EMC²
where information lives®

EMC PERSPECTIVE

The Efficient, Green Data Center

October 2008

Delivering IT with Financial and Environmental Consciousness

Table of Contents

Preface

About this Document

This white paper will help you to build an efficient, green data center that will yield financial and environmental benefits. Data center power and cooling, and virtualization are two key strategies to reap these benefits. We will discuss power and cooling optimization, as well as virtualization and other related data center technologies that are required to develop and operate an efficient, green data center. This discussion will occur in the context of a virtualization-leveraged data center.

Virtualization has been positioned as the core enabler for driving broader efficiencies. It includes server and storage procurement and utilization, information protection (backup and recovery), business continuity and disaster recovery (local and remote replication), infrastructure management, and infrastructure consolidation and automation. This paper provides you with a breadth of knowledge about the major data center functions that have a direct or indirect impact on operational and environmental efficiency.

The discussion in this document is not limited to technology, but includes other relevant areas of a data center that are either impacted by or have an impact on the technology infrastructure and IT operations. It identifies major areas of consideration as well as step-by-step guidance about how to implement power and cooling, virtualization, and other related technologies. Designs and architectural drawings for optimization are included.

How is this Document Organized?

This paper consists of three sections:

- Introduction: Focuses on financial and environmental impacts of inefficient data centers and the case for building efficient and green data centers.
- Considerations for Building Efficient, Green Data Centers: Focuses on high-level data center operating environments, IT processes, and technology considerations for data center efficiency.
- Implementing an Efficient, Green Data Center: Focuses on implementation processes, steps, and technologies; describes designs and architectural drawing in detail.

Who Will Benefit from this Document?

This paper is written for IT professionals including senior-level IT managers (CxOs, VP/Line of Business Managers) who define the strategic direction of an organization's data center operations, middle-level managers (IT directors, operations managers, project and program managers) who execute and implement strategy, and technical staff (architects, administrators) who evaluate, design, build, and manage the data center's day-to-day operations.

Any organization or individual with the following interests will find value in this paper:

- Would you like to better understand data center efficiency opportunities?
- Have you started or are you about to start IT and non-IT-related data center efficiency projects such as data center consolidation, server virtualization and consolidation, storage virtualization for data migration and mobility, storage consolidation, equipment placement planning for new or existing data centers, and new design and implementation of a cooling system?
- Have you completed efficiency projects? Are you looking for additional ideas?
- Would like to learn what steps to consider prior to and during a project?
- Are you interested in better understanding technology requirements and considerations for improving asset efficiency?

If you've answered yes to any of these questions, you will learn from this paper.

Acknowledgements

I would like to thank a few people who helped me to write and publish this paper. First, I would like to thank Tom Clancy and Christine McCarthy for motivating me to write about this topic. Next I would like to thank Tom Downey, Dick Sullivan, and those EMC® Champions who took the time to review this document and provide comments. I would also like to thank VMware® for their participation and interest in working with me on this paper. In particular, I would like to thank Rod Gilbert and Amy Lu from VMware. Rod made sure that I had access to the right individuals at VMware, and Amy provided information and reviewed this paper. Finally, I would like to thank my good friend Alok Shrivastava and his team for publishing this document; without their support you would not be reading this.

About the Author

Raza Syed is a senior manager responsible for technology sales and channel development at EMC Corporation, based in Santa Clara, California. Mr. Syed, the author of several storage industry white papers and a frequent presenter at storage industry events, has over 14 years of storage industry experience and has held several senior product management and marketing positions at Mercury Interactive/HP, VERITAS/Symantec, and Sun Microsystems. He has also held engineering positions at Sun Microsystems, Data General/CLARiiON®, Siemens/Pyramid Technology, and Seagate Technology Incorporated. Mr. Syed frequently meets with senior executives and IT specialists from different industries in the U.S. and provides counseling on various IT and storage-related topics including data center efficiency and green initiatives. Mr. Syed has also been a member of SunTone Architecture Council and has helped define IT services architecture for the industry. He has been awarded numerous system engineering awards including the achievement award for sales and business development at Sun Microsystems. Mr. Syed holds B.Sc. degrees in mathematics and physics from overseas educational institutions, and a B.Sc. in economics from the University of San Francisco, CA. He can be contacted via e-mail at: syed_raza@EMC.com.

