VMotion.

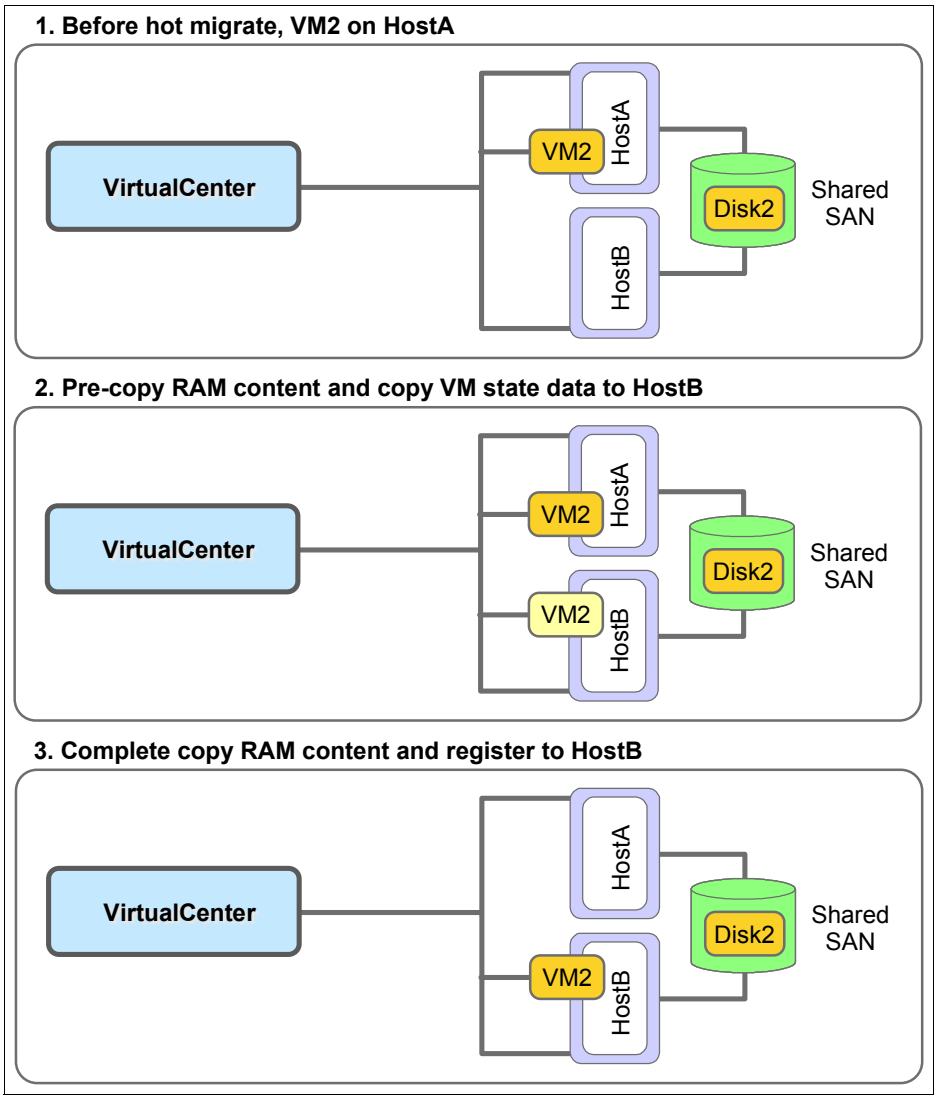


Figure 5-11 Performing a migration with VMotion

Migration

Migration is the process of moving a VM from one host to another. With a migration, you also have the option of moving the associated disks from one host to another. Migration requires that the VM being migrated is turned off prior to beginning the migration process. A migration consists of the following steps:

1. A copy of the configuration file and the associated virtual disks are placed on the new host and in the associated volume storage area.
2. The virtual machine is managed on the target host and unregistered on the source host.
3. After the migration is completed, the old version of the virtual machine is deleted from the source host.

If any error occurs during migration, the VMs revert to their original states and locations.

Figure 5-12 illustrates the process for migrating a powered off virtual machine from one host to another.

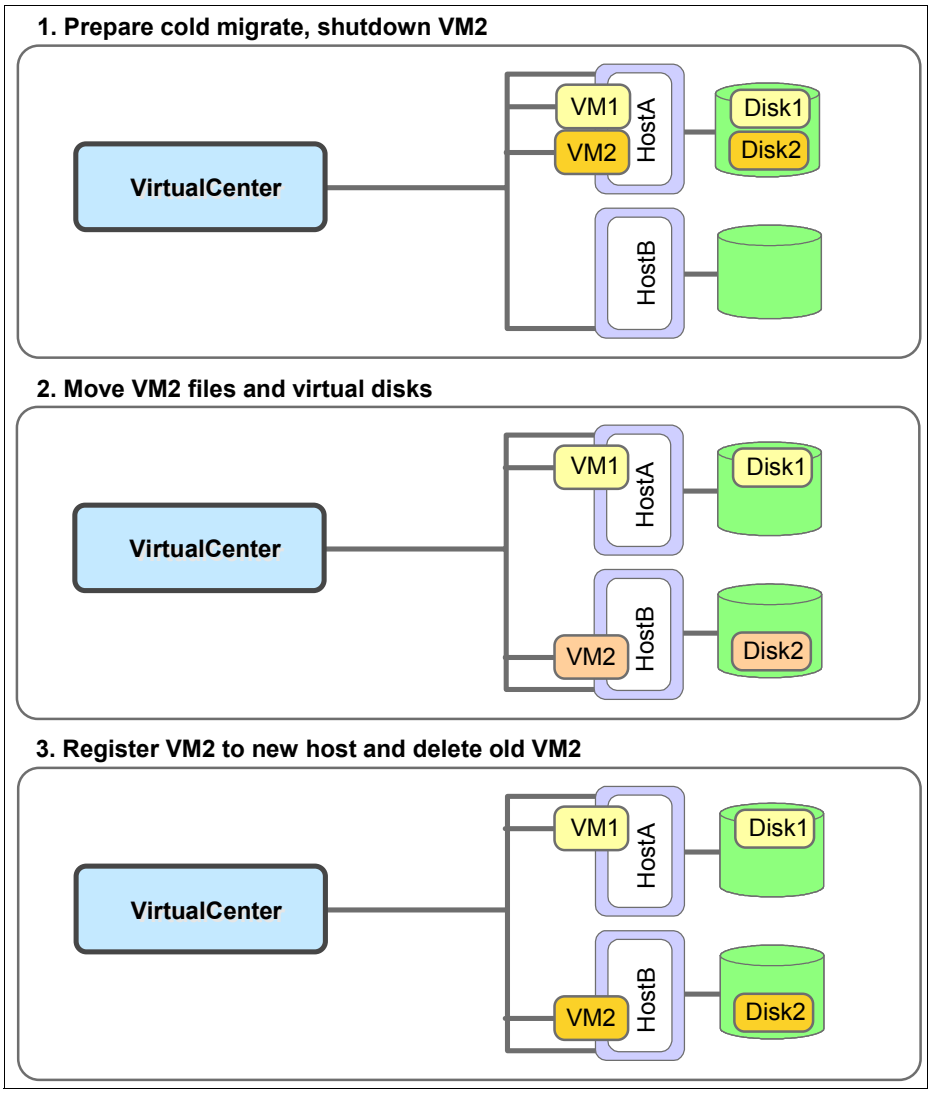


Figure 5-12 Performing migration of a powered off virtual machine

5.3.6 Creating and configuring new virtual machines

In addition to providing multiple methods for moving, migrating, deploying, and cloning virtual machines, VirtualCenter also incorporates the creation capabilities of the virtualization platforms. Through the New Virtual Machine wizard and the Virtual Machine Properties dialog box, you may create new virtual machines from scratch and make configuration modifications to them once they are created.

5.3.7 VirtualCenter scheduled tasks

VirtualCenter allows you to create scheduled tasks to automatically perform operations at specified times. You may configure the tasks to occur once or routinely. Apply tasks to groups, farms, or individual virtual machines.

The VirtualCenter New Task wizard provides the following task options:

- Deploy a virtual machine from a template
- Clone an existing virtual machine
- Change the power state of a virtual machine
- Migrate a virtual machine
- Migrate a virtual machine with VMotion
- Change resource settings of a virtual machine

5.3.8 VirtualCenter events

VirtualCenter events are logged in the VirtualCenter database. Each event panel displays the events relevant to the selected VirtualCenter object. With the Events option in the navigation bar selected, the Events panel displays all the events for the current session. With a farm or host selected, only the events that occurred on that farm or host appear in the Events panel.

5.3.9 VirtualCenter alarms

Alarms are preconfigured actions that can be triggered when selected events occur. Apply alarms to any VirtualCenter object. View, modify, and remove an existing alarm through the Alarms tab of the object where the alarm was defined.

5.4 Scaling out with VMotion and BladeCenter

The benefits of VMotion in a blade server environment are:

- Improving overall utilization across the data center
  - Physical server boundaries are no longer relevant. Treat all server resources as a single pool of computing power.
  - Move applications flexibly across the computing pool to maximize application throughput and achieve high utilization levels.
- Reacting more quickly to changing business demands
  - Repurpose hardware resources on the fly without requiring extensive change management procedures.
  - Perform zero-downtime hardware upgrades that are transparent to users.
- Creating alternate form-factors for cost-effective consolidation
  - Build scale-up solutions using low-cost server blades.

Figure 5-13 on page 94 illustrates VMotion scale out with BladeCenter.

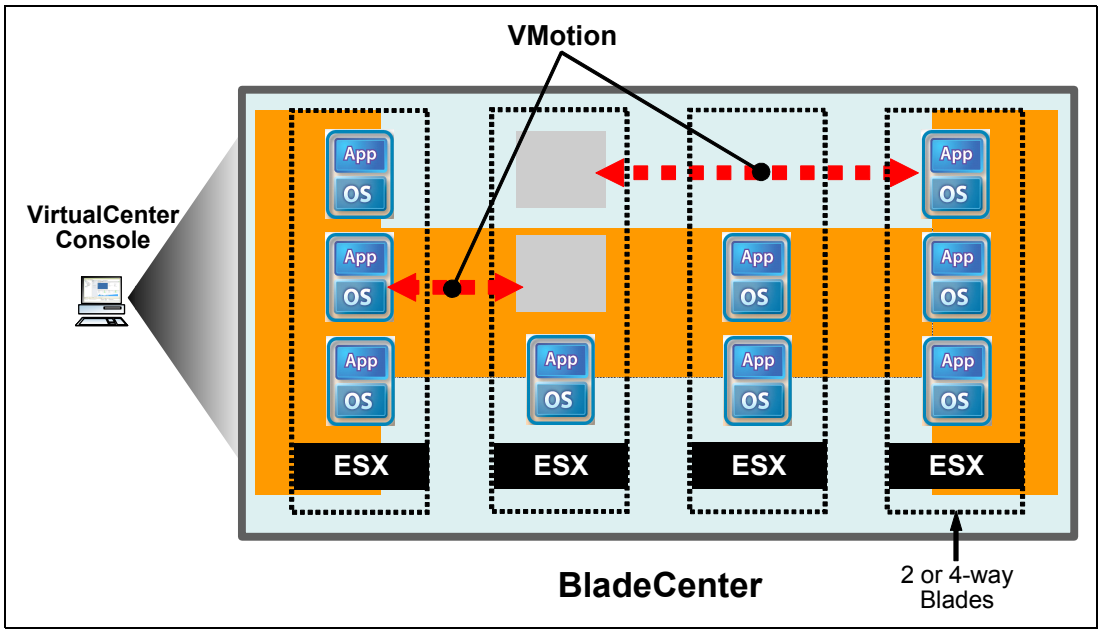


Figure 5-13 Managing VMs in a blade server environment

Enabling a virtual scale-out solution using multiple server blades has the following benefits:

- Achieve 60-80% average utilization
- Share headroom across multiple server blades to handle spikes
- Applications are mobile across server blades
- Transparently shift systems across blades to match complementary workloads

The IBM BladeCenter is covered in Chapter 4, “IBM Eserver BladeCenter with VMware ESX Server” on page 65.