Executive Summary

Estimates reveal that information is growing at about 55 percent (CAGR). Organizations are continuously investing in IT capital and human resources to manage this massive influx of information. The net result is that organizations are struggling to keep up with IT requirements from the business. This forces the IT team to implement tactical strategies rather than building and maintaining broader, visionary data centers. The “block and tackle” approach is leading to large-scale inefficiencies across data centers, resulting in financial loss and environmental impact.

The expanding gap between a data center’s physical and operational assets and their sub-optimal use is resulting in higher costs for hardware acquisition and maintenance, energy consumption, data center real estate, infrastructure management, and human resources. There is also a corresponding environmental cost and impact. Organizations have an opportunity to improve their data center operations and reduce cost and environmental impact by better understanding all aspects of an operational data center. This includes knowledge of data center power supply, cooling systems and infrastructures, equipment placement plans, IT infrastructure planning and design processes, and the use and operation of IT.

Introduction

Building an efficient, green data center requires increasing the utilization of existing and new IT resources, containing and improving equipment power and cooling usage and requirements, reducing the data center's physical infrastructure footprint, reducing IT administrative and maintenance costs, and optimizing staff productivity.

The two drivers for efficient, green data centers are the:

1. opportunity to realize significant, near and long-term financial benefits
2. ability to minimize the environmental impact of inefficient data centers

The Financial Impact

The average utilization of physical IT assets, including servers and storage, remains well below the 75 to 80 percent desired rates. Studies from equipment manufacturers and independent analysts reveal that the average utilization of servers is approximately 8 to 20 percent, while the average utilization of storage is approximately 35 to 40 percent. On the other end of the spectrum, the amount of new information being created continues to grow at approximately 55 percent (CAGR), requiring organizations to invest in new infrastructure to keep up with growth.

This expanding gap of inefficiency between the data center infrastructure and the used capacities results in far greater dollar-to-value cost for the infrastructure than originally anticipated or desired. Hardware inefficiency also results in inefficient energy consumption and increased management efforts. In combination, the net result is increased operating costs.

While the initial hardware (server, network, and storage equipment) procurement cost is becoming more affordable, it still remains one of the major expenses and has a trickle-down effect on other areas of data center operations that are increasingly becoming less affordable. These new areas include data center power and cooling and data center real estate. Infrastructure administration and maintenance remain the top operations expenses. The opportunity to realize financial benefits by building efficient IT can be achieved not only by improving equipment utilization rates, but also by considering associated resources and activities.

Data center electricity costs, as an example, have become the second highest expense in data center operations at 13 percent, followed by the cost of data center maintenance and administration, which is about 67 percent of the overall operations expense.(1) Businesses paid about 20 percent more for electricity in 2005 than they did in 2004. It should also be noted that data centers in 2006 used 61 billion kilowatt-hours of electricity—twice the energy they consumed in 2000 and an output equivalent of about 15 power plants. It is estimated that if this trend continues, an additional 10 power plants will be required by 2011 to support data center operations.(2)

Many IT organizations are focused on transforming from building core information infrastructures that support business operations to building highly optimized, efficient, flexible, and lightweight infrastructures.