5.5 Why Use VMware VirtualCenter?

There are several advantages and benefits to using VirtualCenter:

- Dramatically reduce total cost of ownership (TCO)
  - Reduce the cost of hardware by achieving even higher utilization levels
  - Reduce substantially the cost of managing a heterogeneous environment
  - Enhance the value proposition of server blades
  - Improve availability of the existing computing infrastructure
  - Eliminate planned hardware downtime
  - Improve proactively application performance by balancing workloads with VMotion
- Achieve unparalleled levels of business flexibility
  - Instantly repurpose hardware resources across the datacenter
  - Reduce the number of operations needing lengthy change management procedures
- Take a major step towards utility computing
  - Apply virtual computing to legacy and next-generation applications alike

These advantages and benefits are discussed more fully detailed in the Table 5-1 on page 95.

Table 5-1 VMware VirtualCenter usage scenarios

Usage scenarios	Benefits
Dynamically move workloads across distributed physical servers ► Migrate virtual machines across an environment to optimize workloads ► Perform zero downtime and rolling hardware upgrades ► Migrate virtual machines to new physical servers in response to a hardware failure	► Reduce hardware costs by increasing utilization ► Cut operational costs by reducing downtime windows ► Ensure application availability and end-user continuity in the event of a hardware failure
Streamline server provisioning and management ► Centrally manage all virtual machine images ► Deploy new virtual machines using standardized templates ► Enforce configuration standards through pre-built virtual machines	► Reduce operational costs ► Instantly provision new virtual machines ► Keep up with the demand for new servers while controlling costs
Monitor system availability and performance ► Monitor individual virtual machine usage of system components ► Set triggers and alerts for key performance and availability metrics ► Leverage pre-built reports to proactively identify resource contention issues	► Increase availability and lower hardware costs ► Set policies to automatically manager performance and availability ► Reduce operational costs ► Identify resource issues or hardware failures earlier
Manage distributed servers as a single pool of resources ► View entire system and virtual machine inventory from a central console ► Manager access control globally across servers and virtual machines ► Remove hardware-related dependencies from OS management	► Manage software independently from hardware ► Manager application workloads efficiently and seamlessly ► Provide continuous hardware availability to all applications

5.6 VMware VirtualCenter requirements

This section discusses several prerequisites for VMware VirtualCenter to operate.

5.6.1 Server requirements

The VirtualCenter server must have:

- Administrator privileges on the installing system to install the VirtualCenter server

  The VirtualCenter server installation adds VirtualCenter as a Windows service.
- Microsoft Windows 2000 Server, Windows 2000 Advanced Server, Windows XP Professional, or Windows Server 2003 (Web, Standard, and Enterprise)
- Minimum 2GB RAM for VirtualCenter configurations managing 50 hosts or less
  - For configurations greater than 50 hosts, use 3GB RAM.
  - For configurations with 100 hosts running 2000 virtual machines, use 4GB RAM.
- Minimum Pentium IV 2.0Ghz processor

  Dual processors are recommended for deployments with greater than 25 hosts.
- Minimum 1 10/100Mbps NIC (1Gbps NIC recommended)

- Windows Script version 5.6 or later

  If you do not have this version, VirtualCenter installer automatically updates to Windows Script version 5.6.

- Disk space sufficient on the machine to support the VirtualCenter database and the template upload directory

  The VirtualCenter server can run on the same machine as the VirtualCenter client, or it can be installed separately on another Windows system. The VirtualCenter server can also be installed in a VM.

5.6.2 Client requirements

The VirtualCenter client must have:

- .NET Framework version 1.1

  If you do not have this version, VirtualCenter automatically updates to .NET Framework version 1.1.4322.573.

- Windows 2000 (all versions), Windows XP Professional, Windows XP Home Edition, Windows Server 2003 (all versions), Windows 98, Windows 98 SE, Windows ME, or Windows NT4 (SP6a required)
- Minimum 256MB RAM (512 MB recommended)

The VirtualCenter client can be installed on multiple Windows systems and access the VirtualCenter server through the network. These Windows systems can be on your desktop, mobile computer, or another VM.

Note: If you install your VirtualCenter server in a virtual machine, do not install the VirtualCenter client in that same virtual machine. Viewing a Console recursively from that virtual machine might cause the host to fail.

5.6.3 Supported databases

The VirtualCenter database must have one of the following compatible database formats:

- Microsoft Access (default)

  You do not need to have a Microsoft Access database installed to use this option. VirtualCenter stores the database in Microsoft Access format.

- Microsoft SQL Server 2000
- Microsoft SQL Server 7
- Oracle 8i, version 8.1.5 and Oracle 9i, version 9.0.1

Note: VMware recommends using the Microsoft Access Database for demonstration environments and proof of concepts only. This database is not recommended for production deployments. For production environments use either the SQL Server or Oracle database.

5.6.4 Managed hosts requirements

Currently, the VirtualCenter registered and managed hosts must be running the following virtualization platform:

- VMware ESX Server 2.0.1, 2.1
- VMware GSX Server 3.1

5.6.5 Networking requirements

VirtualCenter host and VM networking capabilities are as follows:

- Up to four virtual Ethernet network adapter cards (NICs)

  Each virtual NIC can be a high-performance VMware virtual NIC or AMD PCnet-PCI II compatible virtual NIC.

- Any protocol that the guest operating system supports over Ethernet

  You can even use multiple high-performance Ethernet-compatible virtual networks.

- A TCP/IP address configured on an extra NIC sufficient to support migration
- Minimum of two NICs

  How you assign the NICs differs depending upon the version of the VMware virtualization platform you use:

  – For VMware ESX Server V2.1, create a bond including both NICs and give the bond to all three entities: service console, VMkernel, and VMs. Use VLANs to create separate logical networks.

  – For VMware ESX Server V2.0.1, dedicate one NIC to the service console and VMotion. Dedicate the other NIC to the VMs. The preferred minimum number is three:

    - One is dedicated to the host.
    - One (or more) are dedicated to the VMs.
    - One is dedicated to VMotion activity.

5.6.6 VMotion requirements

To perform a migration with VMotion, you must have a VMotion license for each host and VMotion must be active on each host. The hosts you are planning to migrate between must meet the following requirements:

- Hosts must share a storage area network (SAN) infrastructure.
- Hosts must be configured to use a shared Virtual Machine File System (VMFS) volume.
- Clients are advised to deploy a Gigabit Ethernet network for VMotion.
- Source and target hosts can be from different hardware vendors, but the CPU vendor and family need to be consistent.
- Source and target hosts need to run ESX Server V2.x or later.

5.7 VMotion usage scenario: Zero-downtime maintenance

VMotion can be used to move running virtual machines from one physical server to another to allow hardware maintenance to be performed. This scenario provides a method for upgrading, replacing or upgrading hardware without impacting on the availability of the services.

Figure 5-14 on page 98, Figure 5-15 on page 98, Figure 5-16 on page 99, and Figure 5-17 on page 99explain how this process works.

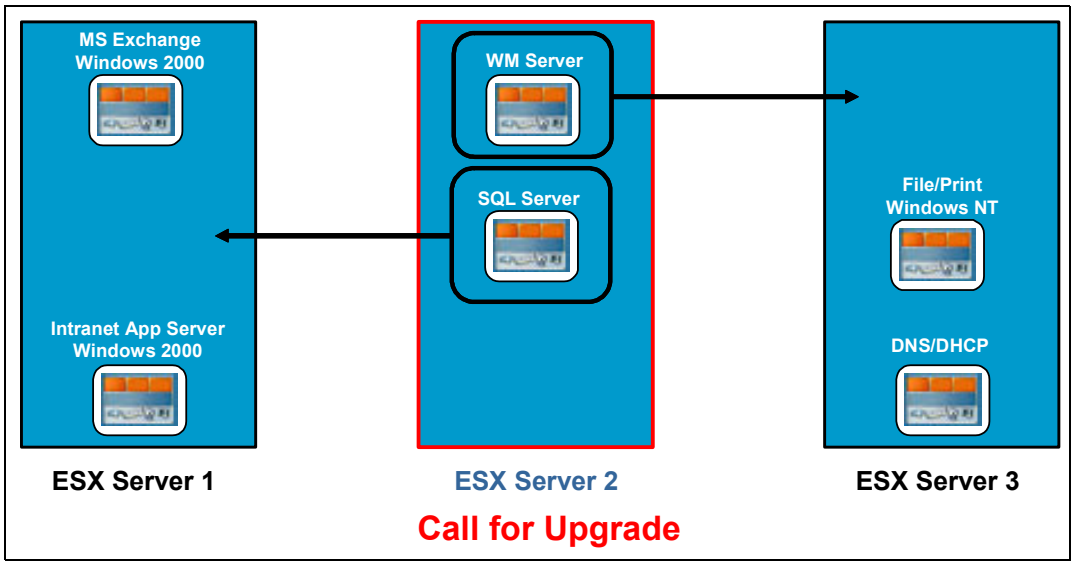


Figure 5-14 HW maintenance scenario

In this somewhat simplified environment in Figure 5-14, ESX Server 2 is to be serviced. Using VMotion, it is a simple matter of moving the active VMs from that server to servers with available capacity. In this case, ESX Servers 1 and 3 have available capacity.

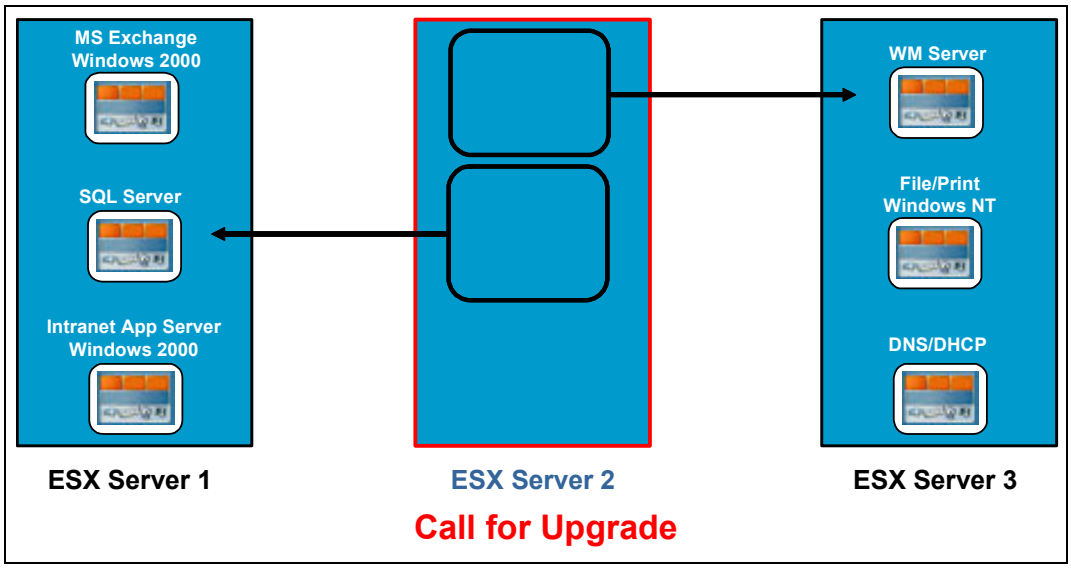


Figure 5-15 Transferring affected VMs

Figure 5-15 shows that these VMs are active and accessible even during the migration. Once migrated, there are no virtual machines running on ESX Server 2.

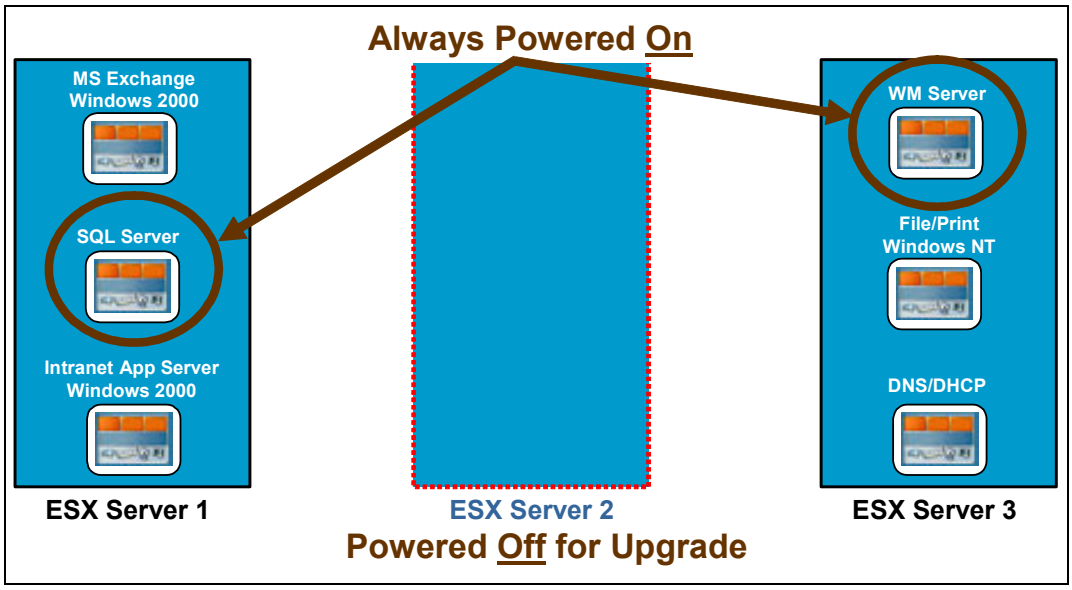


Figure 5-16 VMs in active state during transit

Figure 5-16 shows that ESX Server 2 can be turned off for upgrades or maintenance.

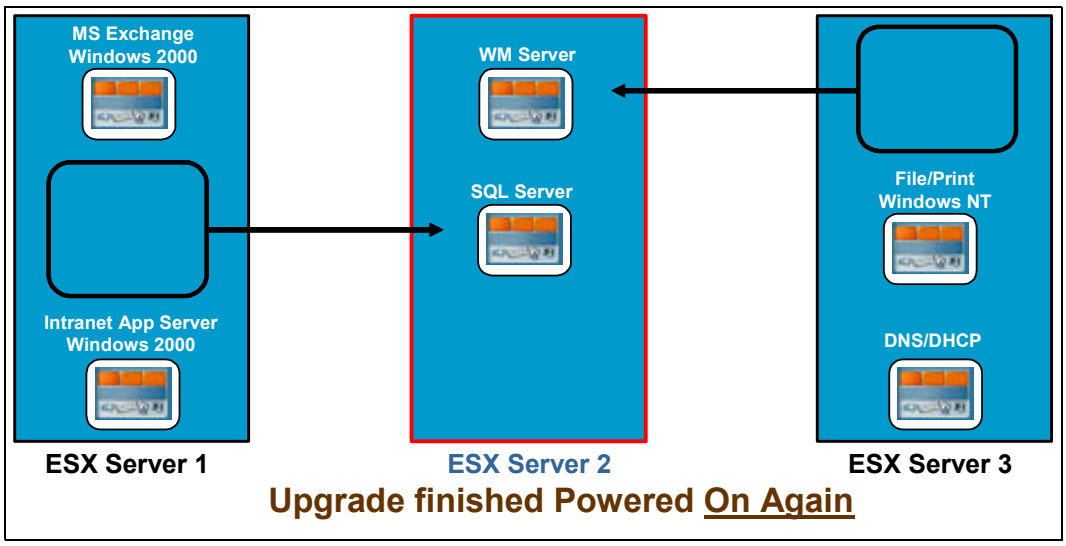


Figure 5-17 VMs moved back to original host

Figure 5-17 shows that once the work is completed on ESX Server 2, the VMs can be migrated back without impact to the users.

While it is possible to perform this task on production servers, it would be prudent to schedule this activity for off-peak times to minimize any risk. The benefit of this scenario is that it is possible to perform hardware maintenance at times which may be more convenient for support staff and contractors. For example, hardware maintenance could be performed on a weekday at 6pm, rather than on a weekend or very late at night.

6

VMware management with IBM Director and VMM

IBM Director Version 4.2 is a workgroup management application for Intel processor-compatible systems. IBM Director consists of a suite of systems management tools that help users manage hardware to achieve maximum system availability and lower IT costs.

In this chapter, we provide a brief overview of IBM Director, the Virtual Machine Manger (VMM) extension, and how they fit into a VMware ESX Server management environment.

6.1 IBM Director overview

IBM Director is a comprehensive systems-management solution. Based on industry standards, it can be used with most Intel microprocessor-based systems and certain IBM @server iSeries™ and IBM @server pSeries® servers.

A powerful suite of tools and utilities, IBM Director automates many of the processes that are required to manage systems proactively, including capacity planning, asset tracking, preventive maintenance, diagnostic monitoring, troubleshooting, and more. It has a graphical user interface that provides easy access to both local and remote systems.

IBM Director can be used in environments with multiple operating systems, or *heterogeneous environments*, and integrated with robust workgroup and enterprise management software from IBM such as Tivoli® software, Computer Associates, Hewlett-Packard, Microsoft, NetIQ, and BMC software.

IBM Director consists of a suite of systems management tools that deliver hardware manageability. IBM Director's industry-standard foundation enables heterogeneous hardware support and works with a variety of operating systems and network protocols.

As the sole systems management application in an environment, it can provide a complete management solution for inventorying hardware features and settings, obtaining general system information, invoking proactive systems management functions, and providing a direct link to IBM service and support Web pages. Leveraging industry standards allows for easy integration with other systems management tools and applications.

IBM Director V4.2 Server Plus Pack is a set of tools that extends the capabilities of IBM Director for advanced server management, resulting in optimal performance and maximum availability. These five additional tools are tightly integrated with IBM Director for consistent management from a single console.

The information provided in the following sections is based on IBM Director version 4.20 unless otherwise specified.

6.1.1 Capabilities

The list of capabilities provided by IBM Director is long, with a range of features including easy installation and setup, self-managing smart tools, and seamless integration into leading workgroup and enterprise systems management environments. Because IBM Director is based on industry standards, you can use many of the tools it provides to manage non-IBM hardware.

IBM Director's capabilities can be divided into the following categories:

- Inventory

  A critical first step in any systems management strategy is to understand exactly what hardware exists in the environment and how it is configured. IBM Director performs a thorough inventory scan of each managed system it discovers. Hundreds of hardware and software data points are collected and stored in the IBM Director database. Inventory collection can be repeated both manually and through multiple automated processes.

- Hardware status

  The moment you install IBM Director, it starts working to let you know about hardware problems that occur on managed IBM xSeries systems. If there is a problem with a power supply, fan, voltage regulator module (VRM), network interface card (NIC), or other hardware, IBM Director informs you of the problem and identifies the affected system.

- Event management

  At the heart of any systems management solution is the ability to alert the IT staff in the event of a system problem. IBM Director provides a unique and very powerful method of alerting called Event Action Plans, which enables you to define event triggers independently from actions that might be taken. Simply combine these two items into customized plans for action and assign them to individual or groups of managed systems.

- Process management

  Using the Process Management task in IBM Director, you can keep track of all important processes running in your environment. IBM Director can alert you if any monitored process starts, stops, or fails to start. On managed systems running a supported Windows operating system, you can also see the status of every service and device driver installed.

- Resource management

  Resource management is an important aspect of keeping an IT environment running at peak efficiency. It is important to know whether any given system is overloaded and not able to keep up with the workload demand. IBM Director provides the ability to monitor hundreds of system resources, set individual or group thresholds for these resources, and alert you in the event that a resource threshold has been exceeded.

- Remote management

  IBM Director was built to perform remote management. Any management task that can be performed on a local system can also be performed on a system thousands of miles away, provided network connectivity is available. In addition, the Remote Control task in IBM Director allows you to take control of any managed system in your managed environment.

- Update management

  IBM has provided exceptional update management through the Update Xpress CD. This CD contains a self-starting program that allows you to maintain your system firmware and device drivers at the most current levels. A key enhancement in IBM Director 4.1 is the addition of the Update Assistant, a wizard-based tool that allows you to import Update Xpress updates for distribution to any managed systems in your environment.

- Mass configuration

  One of the advantages of managing systems using IBM Director is in its ability to make certain configuration changes on multiple managed systems at once. Even in a dynamic host control protocol (DHCP) environment, many critical servers tend to use static addresses. Using mass configuration profiles, you can, for example, change the IP address these managed systems use to locate their primary DNS server, all without having to physically visit each system.

- SNMP management

  In addition to the sophisticated management capabilities IBM Director enables for systems running the IBM Director Agent, any SNMP device can be discovered and managed as well. IBM Director can send and receive SNMP traps and convert these traps into native IBM Director alerts, delivering more helpful information than a raw SNMP trap normally can provide.

6.1.2 IBM Director environment

IBM Director is designed to manage a complex environment that contains numerous servers, desktop computers, workstations, mobile computers and assorted devices. IBM Director can manage up to 5000 systems.

An IBM Director environment contains the following groups of hardware:

- One or more servers on which IBM Director Server is installed

  These servers are called management servers.

- Servers, workstations, desktop computers, and mobile computers that are managed by IBM Director

  These systems are called managed systems.

- Network devices, printers, or computers that have Simple Network Management Protocol (SNMP) agents installed or embedded

  These devices are called SNMP devices.

6.2 IBM Director components

The IBM Director software has three components:

- IBM Director Server (6.2.1, “IBM Director Server” on page 105)
- IBM Director Agent (6.2.2, “IBM Director Agent” on page 105)
- IBM Director Console (6.2.3, “IBM Director Console” on page 106)

IBM Director Server must be installed on the management server. When you install IBM Director Server on Microsoft Windows or Linux, IBM Director Agent and IBM Director Console are installed automatically also. When you install IBM Director Server on IBM i5/OS™, IBM Director Agent is installed automatically also.

IBM Director Agent must be installed on each system that you want to manage.

IBM Director Console must be installed on each system from which a system administrator will remotely access the management server using the graphical user interface (GUI). A system on which IBM Director Console is installed is a management console.

Figure 6-1 on page 105 shows where the IBM Director software components are installed in a basic IBM Director environment.

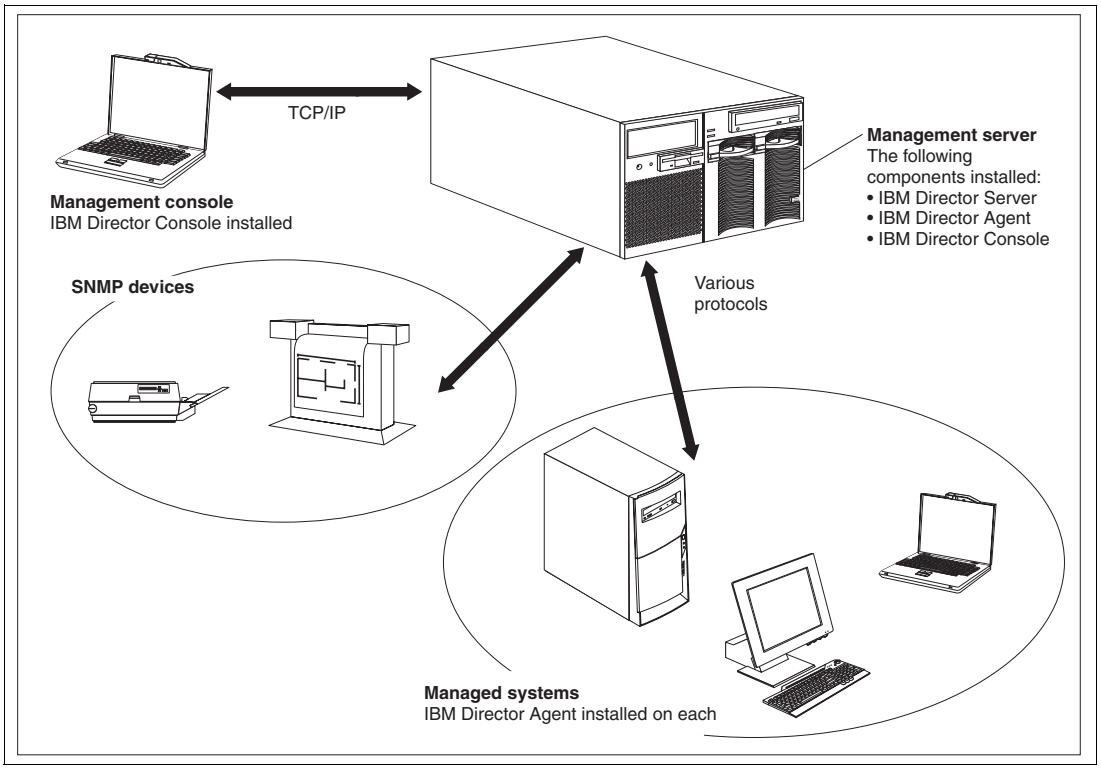


Figure 6-1 Typical IBM Director management environment

6.2.1 IBM Director Server

IBM Director Server is the main component of IBM Director. It contains the management data, the server engine, and the application logic. IBM Director Server provides basic functions such as discovery of the managed systems, persistent storage of configuration and management data, an inventory database, event listening, security and authentication, management console support, and administrative tasks.