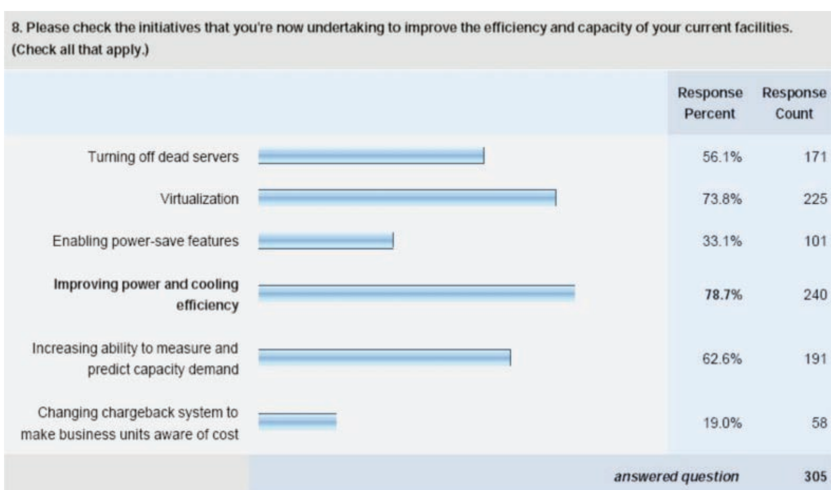


Figure 1: A study of over 300 IT professionals including senior and middle managers, technology analysts, and consultants conducted and released by the Uptime Institute in March, 2008.

The Environmental Impact

Data centers impact the environment in two ways:

• the direct generation of heat used to operate IT and cooling equipment

• the on-going process of power generation, supply, and consumption

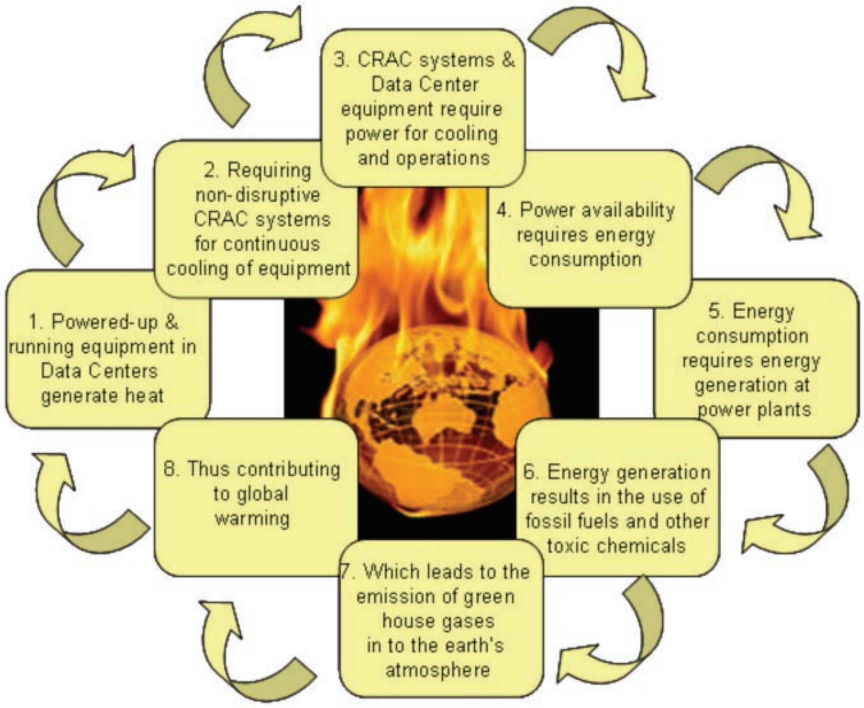


Figure 2: A conceptual diagram illustrating how the use of power in data centers impacts the environment.

As illustrated in figure 2, powered-up and operating data center equipment generates heat. Cooling is applied through large-scale (commercial use) computer room air conditioning (CRAC) systems to keep equipment operating within normal temperatures and prevent it from overheating. Both the IT equipment and CRAC systems require operating power and consume energy; the availability of energy requires power generation at power plants. Power plants use fossil fuels and other chemicals that result in the emission of greenhouse gases into the earth's atmosphere.

The Bottom Line

Technology has matured. Organizations can realize near and long-term financial benefits and demonstrate environmental responsibility by exercising diligence and prudence in technology evaluation, purchase, implementation, and operations and management. Our industry is now in a position to deliver IT with financial and environmental consciousness.

Considerations for an Efficient, Green Data Center

In this section, we will discuss some of the key considerations when transforming an existing data center or building new data centers. The goal is to feature efficiency and flexibility, make better use of existing and new infrastructure and resources, and provide headroom to keep up with on-going growth while minimizing the environmental burden.