IBM Director Server stores the inventory data in a Structured Query Language (SQL) database. You can access information that is stored in this relational database even when the managed systems are not available.

Every IBM xSeries server and IBM @server BladeCenter unit comes with an IBM Director Server license. You can purchase additional IBM Director Server licenses for installation on non-IBM servers.

6.2.2 IBM Director Agent

IBM Director Agent provides management data to IBM Director Server. Data can be transferred using several network protocols, including Transmission Control Protocol/Internet Protocol (TCP/IP), Network Basic Input/Output System (NetBIOS), Internetwork Package Exchange (IPX), and Systems Network Architecture (SNA). IBM Director Server can communicate with all systems in your network that have IBM Director Agent installed.

The IBM Director Agent features vary according to the operating system on which IBM Director Agent is installed. For example, you can install Web-based Access only on Windows 32-bit operating systems.

All IBM @server Intel-compatible servers, IBM @server JS20 blade servers, IBM NetVista™ desktop computers, IBM ThinkCentre™ desktop computers, IBM PC desktop computers, IBM IntelliStation® workstations, IBM ThinkPad® mobile computers, IBM TotalStorage® Network Attached Storage (NAS) products, and IBM SurePOS™ point-of-sale systems come with a license for IBM Director Agent. You can purchase additional licenses for non-IBM systems.

6.2.3 IBM Director Console

IBM Director Console is the GUI for IBM Director Server. Data is transferred between IBM Director Console and IBM Director Server through TCP/IP. Using IBM Director Console, you can conduct comprehensive systems management using either a drop-and-drag action or a single click. When you install IBM Director Console on a system, IBM Director Agent is not installed automatically. If you want to manage the system on which you have installed IBM Director Console, a management console, you must install IBM Director Agent on that system also. You may install IBM Director Console on as many systems as needed. IBM Director includes an unlimited-use license for IBM Director Console.

6.2.4 IBM Director Agent features

When you install IBM Director Agent, you have the opportunity to install the following features.

ServeRAID Manager

ServeRAID Manager works with xSeries servers that contain a ServeRAID adapter or an integrated small computer system interface (SCSI) controller with redundant array of independent disks (RAID) capabilities. Using ServeRAID Manager, you can monitor and manage RAID arrays without taking the servers offline.

Note: ServeRAID Manager is not supported on VMware console or guest operating systems.

Management Processor Assistant Agent

Management Processor Assistant (MPA) Agent works with xSeries and @server servers that contain one of the following service processors or adapters:

- Advanced System Management processor (ASM processor)
- Advanced System Management PCI adapter (ASM PCI adapter)
- Integrated system management processor (ISMP)
- Intelligent Platform Management Interface (IPMI) baseboard management controller
- Remote Supervisor Adapter
- Remote Supervisor Adapter II

You must install MPA Agent in order to use the MPA task to configure, monitor, and manage the service processors.

MPA Agent handles in-band communication between service processors and IBM Director Server. MPA Agent also provides in-band alert notification for certain managed systems running Linux and NetWare. For managed systems running Linux, if System Health Monitoring is not supported on a server, MPA Agent handles in-band alert notification. For managed systems running NetWare, if supported by the service processor, MPA Agent handles in-band alert notification.

IBM Director Remote Control Agent

You can use IBM Director Remote Control Agent to perform remote desktop functions on managed systems. From IBM Director Console, you can control the mouse and keyboard of a managed system on which IBM Director Remote Control Agent has been installed. This feature is supported only on Windows 32-bit and 64-bit operating systems.

Web-based Access

When you install Web-based Access on a managed system, you can access IBM Director Agent and view real-time asset and health information about the managed system from either a Web browser or the Microsoft Management Console (MMC). This feature is supported only on Windows 32-bit operating systems.

Web-based Access help files

These are the help files for the Web-based Access interface. They provide information about the managed-system data that is available when you use Web-based Access, as well as instructions for performing administrative tasks. Web-based Access is supported only on Windows 32-bit operating systems.

System Health Monitoring

System Health Monitoring provides active monitoring of critical system functions, including system temperatures, voltages, fan speeds, and power state. It produces and relays hardware alerts to the operating-system event log, IBM Director Server, and other management environments. This feature can be installed only on Windows 32-bit operating systems.

Notes: System Health Monitoring is operating-system dependent. You must choose the feature based on your operating system.

- For managed systems using Windows, you must install System Health Monitoring if you want to monitor the system hardware and send in-band alerts.
- For managed systems using Linux, System Health Monitoring is supported on some xSeries servers. It is not an installable IBM Director Agent feature, but is built into IBM Director Agent.

SNMP Access and Trap Forwarding

This feature enables SNMP as a protocol for accessing managed-system data. This enables SNMP-based managers to poll managed systems and receive their alerts. If System Health Monitoring is enabled also, this feature enables hardware alerts to be forwarded as SNMP traps.

Note: For managed systems running Linux, SNMP Access and Trap Forwarding is not an installable IBM Director Agent feature but is built into IBM Director Agent.

6.3 IBM Director extensions

Extensions are tools that extend the functionality of IBM Director. IBM Director extensions include the IBM Director Server Plus Pack, IBM Director Software Distribution (Premium Edition), IBM Remote Deployment Manager, IBM Scalable Systems Manager, IBM Virtual Machine Manager, and others.

IBM Director Server Plus Pack

The IBM Director Server Plus Pack contains a portfolio of tools that extend the functionality of IBM Director. These advanced server management tools are specifically designed for use on xSeries and Netfinity servers. The Server Plus Pack contains the following extensions:

- Active PCI Manager
- Capacity Manager
- Rack Manager
- Software Rejuvenation
- System Availability

The IBM Director Server Plus Pack CD is offered for purchase at an additional fee. For more information, contact your IBM marketing representative.

Active PCI Manager

Active PCI Manager works with the xSeries 235, 255, 345, 360, 365, 440 and 445 servers and the RXE-100 Remote Expansion Enclosure.

Using Active PCI Manager, you can manage peripheral component interconnect (PCI) and peripheral component interconnect-extended (PCI-X) adapters. Active PCI Manager contains two subtasks: Fault Tolerant Management Interface (FTMI) and Slot Manager (previously released as Active PCI Manager). Using FTMI, you can view network adapters that are members of fault-tolerant groups. You also can perform offline, online, failover, and eject operations on the displayed adapters. Using Slot Manager, you can display information about PCI and PCI-X adapters, analyze PCI and PCI-X performance, and determine the best slots in which to install PCI and PCI-X adapters.

Capacity Manager

Using Capacity Manager, you can monitor critical resources such as processor utilization, hard disk capacity, memory usage, and network traffic. Capacity Manager can identify current or latent bottlenecks for an individual server or a group of servers. It generates performance analysis reports that recommend ways to prevent diminished performance or downtime. It also forecasts performance trends.

Rack Manager

Using the Rack Manager drag-and-drop interface, you can build a realistic, visual representation of a rack and its components. By clicking an element in the visual representation, you can access detailed information (such as system health and inventory data) for the rack component.

Software Rejuvenation

Using Software Rejuvenation, you can avoid unplanned system outages due to resource exhaustion. As software runs over long periods of time, operating systems steadily consume resources and might fail to relinquish them properly. This phenomenon, known as *resource exhaustion* or *software aging*, can eventually lead to ineffective operation or even system failure. Software Rejuvenation monitors operating-system resources, predicts system outages, and generates resource exhaustion events. After being notified, you can take corrective action before a failure occurs.

You also can use Software Rejuvenation to automate the process of restarting operating systems, applications, and services at convenient times and in advance of actual failures. Because Software Rejuvenation is cluster aware, you can use it to restart a node without taking the cluster offline.

System Availability

Using System Availability, you can document and track server availability. System Availability accurately measures server uptime and downtime and provides several graphical representations of this information. It helps you to notice patterns concerning system availability.

IBM Director Software Distribution Premium Edition

IBM Director Software Distribution Premium Edition adds functions to the IBM Director Software Distribution task. You can use the base IBM Director Software Distribution task to import IBM software and build software packages using the Update Assistant wizard. When you purchase and install IBM Director Software Distribution Premium Edition, you can accomplish the following additional tasks:

- Import both IBM and non-IBM software and build software packages using the wizards that are designed for the following platforms: AIX®, i5/OS, Linux, and Windows.
- Back up or export a software package for use on another management server.
- Import a software package that was created by another management server.

IBM Software Distribution Premium Edition is offered for purchase for an additional fee. For more information, contact your IBM marketing representative.

IBM Remote Deployment Manager

IBM Remote Deployment Manager (RDM) is a flexible and powerful tool for configuring, deploying, and retiring systems. Using RDM, you can accomplish the following deployment tasks:

- Update system firmware
- Modify configuration settings
- Install operating systems
- Back up and recover primary partitions
- Securely erase data from disks

RDM supports both customized and scripted deployments. In addition, because it uses industry-standard protocols to wake and discover target systems, RDM does not require an agent component.

RDM is offered for purchase for an additional fee. For more information, contact your IBM marketing representative.

IBM Scalable Systems Manager

You can use Scalable Systems Manager (SSM) for viewing, configuring, and managing static hardware partitions on supported xSeries servers. Using Scalable Systems Manager, you can perform the following tasks:

- View information about predefined scalable systems and scalable partitions that is saved in non-volatile random access memory (NVRAM)
- Configure and manage additional scalable systems and scalable partitions
- Configure RXE-100 Remote Expansion Enclosures that are attached to servers that are used in scalable partitions

Because SSM communicates with servers out-of-band through their service processor, it does not require an agent component. You can download SSM from the IBM Support Web site:

`http://www.ibm.com/servers/eserver/xseries/systems_management/ssm.html`

IBM Virtual Machine Manager

IBM Virtual Machine Manager (VMM) enables the use of VMware VirtualCenter and Microsoft Virtual Server in an IBM Director environment. When VMM and these virtualization applications are installed, you can perform the following tasks from IBM Director Console:

- Correlate relationships between physical platforms and virtual components
- Report status of physical platforms and their corresponding virtual components
- Log in to the management interface of the virtualization application
- Discover virtual components
- Perform power operations on virtual machines
- Create event action plans that involve virtual objects

Additionally, for environments running VMware VirtualCenter, VMM provides the ability to move a running virtual machine between two physical hosts.

You can download VMM from the IBM Support Web site:

http://www.ibm.com/servers/eserver/xseries/systems_management/xseries_sm/vmm.html

Additional IBM Director extensions

IBM provides additional IBM Director extensions that you can download from the IBM Support Web site:

http://www.ibm.com/servers/eserver/xseries/systems_management/xseries_sm.html

Electronic Service Agent™

Electronic Service Agent tracks and captures system-inventory data, and if the system is under a service agreement or within the warranty period, automatically reports hardware problems to IBM

Real Time Diagnostics

Real Time Diagnostics enable you to run industry-standard diagnostic utilities on xSeries servers while they are running IBM can add or withdraw extensions on the IBM Support Web site without notice.

Tape Drive Management Assistant

Tape Drive Management Assistant integrates with IBM Director and generates alerts so that all network management tasks can be performed through IBM Director. Regular cleaning helps to extend the life of tape drives, and ensures that your data will be available when you need it.

Dynamic System Analysis

IBM Dynamic System Analysis (DSA) collects and analyzes system information to aid in diagnosing system problems.