Consideration 1: Data Center Equipment Placement

Today's high-performance equipment has densely packed electronics in a compact footprint. This requires greater heat dissipation and higher cooling requirements. Hence, it is important to pay close attention to the placement of densely populated racks in the data center. Misplacement of equipment results in higher power and cooling usage and higher energy costs. Organizations can take some simple steps to reduce cooling requirements and costs. These include creating hot aisles and cold aisles when configuring equipment placement. This step is discussed in greater detail later in the "Implementing an Efficient, Green Data Center" section.

Similarly, separating high-density equipment across the data center helps to prevent the build-up of major hot spots and maintains a relatively even temperature gradient. The data center can then be cooled with relatively higher room temperature requiring less cooling power. This step is also discussed in the "Implementing an Efficient, Green Data Center" section.

Consideration 2: Data Center Power and Cooling

Assess Power and Cooling requirements

When considering a green data center initiative, begin with an inventory of all IT assets to assess and understand current power usage patterns. For increased assessment control and accuracy, consider logical division of the data center into more than one section, where power and cooling readings can be made in each section. These logical divisions can be made along the lines of service to business units; an assessment can be performed for a complete infrastructure serving a given business unit.

Similarly, a data center asset division can be made by asset type such as servers, networks, or storage resources. The goal of an assessment is to identify inefficiencies in existing power and cooling patterns, and areas of opportunity for the greatest impact. An assessment is only as good as the tools used to perform it; hence it is important to use tools that provide granular power and cooling information. As an example, a tool that is designed to use vendor equipment specifications as input will likely lead to inaccuracies. It may not be sufficient to capture the total potential power demand of a rack of blade servers, and the total potential heat dissipation of the rack in a situation where the actual count of servers in the rack may be different. Using vendor power and cooling specifications for a rack of storage alone in an assessment will likely lead to inaccuracies. Storage usage within the rack may require less than the full rack.

It is also important to project (within a reasonable margin of error) the short and relatively long-term power and cooling requirements. Anticipate growth and manage the risk of running out of data center power. Assessment tools that perform trend analyses are useful to properly plan for current and future power and cooling requirements.

An energy assessment should also clearly identify the energy and cost-savings over varying lengths of time.

Design Optimal Cooling Plans

The key objective when designing an optimal data center cooling system is to create a clear path of air flow from the source of cooled air to the intake of equipment, and from the hot air exhaust of the equipment to the exhaust system's return air duct. While there are a number of things to consider when designing an optimal cooling system, a large majority revolve around the optimal operation of the computer room air conditioning (CRAC) system. Consider:

Ongoing Health-Check Inspection and Cooling System Maintenance

A regular health-check inspection ensures that the cooling demands in the data center are sufficiently met through the CRAC system's full potential output utilization. The health check inspection and maintenance must include the entire CRAC system including chillers and condensers, pumping systems, cooling loops, and direct expansion (DX) systems. Sub-floor conditions, in the case of raised-floor cooling, need to be inspected to ensure that there is no obstruction to air circulation due to improper equipment cabling and wiring. Dirt and dust in the sub-floor area will blow up into the equipment; dirty and blocked coils and air filters also impact cooling performance.

CRAC Positioning with IT Equipment

Data center CRAC units must align with equipment placement. Once equipment aisle (hot aisle/cold aisle) configurations are determined, plan the placement of CRAC units inline with and on both sides of hot aisles. We have greater concentrations of heat in these aisles by design. Also, ensure that blank panels are installed in all racks for any unused and blank space. Missing blank panels create openings in the rack and allow the hot air from the back end of the equipment (from hot aisles) to enter the cool air intake system of the equipment (in the cold aisles) and adversely impact cool air circulation.

Consideration 3: IT Infrastructure Virtualization

Virtualization is the single most important consideration in making a data center efficient and green. It addresses the most chronic inefficiency—low equipment utilization rates and assets that consume power and generate heat. Virtualization enables infrastructure consolidation, provides flexible resource provisioning, and yields tremendous resource utilization increases. This allows IT to meet business requirements with fewer resources—good for the business and the environment. The ability to do more with less equipment means less power consumption, which in turn means reduction in the data center's operating cost. Less power consumption also means lower heat dissipation. This in turn means a lower burden on power grids and power plants, which reduces greenhouse gas emissions and the impact on the earth's ecological system.

Consider these issues when implementing an IT infrastructure virtualization:

Consideration 1: Start with Analysis

When virtualizing an infrastructure, perform a detailed technical, operational, and business analysis at the start of the initiative. Create a map of business users and their applications along with the usage pattern by departmental or business unit. This will define the functional and operational requirements of virtualization solutions.

Focus your analysis on discovery of performance, availability, and growth requirements under normal and special conditions. A special condition for an application can be a particular time of the day, week, month, or year that requires increased availability and/or performance. Once this information has been gathered, map it against the available IT infrastructure to identify how best to meet these requirements with available resources. This mapping can be used later to build a charge-back model.

	Tier Storage Considerations						
Criteria	Symmetrix (Tier 1)	CX-FC (Tier1/2)	CX-ATA (Tier2/3)	NS I/G/X (Tier 1)	Centera (Tier 2/3/4)	EDL (Tier 3/4/5)	Tape (Tier 4/5/6)
Application Performance (Block Level)	X	X					
Application Performance (File Level)				X			
Application Availablity	X	X	X	X			
RAID Protection	X	X	X	X	X	X	
HA Capabilities	X	X	X	X	X	X	
Security	X	X	X	X	X	X	
R.P.O.	X	X	X	X			
R.T.O.		X	X			X	X
Disk to DiskBackup			X			X	
De-duplicated Disk-based Backup						X	
Disk to Disk to Tape			X			X	X
Tape Consolidation with Emulation						X	
Regulatory Compliance					X	X	X
Infrequent Data Access Management			X		X		
Inactive Data Management			X		X	X	X
Data Vaulting					X	X	X
Replication Capabilities	X	X	X	X	X	X	
Disaster Recovery Requirements	X	X	X	X	X	X	X

Figure 3: Above, a sample of mapping between business criteria and the use of tiered storage to align storage use with business requirement. Similar mappings for other components of infrastructure such as servers and networks can assist in building an efficient and well-leveraged virtualized infrastructure.

Consideration 2: Plan for Future Requirements

Include future growth requirements in your virtualization plan to build an infrastructure with sustainable efficiency. An application that becomes increasingly important in the day-to-day operations of a company may require additional server resources and/or higher storage performance over time. Investment in server and storage virtualization technology that features flexibility to scale and seamlessly adapt to evolving business requirements without disruption achieves the greatest efficiencies. The as-needed scaling of the existing infrastructure ensures that the ratios of efficiency for power, cooling, and utilization derived in the initial deployment can be retained in the expanding infrastructure and that growth occurs in a controlled and accountable manner.

Consideration 3: Start with the Easiest Resources to Virtualize and Consolidate

Servers and storage are the IT assets with the highest power consumption and lowest utilization rates. They consume approximately 50 percent of the overall data center power; however, the average server utilization is below 20 percent and the average direct-attached storage utilization remains below 40 percent. Virtualization and consolidation of these resources can have a dramatic impact in the cost of energy consumption, data center real estate, procurement of additional hardware, and management of infrastructure. The spare energy in the new efficient virtualized infrastructure can be conserved and used for future growth without new (additional) power requirements.

Power and Cooling

With the proposed VMware solution, servers can be consolidated helping to significantly reduce the power and cooling requirements and costs.

Power and Cooling Savings	Year 1	Year 2	Year 3
Current (As Is)	$ 291,435	$ 320,579	$ 352,637
With VMware (Projected)	$ 19,705	$ 21,676	$ 23,844
Total savings	$ 271,730	$ 298,903	$ 328,793

Figure 4: Above, a three-year break down of financial savings in power and cooling of servers by consolidating on VMware server virtualization infrastructure. Compare these savings against continuation of operations without consolidation.

Consideration 4: Use Energy Efficient/Green Equipment in the Data Center

Consider energy-efficient hardware when purchasing new equipment. All major server and some storage vendors are manufacturing energy-efficient hardware. These systems feature low-voltage CPUs designed to deliver more performance per watt of power, adaptive cooling fans that automatically throttle fan speed based on workloads, and high-efficiency power supplies that draw significantly less energy and reduce internal equipment temperatures. High-capacity, low-power disk drives with medium to high-performance disk drives in tiered storage subsystems, and disk drive spin-down features reduce power and cooling requirements and increase overall efficiency.