Diagnostic Data Capture

IBM Diagnostic Data Capture is a program that captures, analyzes and manages memory dumps that are generated from operating system failures on supported xSeries systems. With these features, Diagnostic Data Capture can help you reduce problem-determination time and enable trend analysis.

6.4 IBM Director licensing and requirements

Every IBM xSeries server and @server BladeCenter unit comes with an IBM Director Server license. This license includes authorizations for the following installations:

- One installation of IBM Director Server
- 20 installations of IBM Director Agent on non-IBM systems
- Unlimited installations of IBM Director Console

Most IBM Intel-compatible systems come with a license for IBM Director Agent. For a complete list of IBM Intel-compatible systems and @server JS20 blade servers that are entitled to an IBM Director Agent license, see the *IBM Director Hardware and Software Compatibility* document. Download this PDF file from the IBM Director Agent Web page at this URL:

hhttp://www-1.ibm.com/servers/eserver/xseries/systems_management/sys_migration/ibmdiragent.html

You can purchase additional licenses for non-IBM systems, if needed. For more information, contact your IBM marketing representative.

The license to install IBM Director Server also includes the right to install the Server Plus Pack on the management server. This allows you to use the Server Plus Pack extensions, except for Rack Manager, on the management server only. To install the Server Plus Pack on managed systems or Rack Manager on the management server, you must purchase additional licenses. Contact your IBM marketing representative for more information.

6.4.1 Upgrading from previous releases of IBM Director

If you are running one of the following versions of IBM Director on a supported operating system, you can upgrade to IBM Director 4.20:

- IBM Director 3.1
- IBM Director 3.1.1
- IBM Director 4.1
- IBM Director 4.10.2
- IBM Director 4.11
- IBM Director 4.12

Versions of IBM Director earlier than IBM Director 3.1 are not compatible with IBM Director 4.20. IBM Director Server 4.20 can manage systems running IBM Director Agent, version 3.1 or later. This enables you to manage systems that are running operating systems that are not supported by IBM Director 4.20.

IBM Director Server and IBM Director Console must be at the same release level. If you upgrade IBM Director Server, you must upgrade IBM Director Console also.

If IBM Director Console and IBM Director Agent are installed on the same system, both software components must be at the same release level as IBM Director Server.

If the IBM SMBus device driver for Linux, version 4.1, 4.11, or 4.12, is installed on a managed system, you must uninstall the device driver and then install the IBM SMBus device driver, version 4.20.

6.4.2 Operating systems supported by IBM Director components

This section lists the operating systems upon which IBM Director Server, IBM Director Agent, and IBM Director Console are supported. Consider the following restrictions concerning operating-system support:

- To install IBM Director Agent on the following operating systems, you can use either IBM Director Multiplatform or the IBM Director software that came with your BladeCenter unit:
  - AIX 5L™, Version 5.2
  - Red Hat Enterprise Linux AS, version 3.0, for IBM PowerPC® (iSeries and pSeries)
  - SUSE LINUX Enterprise Server 8 for IBM pSeries and IBM iSeries

  The software for these installations also can be downloaded from the IBM Support Web site.
- To install IBM Director Agent or IBM Director Server on i5/OS (formerly OS/400®), you must use IBM Director Multiplatform, which is installed using the IBM Virtualization Engine™.

For the most recent list of supported operating systems, see the *IBM Director Hardware and Software Compatibility* document. This PDF file is updated every six to eight weeks. You can download it from:

http://www.ibm.com/servers/eserver/xseries/systems_management/sys_migration/ibmdiragent.html

IBM Director Server

You can install IBM Director Server on the following operating systems:

- i5/OS, Version 5 Release 3
- Red Hat Linux Advanced Server, version 2.1 (Update 3 required)
- Red Hat Enterprise Linux AS, version 2.1 (Update 3 required)
- Red Hat Enterprise Linux AS, version 3.0, for Intel x86
- Red Hat Enterprise Linux ES, versions 2.1 and 3.0
- SUSE LINUX Enterprise Server 8 for x86 (Service Pack 3 required)
- Windows 2000, Advanced Server and Server Editions (Service Pack 3 required)
- Windows Server 2003, Enterprise, Standard, and Web Editions.

IBM Director Agent

You can install IBM Director Agent on the following operating systems:

- AIX 5L, Version 5.2 (Recommended Maintenance Package 5.2.00-03 or later required)
- i5/OS, Version 5 Release 3
- Novell NetWare, versions 6.0 and 6.5
- Red Hat Linux Advanced Server, version 2.1 (Update 3 required)
- Red Hat Enterprise Linux AS, version 2.1 (Update 3 required)
- Red Hat Enterprise Linux AS, version 3.0, for Intel x86
- Red Hat Enterprise Linux ES and WS, versions 2.1 and 3.0
- Red Hat Enterprise Linux AS, version 3.0, for AMD64 (64-bit)
- Red Hat Enterprise Linux AS, version 3.0, for IBM PowerPC (on iSeries and pSeries)
- Red Hat Enterprise Linux AS, version 3.0, for Intel Itanium (64-bit)
- SUSE LINUX Enterprise Server 8 for AMD64 (Service Pack 3 required)

- SUSE LINUX Enterprise Server 8 for IBM pSeries and IBM iSeries (Service Pack 3 required)
- SUSE LINUX Enterprise Server 8 for Itanium Processor Family (Service Pack 3 required)
- SUSE LINUX Enterprise Server 8 for x86 (Service Pack 3 required)
- VMware ESX Server Version 1.5.2 (Patch 3 required) with the following guest operating systems:
  - Red Hat Linux Advanced Server, version 2.1 (Update 3 required)
  - Windows NT 4.0 Workstation (Service Pack 6a or later required)
  - Windows NT 4.0 Server, Enterprise and Standard Editions (Service Pack 6a or later required)
  - Windows 2000, Advanced Server, Professional, and Server Editions (Service Pack 3 or later required)
  - Windows Server 2003, Enterprise, Standard, and Web Editions
- VMware ESX Server Version 2.0, with the following guest operating systems:
  - Red Hat Linux Advanced Server, version 2.1 (Update 3 required)
  - Red Hat Enterprise Linux AS, version 2.1 (Update 3 required)
  - SUSE LINUX Enterprise Server 8 for x86 (Service Pack 3 required)
  - Windows NT 4.0 Server (Service Pack 6a or later required)
  - Windows 2000, Advanced Server, Professional, and Server Editions (Service Pack 3 or later required)
  - Windows Server 2003, Enterprise, Standard, and Web Editions
- VMware ESX Server Version 2.0.1, with the following guest operating systems:
  - Red Hat Linux Advanced Server, version 2.1 (Update 3 required)
  - Red Hat Enterprise Linux AS, versions 2.1 (Update 3 required)
  - Red Hat Enterprise Linux AS, version 3.0, for Intel x86
  - SUSE LINUX Enterprise Server 8 for x86 (Service Pack 3 required)
  - Windows NT 4.0 Server (Service Pack 6a or later required)
  - Windows 2000, Advanced Server, Professional, and Server Editions (Service Pack 3 or later required)
  - Windows Server 2003, Enterprise, Standard, and Web Editions
- VMware ESX Server Version 2.1, with the following guest operating systems:
  - Red Hat Enterprise Linux AS, version 2.1 (Update 3 required)
  - Red Hat Enterprise Linux AS, version 3.0, for Intel x86
  - SUSE LINUX Enterprise Server 8 for x86 (Service Pack 3 required)
  - Windows NT 4.0 Server (Service Pack 6a or later required)
  - Windows 2000, Advanced Server and Server Editions (Service Pack 3 or later required)
  - Windows XP Professional Edition (Service Pack 1 required)
  - Windows Server 2003, Enterprise, Standard, and Web Editions
- Windows NT 4.0 Workstation (Service Pack 6a or later required)

- Windows NT 4.0 Server, Standard, Enterprise, and Terminal Server Editions (Service Pack 6a or later required)
- Windows NT 4.0 Server with Citrix MetaFrame (Service Pack 6a or later required)
- Windows 2000, Advanced Server, Datacenter Server, Professional, and Server Editions (Service Pack 3 or later required)
- Windows XP Professional Edition (Service Pack 1 or 1a recommended)
- Windows Server 2003, Enterprise, Datacenter, Standard, and Web Editions
- Windows Server 2003, Datacenter and Enterprise Editions, 64-bit versions

IBM Director Console

You can install IBM Director Console on the following operating systems:

- Red Hat Linux Advanced Server, version 2.1 (Update 3 required)
- Red Hat Enterprise Linux AS, version 2.1 (Update 3 required)
- Red Hat Enterprise Linux AS, version 3.0, for Intel x86
- Red Hat Enterprise Linux ES, versions 2.1 and 3.0
- SUSE LINUX Enterprise Server 8 for x86 (Service Pack 3 required)
- Windows 2000, Advanced Server, Professional, and Server Editions (Service Pack 3 required)
- Windows XP Professional Edition (Service Pack 1 or 1a recommended)
- Windows Server 2003, Enterprise, Standard, and Web Editions

6.5 IBM Director in a VMware environment

Figure 6-2 on page 115 shows where IBM Director fits into a VMware ESX Server environment.

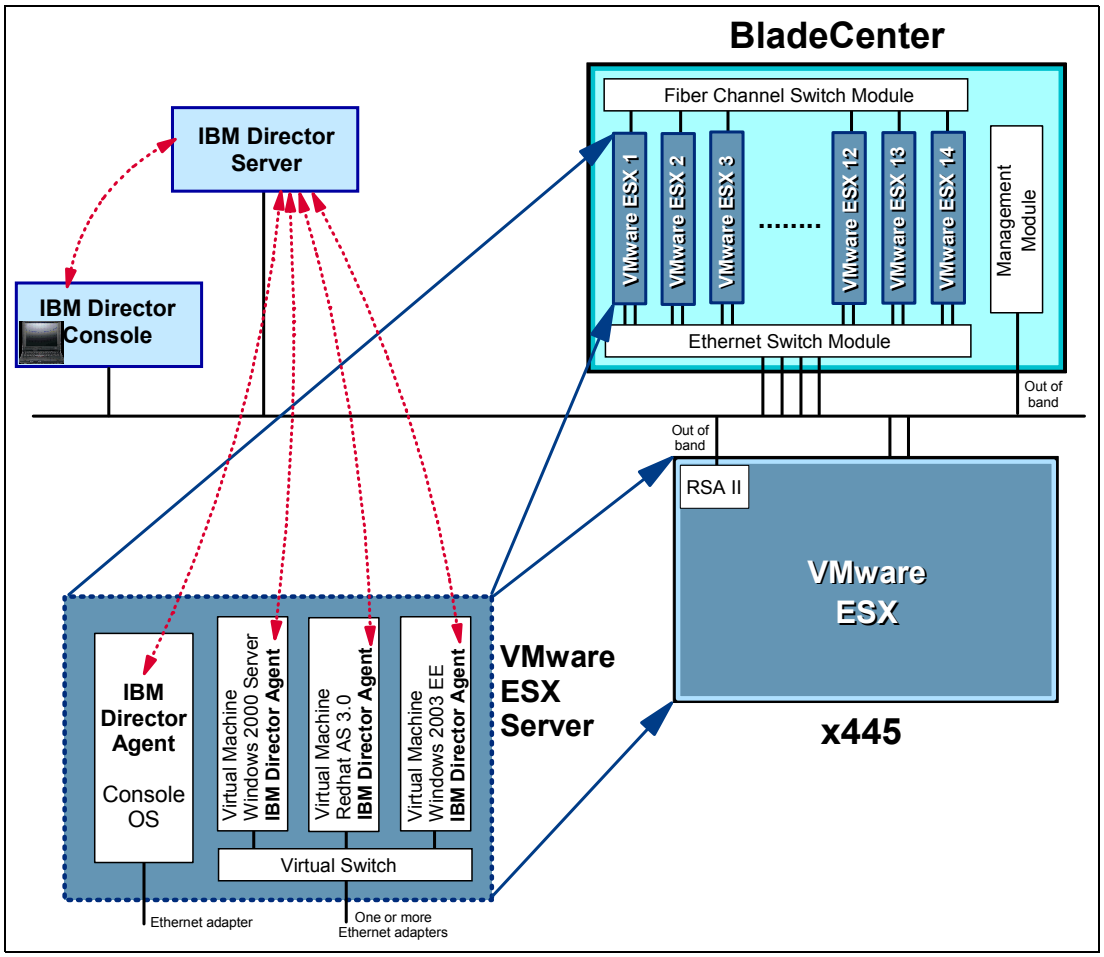


Figure 6-2 IBM Director in a VMware ESX Server environment

6.5.1 IBM Director 4.20 Support for VMware ESX Server V2.1

The following table shows the IBM Director 4.20 features that are supported on the VMware Console OS and within the VMware Guest operating systems. Refer to “IBM Director Agent” on page 112 for a list of supported guest operating systems running under VMware ESX Server V2.1.