Consideration 5: Optimize Infrastructure through Information Lifecycle Management

Leverage tiered storage to perform ongoing information lifecycle management. Classify information and identify its value to the business. Store information in an appropriate type and class of storage device within a storage array. Storage subsystems consume a significant footprint and require data center power. Mixing different types and classes (high, medium, and lower performance) of disk drives in a single storage frame delivers efficiency in data center floor space consumption and power and cooling, as well as efficient management of the storage infrastructure. Tiered storage must not be limited to a given storage array, but rather must be applied across the broader horizontal storage infrastructure to realize storage efficiency gains in the data center.

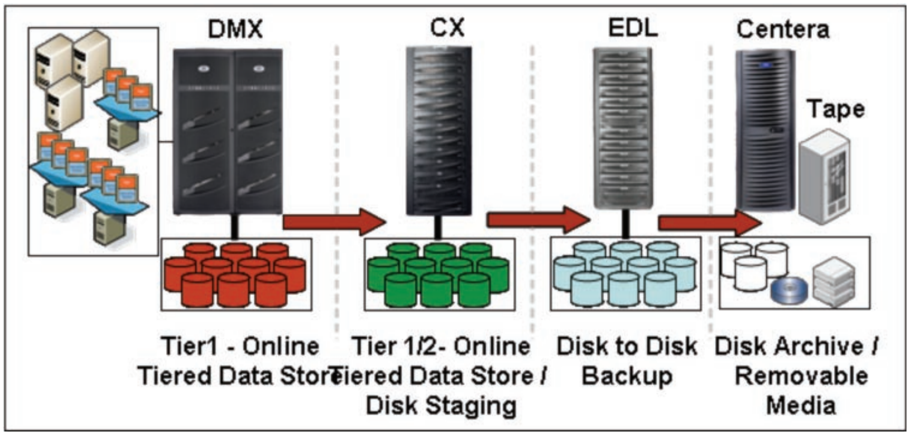


Figure 5: Above, a horizontal tiered storage deployment using storage devices in each tier with built-in tiered storage capabilities for vertical storage tiering inside of each storage device. This model yields maximum, inside-the-box (vertical) and across the storage infrastructure (horizontal), efficiency for power, cooling, and management.

Consideration 6: Optimize Information Archiving and Backup Data Management

Organizations can achieve operational, financial, and environmental efficiencies by storing less-frequently-accessed information, referred to as static or reference information, on lower powered, higher capacity, low-cost storage.

Reference information can be archived on online disks to improve backup and recovery operations across the data center. Since all reference information can be safely protected and made available through the online archive, organizations can eliminate 60 percent or more information from their active backup cycles. This reduction improves backup and recovery windows, requires fewer physical resources, consumes less floor space, and results in reduced power and cooling requirements for the backup and recovery infrastructure.

Data de-duplication can improve the infrastructure's backup and recovery efficiency. This technique is particularly useful for medium to large-size organizations with multiple operation centers and branch offices. Data de-duplication ensures that only unique elements of information are protected and that there is no redundancy in the backup cycle. If a set of information has already been protected in a backup cycle, and only a small portion of this information changes after the cycle's completion, only the new information elements will be protected.

Data de-duplication can significantly reduce the total amount of data requiring backup and drastically improve overall infrastructure efficiency. There are a few different implementation techniques; such as de-duplication of information at the source (de-duplicating information at the location where the information is being picked up from) or de-duplication at the destination (de-duplication of information before placing it in the backup repository). We must understand the business requirements to make the correct implementation choice. For example, is backup required for single or multiple sites? Does each site perform its own backup or is there a centralized backup model in place? Do remote offices require backup over the wide area network? If so, how many remote offices and what type of WAN bandwidth is available? Answering these questions will help to make a technology investment decision that delivers maximum efficiency.