Table 6-1 IBM Director 4.20 features supported with VMware ESX Server V2.1

Feature	VMware Guest OS	VMware Console OS
Software Deployment		
SW Distribution Standard Edition	Yes	Yes
SW Distribution Premium Edition	Yes	Yes
Update Assistant	Yes	Yes[1]
RDM	No (Future)	Yes (With RDM 4.20)
Platform Management		
Inventory - Hardware	Yes	Yes
Inventory - Software	Yes	Yes

Feature	VMware Guest OS	VMware Console OS
AssetID	Yes	Yes
Event Action Plans	Yes	Yes
Hardware Alerts	No	No
Hardware Health Status	Yes	Yes
Resource Monitors	Yes	Yes
Event Log	Yes	Yes
Management Processor Assistant (MPA)	No	No
ServerRAID Manager	No	No
Discovery	Yes	Yes
Process Management	Yes	Yes
Mass Configuration	Yes	Yes
Bladecenter Assistant	n/a	n/a
CIMBrowser	Yes	Yes
DMI Browser	Yes (Windows)	No
SNMP Alerts	Yes	Yes (VMware SNMP Agent)
SNMP Browser	Yes	Yes
Advanced Server Management		
Capacity Manager	Yes	Yes
System Availability	Yes	Yes
Active PCI Manager	No	No
Software Rejuvenation	Yes	No
Rack Manager	n/a	n/a
Application Workload Manager	Yes	Yes
Real Time Diagnostics	No	No
Scalable Systems Manager	n/a	No
Help Desk and Troubleshooting		
Remote Control	Yes (Windows)	No
Remote Session	Yes	Yes
File Transfer	Yes	Yes
Upward Integration		
Tivoli Enterprise™	No	No[2]
Tivoli NetView®	No	No[2]
CA Unicenter	No	No[2]

Feature	VMware Guest OS	VMware Console OS
HP Openview	No	No[2]
Net IQ	No	No[2]
BMC Patrol	No	No[2]
Microsoft SMS	No	No[2]

Notes:
1. Limited by package installation capabilities provided with VMware Console Operating System
2. Out of band management through SNMP trap from RSA I/II/II-EXA

6.6 IBM Virtual Machine Manager extension

IBM Virtual Machine Manager (VMM) is an extension to IBM Director that allows you to manage both physical and virtual machines from a single console. With VMM, you can manage VMware ESX Server environments using IBM Director. VMM also integrates VMware VirtualCenter and IBM Director for advanced virtual machine management.

The benefits of Virtual Machine Manager are:

- Makes it easy to manage instances of VMware ESX Server running on multiple physical platforms
- Provides single glass management of both physical and virtual systems
- Easily administer VMware ESX Server in IBM eServer BladeCenter environments by accessing all VMware ESX Server instances from one view
- Hardware health alerts can drive VMware's VMotion technology
- Enables increased availability and serviceability
- Allows you to perform predictive failure analysis, virtual machine migration (VMware) and provisioning
- Enables enterprise-wide on-demand power conservation
- Integration into IBM Director reduces training costs by providing a consistent look and feel and a familiar single point of management
- Allows you to take advantage of IBM Director's existing facilities for alerting, event action plans, security and system health
- Available at no additional cost to IBM clients

Virtual Machine Manager allows you to:

- See relationships between physical and VMM objects, for example:
  - VirtualCenter management server
  - Farm grouping of physical hosts
  - Physical host, the xSeries server running an OS
  - VM, instances of guest operating systems running on the physical host
- Review status of physical hosts and virtual machines
- See events that have occurred on the physical host or virtual machine
- Move running virtual machines manually or automatically between two physical hosts in the same farm
- Start, shutdown, suspend, resume and restart VMs

- Add and remove physical hosts to and from farms
- Invoke the VMware user interface
- Discover virtualization components

6.6.1 Sample VMware environment with IBM Director and VMM

VMM includes the following software components:

- VMM Server (“VMM Server” on page 118)
- VMM Console (“VMM Console” on page 118)
- VMM Agent (“VMM Agent” on page 119)

Figure 6-3 shows how VMM is integrated with IBM Director, VMware and VirtualCenter.

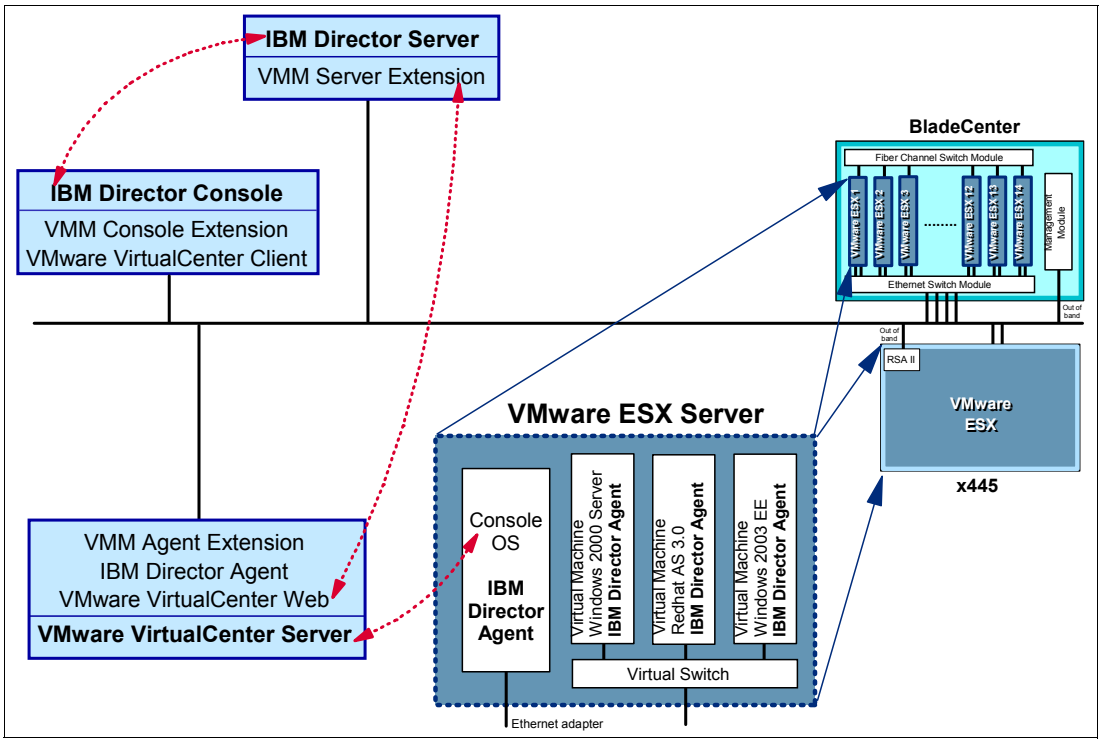


Figure 6-3 IBM Director and VMM with VMware VirtualCenter

VMM Server

VMM Server is the main component of Virtual Machine Manager because it maintains the VMM environment and manages all VMM operations. VMM Server communicates with VMM Agent and VMM Console to display status and perform operations for VMM objects. VMM Server communicates with IBM Director Server to provide event filtering and event actions that support IBM Director event action plans that involve VMM objects.

VMM Server must be installed on the management server. When you install VMM Server, VMM Console is installed automatically.

VMM Console

VMM Console is the graphical user interface (GUI) component of VMM. VMM Console must be installed on any management console from which a system administrator will remotely access the management server and perform VMM tasks.

Note: If you are using VMware VirtualCenter with IBM Director, the VMware VirtualCenter client must be installed on the system where IBM Director Console and VMM Console are installed. Otherwise, you cannot start the VMware VirtualCenter client with the Start Vendor Software subtask of VMM.

VMM Agent

VMM Agent performs the following functions:

- Discovers virtualization application and creates VMM objects
- Relays communication between VMM Server and the virtualization application
- Translates events provided by the virtualization application into events that can be used in IBM Director event action plans

6.6.2 IBM Virtual Machine Manager usage scenarios

VMM is an excellent tool for providing a single and consistent management interface to an organization's infrastructure. Using VMM allows a VirtualCenter architecture to leverage some of the excellent functions of IBM Director and respond to hardware-specific, operating system or administrative events.

Example 1: Predictive Failure Analysis

In this example, a server hosting VMware ESX Server detects that there is a potential problem with a hardware component. This could include the detection of too many hard disk errors, too many memory errors, too many CPU errors, fan problems, power problems, and so on.

In many cases, an IBM server is capable of predicting when hardware is about to fail. This is called Predictive Failure Analysis (PFA). When this happens an alert is stored in the servers log, an alert light is displayed on the server and, if using IBM Director, the alert is transmitted to the Director server.

When the Director server receives an alert it can generate an event which, in turn, can initiate an event action plan. The action plan can be configured to perform various tasks, including a VMM integrated option to automatically and seamlessly migrate some or all virtual machines, using Vmotion, from the failing server to other physical servers. This operation is transparent to the users and will reduce the chance of downtime.

Example 2: Scheduling virtual machine migration

In some cases, it might be desirable to migrate, using Vmotion, one or more VM outside of normal operating hours. This is possible using a similar mechanism as in Example 1: Predictive Failure Analysis.

With IBM Director, it is possible to generate custom events using a command-line tool. Providing IBM Director is configured with the appropriate event action plan, the command-line tool can be initiated at a particular time using the scheduler that comes with Windows.

Example 3: Zero-downtime maintenance

VMotion can be used to move running VM from one physical server to another to allow hardware maintenance to be performed. This scenario provides a method for upgrading, replacing or upgrading hardware without impacting the availability of services.

Once the hardware service is completed, the virtual machines can be migrated back using VMotion. An example of zero-downtime maintenance is provided in 5.7, “VMotion usage scenario: Zero-downtime maintenance” on page 97.

7

VMware P2V Assistant

Migration of an operating system and all its applications from a physical server to a virtual machine can be a complex, and, sometimes daunting, task. VMware provides a migration tool, VMware P2V Assistant, that simplifies the task of transforming an image of an existing physical system into a VMware virtual machine.

This chapter provides a brief description of the VMware P2V Assistant product.

7.1 Introduction to VMware P2V Assistant

Migrating an operating system, and the applications installed on it, from a physical to virtual machine is possible with a product from VMware called P2V Assistant. *P2V* has become a widely used acronym in the IT industry to refer to the process of migrating a *physical* machine's environment to a *virtual* machine.

There are several considerations to keep in mind when migrating from physical machines to virtual machines that determine whether P2V Assistant is appropriate.

7.1.1 Migration by imaging or cloning

There are many situations where P2V Assistant can be used. These include:

- Speed up the migration process
- Unavailability of documentation and media
- No defined installation procedures, including customizations for the organization
- Appropriate skills not available or it costs too much to hire contractors
- Unknown amount of components that are required to make the application work
- You want the software on the VM to work the same way as on the physical machine
- Too costly to figure out what is installed and where, including DLL dependencies
- Minimize the risk
- Test and be confident that the solution will work as a virtual machine
- Significantly reduce the cost of outsourcing the task

A typical implementation strategy when adopting an imaging approach to migration includes the following steps:

1. Take an image of the physical server using the appropriate imaging software.
2. Shut down and turn off the physical server to ensure that no changes are made.
3. Import the image onto a virtual disk.
4. Create a virtual machine that uses the imaged virtual disk.
5. Change all the device drivers within the VM to reference the virtual hardware.
6. Perform applicable testing and put the virtual machine into production.

Leave the physical server functioning for an appropriate amount of time. This allows your organization us this server as a backup for any issues that might occur with the VM.

You can use this strategy as a preliminary step to test that the procedure works with a given physical server. The only difference is that the VM is not on the production network, and the physical server is not left in a turned off state after taking the image, rather it is returned to production immediately.

7.1.2 Migration by reinstallation

In some cases the preferred option is to create a VM, then reinstall the operating system and applications to provide the same functionality as the physical machine you are replacing. It might be preferable not to use P2V Assistant in the following situations:

- Documentation and media are available.
- There is a defined installation procedure, including customizations for the organization.
- Appropriate skills are available to perform the installation in a timely and cost-effective manner, either within the organization or from contractors.
- Support is available from the ISV, if applicable.

- There have been software components or patches added, removed and upgraded over time. One of the objectives is to start fresh.
- When upgrading to a new operating system, it is company policy to install from scratch.

A typical implementation strategy when adopting a reinstallation approach to migration includes the following steps.

1. If documentation is not available, document the environment of the physical server.
2. Create a VM and install the operating system and software.
3. If applicable, migrate the data from the physical server at an appropriate time for testing, while remaining in production on the physical server.
4. Test the new VM without impacting the production system.
5. Schedule a cutover to the VM.
6. If applicable, migrate the data from the physical machine to the VM.
7. Shut down and turn off the physical machine. Leave this server available as it is for an appropriate amount of time. This allows the organization to return to the physical server if there are problems with the VM upon first use in production.

7.2 VMware P2V Assistant overview

VMware P2V Assistant version 2.0 is an enterprise-class migration tool that transforms an image of an existing physical system into a VMware virtual machine. This tool enables fast and reliable physical to virtual machine migration for Microsoft Windows operating systems ranging from Windows NT 4 to Windows Server 2003.

By guiding the administrator through an easy to use GUI wizard, P2V Assistant performs all necessary substitutions to transform a physical system into a production-ready virtual machine. The GUI wizard performs the following tasks:

- Creates an image of the source machine with built-in imaging or a third-party imaging tool
- Performs all necessary disk controller and driver substitutions to boot the VM
- Recommends additional configuration changes to make the new VM production ready

For enterprises that want to take advantage of virtualization in their IT operations VMware P2V Assistant 2 provides:

- An automated solution for fast, reliable physical to virtual migration
- Support for a wide range of Windows platforms from NT 4 to Windows Server 2003
- Ease of use, refined through thousands of client deployments
- Market-proven mature technology supported by consulting services and authorized partners
- Compatibility with the best in class VMware software

The solutions that benefit from this best in class VM creating tool include:

- Fast and Clean Migrations of Existing Applications to VMware Virtual Machines

  VMware P2V Assistant speeds up the migration process, cuts down on set-up time and delivers value on investments in VMware software in the shortest time-frame possible by eliminating the need to re-install software and configure complex application environments. VMware P2V Assistant simplifies migration to VMware virtual infrastructure, the foundation of the next generation computing in the enterprise.

- Efficient quality assurance and debugging

  VMware P2V Assistant can capture images of production systems into VMware virtual machines and redeploy these images in a consolidated environment. Enterprises can minimize disruptions to production servers by troubleshooting problems and testing changes in this exact replica of the production environment.

- Disaster recovery and backup

  With VMware P2V Assistant, users can periodically capture production systems into a library of *offline* VMs that can be activated in the event of a disaster to minimize service disruption.

- Standardizing on virtual infrastructure

  Enterprises can move away from obsolete and orphan platforms and quickly migrate legacy applications to a consolidated VMware virtual machine environment and upgrade these servers to new hardware. With minimal downtime, you, the IT administrator, can migrate all your heterogeneous physical servers and existing VMs (VMware or not) to the best in class VMware platforms to take advantage of flexibility, cost advantages, and streamlined management that virtual infrastructure brings.

Table 7-1 describes the usage scenarios and associated benefits of using VMware ESX Server and VMware P2V Assistant.

Table 7-1 Why use VMware ESX Server and VMware P2V Assistant

Usage Scenarios	Benefits
Fast and clean migration of existing applications to VMware virtual machines	► Reduce downtime during conversion ► Reduce application installation and configuration cycle times ► Reduce risk of incorrect application configuration / data synchronization
Efficient quality assurance and debugging	► Test in realistic environments ► Improve time to market
Disaster recovery and backup	► Create exact replica of the production environment for contingency planning and disaster recovery
Standardizing on virtual infrastructure	► Migrate all your physical and existing virtual machines to the best in class VMware platforms
Legacy server migration	► Migrate legacy applications off obsolete, unsupported hardware ► Eliminate risk and need to locate and follow obsolete installation procedure and replication configuration
Pre-deployment Patch Testing	► Dramatically improve patch testing and staging procedures without exorbitant hardware budgets ► Replicate environments exactly without lengthy configuration cycles

7.3 Practical P2V

The P2V process is performed typically in three major phases:

1. Clone the physical server and import the image to a virtual disk using a helper VM.

2. Perform a system reconfiguration on the operating system within the virtual disk.
3. Create a VM using the cloned virtual disk, start it and complete any additional tasks.

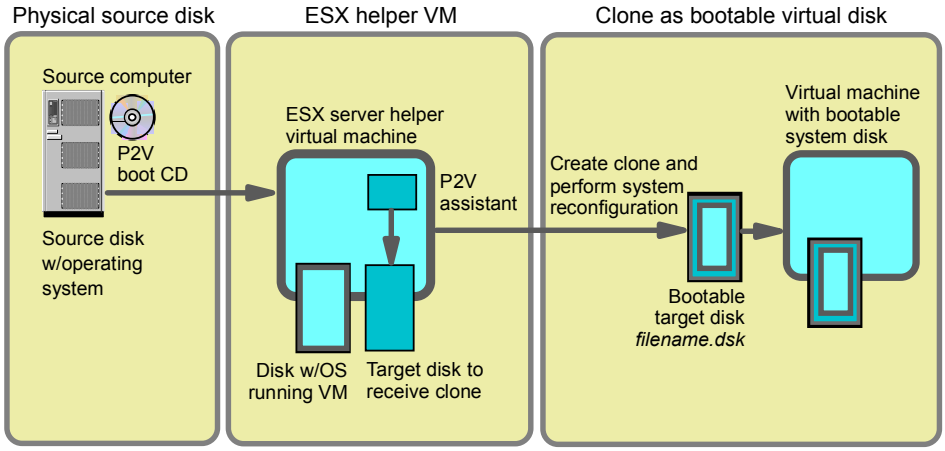


Figure 7-1 Cloning and reconfiguring in a VMware ESX Server helper virtual machine

The basic P2V Assistant processes are discussed in the following sections.

7.3.1 Cloning

Cloning is the process of creating a cloned disk, a virtual disk that is an exact copy of the source physical disk. This involves copying the data on a source machine's hard disk, then transferring that data to the target virtual disk. This process can be performed using the P2V Assistant installed on a helper machine, physical or virtual, or using a third-party disk imaging or backup and restore tool running in a helper VM. The helper machine is an intermediary machine that allows the P2V Assistant to interact with the target machine's virtual disks.

Table 7-2 lists the differences between using the P2V Assistant and a third-party tool for cloning.

Table 7-2 P2V versus third party for cloning

P2V Assistant Cloning	Third Party Imaging
Limited dynamic disk support	Typically can read dynamic disks
Cloned Disk images synchronously	Typically images asynchronously
► Limited SCSI and NIC driver support in the P2V Assistant Boot CD. ► No method to add additional drivers	► Typically requires creation of a special boot disk ► Can build the boot disk to suit specific needs
Preserves drive letter mapping	May not preserve drive letter mapping
Relatively fast imaging	Imaging may not be fast if the tool is DOS-based
Not licensed per cloned disk	Typically licensed per image created
Simplest approach, try this first, if convenient	Fallback method

7.3.2 System reconfiguration

System reconfiguration is the process of adjusting the migrated operating system to enable it to function on virtual hardware. This adjustment is performed on the target virtual disk after cloning and enables the target virtual disk to function as a bootable system disk in a VM. System reconfiguration is not required if the target virtual disk will not be used as a bootable system disk. This process is performed using P2V Assistant installed on a helper machine, physical or virtual. A system reconfiguration does not change the identity of the new VM.

7.3.3 Creating the Virtual Machine

Note: While the P2V Assistant simplifies the physical to virtual migration process by automating several steps, it is not intended to serve as an unattended migration tool for casual use. Migration is complex. Whether using the P2V Assistant or not, only technically-skilled system administrators with a good understanding of and familiarity with operating systems, drivers, imaging tools, and VMware products should perform the physical to virtual migration process.

The new VM can be created using the newly reconfigured virtual disk or disks. If the new VM disk had an operating system and you performed a system reconfiguration on it, the new VM retains the identity and settings of the original, physical machine. This includes: machine name, SID, hard disks including partitions, data files as well as application and user settings. This process requires that the cloned disk be detached from the helper VM, if that was used.

If you plan to run the new VM on the same network as the original source machine, you will need to modify the identity, name, IP address and networking of the VM so they can coexist on the same network.

Note: It is strongly recommended that anyone wanting to employ this tool on a large-scale or mission-critical basis attend VMware server training and P2V training.

A

VMware products from IBM

This appendix provides information about VMware products, licensing and support available through IBM.

VMware software available from IBM

Under a joint global distribution and support agreement, IBM can resell VMware ESX Server VM software on IBM @server xSeries and BladeCenter systems. IBM offers worldwide support for VMware ESX Server software and Linux and Microsoft Windows operating systems running within virtual machines. This support is available through IBM Technical Support Services (ITS), providing customers with a single, integrated support service for xSeries and BladeCenter systems, VMware ESX Server software and operating systems.

VMware ESX Server and VMware VirtualCenter are available through IBM xSeries channels as standard xSeries part numbers.

VMware software support is mandatory and separately available from IBM Technical Support Services. See “IBM support for VMware software” on page 130.

VMware ESX Server software from IBM

Table A-1 lists the VMware ESX Server software part numbers available from IBM.

Table A-1 Part numbers - VMware ESX Server

Description[a]	AG/AP/EMEA Part #	3 year sub Part #	5 year sub Part #
VMware ESX Server - 2 processor version	481722X	4817223	4817225
VMware ESX Server - 4 processor version	481724X	4817243	4817245
VMware ESX Server - 8 processor version	481728X	4817283	4817285
VMware ESX Server - 16 processor version	481726X	4817263	4817265
VMware Virtual SMP – 2 processor version	48173CX		
VMware Virtual SMP – 4 processor version	48173DX		
VMware Virtual SMP – 8 processor version	48173EX		
VMware Virtual SMP – 16 processor version	48173FX		
ESX Server Upgrade (2 to 4 processors)	481739X		
ESX Server Upgrade (4 to 8 processors)	48173AX		
ESX Server Upgrade (8 to 16 processors)	48173BX		
Virtual SMP Upgrade (2 to 4 processors)	4817U1X		
Virtual SMP Upgrade (4 to 8 processors)	4817U2X		
Virtual SMP Upgrade (8 to 16 processors)	4817U3X		

a. All products include a one year software maintenance subscription. Support is available separately through IBM Technical Support Services (ITS)

VMware VirtualCenter software from IBM

A license for VirtualCenter Management Server must be purchased for each managed environment. A single VirtualCenter Management Server can manage 50+ ESX Servers. One license includes unlimited seats of the VirtualCenter client.

VirtualCenter Agents unlock management of an ESX Server. All CPUs on an ESX Server must be licensed. Licenses are additive, meaning two 2 CPU agents can be used to manage a 4-way machine.

A VMotion license unlocks VMotion to and from an ESX Server through the management agent. All CPUs on an ESX Server must be licensed. Licenses are additive, meaning you can use two CPU licenses to VMotion-enable a 4-way machine. Table A-2 lists the VMware VirtualCenter software part numbers available from IBM.