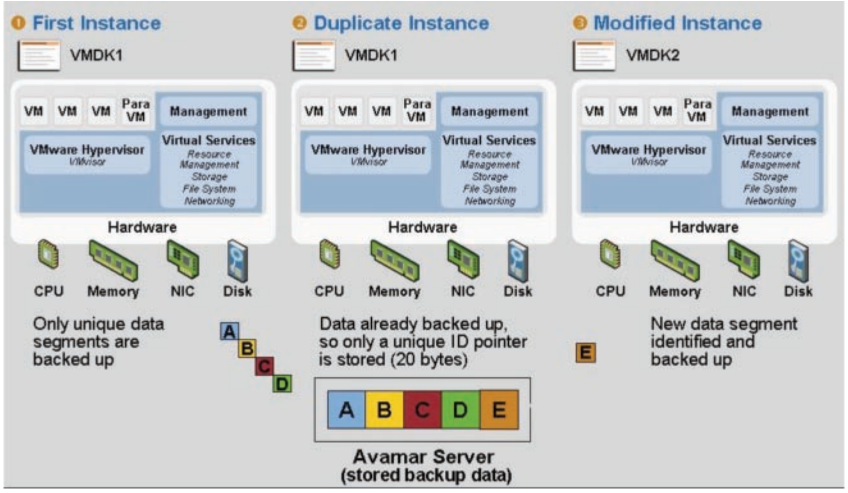


Figure 6: Above, shows how total backup information is reduced through data de-duplication by identifying and protecting only the unique information. This improves resource consumption and achieves daily full backups.

Consideration 7: Secure Virtual Infrastructure Management

Organizations routinely secure perimeters of the data center and control user access with various hardware and software solutions such as firewalls, filters, anti-virus software, VPN, and authentication systems. It is also important to monitor direct access to the management of data center resources such as servers, networks, storage, applications, and databases. Data center operation teams need only role-based access. Access must be designed and restricted/controlled so that no individual has enough access to single-handedly impact overall operations.

Create isolated management VLANs to logically separate the user network (LAN) from the management network. This not only ensures greater security when attempting to access IT assets, but also ensures availability of deterministic network resources (bandwidth) to IT staff for management functions and to users for application functions.

Consideration 8: Set Appropriate Expectations with Your Customers—the Users of Technology

Setting realistic expectations is critical to build an efficient data center. Not every user application or environment is critical to a company's day-to-day business operations. Over-provisioning resources or provisioning expensive resources to operate less-demanding lines of business is inefficient. Implementing charge-back systems, where departments are required to pay for services used, can help to prioritize the use of infrastructure and provide service delivery quality control data.

Service Level: Platinum	Service Number: 10060781
Subscriber Organization: OTD -- Online Trading Department	Cost Center: 20070
Subscription Type: Monthly	Start Date: 8/28/2008
Service Description:	Service Charge
Primary Trading DB Storage -- Grade A+ -- EMC Symmterix DMX	
Tier 0 -- Flash/Solid State Drive Storage	$10/GB
Tier 1 -- Fibre Channel (15K RPM) Storage	$5/GB
Priamry ORACLE Stoage -- Grade A -- EMC Clariion CX4	
Tier 1 -- Fibre Channel (15K RPM) Storage	$5/GB
Tier 2 -- Fibre Channel (10K RPM) Storage	$4/GB
Tier 4 -- Low Power Serial-ATA (5.4K RPM) Storage	$2.50/GB
Microsoft Exchange Operations	
200 Accounts	
Grade A Storage (CX4-240)	$100/Account
Backup to Disk -- EMC Networker/EDL	
Standard Data Protection	$2/GB
- Recoverfy Point Objective (R.P.O)	< 4 Hrs.
- Recovery Time Objective (R.T.O)	< 8 Hrs.
Continuous Data Protection	$4/GB
- Recovery Point Objective	< 2 Minutes
- Recovery Time Objective (R.T.O)	< 1 Hr.
Active Archive using EMC Centera	$1.50/GB
Replication - Local	
For Trading DB Multi-application Processing using DMX TimeFinder Mirror	$4/GB
For ORACLE Multi-application Processing CX4 SnapView Clone	$3/GB
Replication - Remote	