Table A-2 Part number - VMware VirtualCenter

Description[a]	AG/AP/EMEA Part #	3 year sub Part #	5 year sub Part #
VirtualCenter Management Server	4817M1X	4817M13	4817M15
VirtualCenter Agent - 2 CPU	4817M2X		
VirtualCenter Agent - 4 CPU	4817M3X		
VirtualCenter Agent - 8 CPU	4817M4X		
VirtualCenter Agent - 16 CPU	4817M5X		
VirtualCenter VMotion - 2 CPU	4817M6X		
VirtualCenter VMotion - 4 CPU	4817M7X		
VirtualCenter VMotion - 8 CPU	4817M8X		
VirtualCenter VMotion - 16 CPU	4817M9X		
Virtual Infrastructure Node - 2 CPU	4817MAX	4817MA3	4817MA5
Virtual Infrastructure Node - 4 CPU	4817MBX	4817MB3	4817MB3
Virtual Infrastructure Node - 8 CPU	4817MCX	4817MC3	4817MC3
Virtual Infrastructure Node - 16 CPU	4817MDX	4817MD3	4817MD3

a. All products include a one year software maintenance subscription. Support is available separately through IBM Technical Support Services (ITS).

BladeCenter VMware bundles

Special bundling packages are available for VMware software installed on IBM BladeCenter hardware. For example, a VMware Virtual Infrastructure Node license pack for BladeCenter is available. A VMware Virtual Infrastructure Node (VIN) is comprised of ESX Server, VMware Virtual SMP, VMotion and a VirtualCenter Agent. ESX Server Virtual Infrastructure Nodes can be deployed and managed with VMware VirtualCenter to transform your IT infrastructure into virtual infrastructure.

These bundled software options must be sold with IBM BladeCenter hardware. Other VMware software such as Virtual SMP, VirtualCenter and VMotion software can be sold separately from the hardware.

Table A-3 on page 130 shows the BladeCenter VMware packages.

Table A-3 Part numbers for VMware BladeCenter offerings

Description[a]	AG/AP/EMEA Part #	3 year sub Part #	5 year sub Part #
VMware Virtual Infrastructure Node - 2 CPU license pack (fourteen 2 CPU Virtual Infrastructure Nodes)	4817BEV	4817B3X	4817B5X
VMware ESX Server and VirtualCenter Agent license pack (fourteen 2 CPU ESX Server & fourteen 2CPU VirtualCenter Agent licenses)	4817BE2		
VMware ESX Server and VirtualCenter Agent license pack for a single blade (Single 2 CPU ESX Server & 2CPU VirtualCenter Agent)	4817B12		
VMware Virtual Infrastructure Node - 4 CPU license pack (seven 4 CPU Virtual Infrastructure Nodes)	4817BAX	4817BBX	4817BCX

a. All products include a one year software maintenance subscription. Support is available separately through IBM Technical Support Services (ITS).

Purchasing VMware software from IBM

When purchasing VMware software from IBM, you receive a License Activation Certificate for each VMware product. The certificate contains an Activation Code and instructs you to register the Activation Code at the VMware Web site. After activation, you receive a product license key and access to download the latest version of VMware software.

Registration entitles you to a one year software maintenance subscription. This subscription provides you with notification and updates to purchased VMware products. Software support is not included in the base purchase price of VMware ESX Server from IBM.

IBM support for VMware software

Support for VMware software ordered through IBM is mandatory and is separately available from IBM Technical Support Services. VMware software support available through IBM includes support for VMware software and:

- Microsoft Windows or Linux running in a VMware virtual machine
- IBM Director and other xSeries systems management, including the Virtual Machine Manager plug-in to IBM Director
- IBM software supported within a VM such as WebSphere® and DB2. For IBM support for software such as DB2 and WebSphere you need support for both VMware and the IBM software application itself. In other words, you need to purchase support for WebSphere and DB2 in addition to ITS support for VMware.

VMware software is transparent to the middleware and applications that operate above the VMware guest operating system. If an IBM software problem occurs only within a VMware ESX environment, it is considered a transparency problem and directed to VMware for resolution. The IBM VMware supportline is available to provide assistance in working with our clients and our VMware IBM Business Partner to resolve this issue.

Clients using software not covered by an IBM VMware supportline agreement are required to recreate the problem in a native environment without VMware to use IBM software maintenance support services for the IBM software experiencing the problem.

Abbreviations and acronyms

AIX	Advanced Interactive eXecutive
AMD	Advanced Micro Devices
API	application programming interface
ASM	Advanced Systems Management
BIOS	basic input/output system
CDAT	consolidation discovery and analysis tool
CEC	Central Electronics Complex
CFO	chief financial officer
CIO	chief information officer
COE	common operating environment
CPU	central processing unit
CRM	customer relationship management
CTO	chief technology officer
DDR	double data rate
DIMM	dual inline memory module
DLL	dynamic link library
DLT	digital linear tape
DMZ	de-militarized zone
DNS	domain name system
DSA	dynamic system analysis
ECC	error checking and correcting
ERP	enterprise resource planning
FC	fibre channel
FTMI	Fault Tolerant Management Interface
GUI	graphical user interface
HBA	host bus adapter
HTTP	hyper-text transport protocol
IDE	integrated drive electronics
IP	Internet protocol
IPMI	Intelligent Platform Management Interface
IPX	Internetwork Package Exchange
ISMP	integrated system management processor
ISV	Independent Software Vendor
IT	information technology
ITS	Integrated Technology Services
ITSO	International Technical Support Organization
LAN	local area network
LED	light emitting diode
LUN	logical unit number
MAC	media access control
MMC	Microsoft Management Console
MP	multi-processor
MPA	Management Processor Assistant
MSCS	Microsoft Cluster Server
NAS	network attached storage
NIC	network interface card
NUMA	Non-uniform Memory Access
NVRAM	non-volatile random access memory
OPM	optical passthrough module
OS	operating system
PCI	peripheral component interconnect
PDF	portable document format
PFA	Predictive Failure Analysis
PXE	Pre-boot Extension Environment
QA	quality assurance
RAID	redundant array of inexpensive disks
RAM	random access memory
RAS	reliability, availability, servicability
RDBMS	remote DB management system
RDM	raw disk map
ROI	return on investment
RSA	Rivest, Shamir, & Adleman
RXE	Remote eXpansion Enclosure
SAN	storage area network
SCON	server consolidation
SCSI	small computer system interface
SDK	software development kit
SFP	Small Form Pluggable
SID	security ID
SMP	symmetric multiprocessing
SNA	Systems Network Architecture
SNMP	simple network management protocol
SOE	standard operation environment
SQL	structured query language
SSH	secure shell
SSM	Scalable Systems Manager
TCO	total cost of ownership
TCP/IP	transmission control protocol/internet
URL	Uniform Resource Locator

USB universal serial bus

VIN Virtual Infrastructure Node

VLAN virtual LAN

VM virtual machine

VMFS Virtual Machine File System

VMM IBM Virtual Machine Manager

VRM voltage regulator module

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this Redpaper.

IBM Redbooks

For information about ordering these publications, see “How to get IBM Redbooks” on page 137. Note that some of the documents listed here might be available in softcopy only.

- *The Cutting Edge: IBM @server BladeCenter*, REDP-3581
- *Windows Server 2003 Datacenter Edition on the IBM @server xSeries 445*, REDP-3700
- *IBM @server xSeries Server Consolidation: An Introduction,* REDP-3785
- *VMware ESX Server: Scale Up or Scale Out?*, REDP-3953
- *Implementing Systems Management using IBM Director*, SG24-6188
- *Implementing VMware ESX Server with IBM TotalStorage FAStT*, SG24-6434
- *Server Consolidation with the IBM @server xSeries 440 and VMware ESX Server*, SG24-6852
- *Server Consolidation on IBM @server pSeries Systems*, SG24-6966
- *IBM @server xSeries 445 Planning and Installation Guide*, SG24-8870

Other IBM publications

These publications are also relevant as further information sources.

- *IBM Director 4.2 Installation and Configuration Guide*
- *IBM Director 4.2 Systems Management Guide*
- *IBM Director 4.2 Virtual Machine Manager 1.0 Installation and User’s Guide*
- *VMware ESX Server: Scale Up or Scale Out*

VMware publications

The following are VMware publications:

- *VMware ESX Server 2 Administration Guide*
- *Guest Operating System Installation Guide*
- *VMware VirtualCenter User’s Manual*
- *VMware P2V Assistant User’s Manual*

The following are VMware white papers:

- *Building Virtual Infrastructure with VMware VirtualCenter*
- *Configuring and Installing IBM BladeCenter*
- *Converting Image Files into Virtual Machine Disks*
- *VMware ESX Server 2 NUMA Support*

The following are VMware product datasheets

- *VMware ESX Server*t
- *VMware VirtualCenter*
- *VMware P2V Assistant*

The following are VMware technical notes:

- *VMware ESX Server Backup Planning*
- *VMware ESX Server Configuration for Clustering*
- *VMware ESX Server 2 NUMA Support*
- *Support for Microsoft Software in VMware Virtual Machines*, VMware support statement

Online resources

These Web sites and URLs are also relevant as further information sources:

- IBM VMware Web page:

  http://www.ibm.com/servers/eserver/xseries/vmware

- VMware Web page for IBM customers:

  http://www.vmware.com/ibm

- VMware Website homepage:

  http://www.vmware.com

- VMware Product information:

  http://www.vmware.com/products/server/esx_features.html

- VMware solutions and user scenarios:

  http://www.vmware.com/solutions

- Interoperability matrixes for ESS, DS4000, DS6000 and DS8000:

  http://www.ibm.com/servers/storage/disk/ess/interop-matrix.html
  http://www.ibm.com/servers/storage/disk/ds4000/interop-matrix.html
  http://www.ibm.com/servers/storage/disk/ds6000/interop.html
  http://www.ibm.com/servers/storage/disk/ds8000/interop.html

- IBM @server xSeries literature:

  http://www.ibm.com/servers/eserver/xseries/literature.html

- IBM @server BladeCenter literature:

  http://www.ibm.com/servers/eserver/bladecenter/literature/index.html

- IBM Server Consolidation White Papers

  http://www.ibm.com/servers/eserver/literature/

- IBM Director homepage:

  http://www.ibm.com/servers/eserver/xseries/systems_management/director_4.html

- IBM ServerProven homepage:

  http://www.pc.ibm.com/us/compat/index.html

IBM Internal Resources

These Web sites and URLs are also relevant as further information sources:

- IBM VMware:

  http://vmware.ibm.com/v_dir/vmware.nsf/index.html?OpenForm

- ATS EMEA:

  http://ats.greenock.uk.ibm.com/vmware_new.htm

- L3 Kirkland:

  http://level3.kirkland.ibm.com/VMWare/

- Beaverton Support Center:

  http://w3.svc.beaverton.ibm.com/vmware/

- Beaverton SWAT page:

  http://w3.beaverton.ibm.com/groups/SCON_SWAT/

How to get IBM Redbooks

You can search for, view, or download Redbooks, Redpapers, Hints and Tips, draft publications and Additional materials, as well as order hardcopy Redbooks or CD-ROMs, at this Web site:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Index

X

Server Consolidation with VMware ESX Server

Redpaper

How to consolidate your production servers

How to simplify your server management

How to reduce your server costs

Today, IT infrastructure organizations are working hard to solve the problems created by the explosion in the scope and complexity of IT platforms adopted in the 1990s. The migration of application architectures to thin-client multi-tier architectures, the rapid introduction of four generations and multiple editions of Microsoft Windows servers and the growth of Linux have swept across IT organizations in successive waves over the last ten years. These waves caused explosive growth in server counts, network complexity and storage volumes throughout geographically distributed IT organizations. The policies and procedures adopted to gain back control of the infrastructure have often introduced their own challenges.

IT has started to regain the upper hand in the battle against costly, inflexible and disorderly infrastructure. As a first step, IT organizations have generally centralized their IT infrastructure into fewer locations for better visibility. As a second step, they are adopting a new generation of infrastructure technologies and methodologies including server consolidation.

This IBM Redpaper discusses server consolidation options and considerations using VMware ESX Server on IBM eServer BladeCenter and IBM eServer xSeries 445 hardware. In addition, systems management options using VMware VirtualCenter, VMotion, VMware P2V Assistant, and IBM Director with VMM are also discussed.

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks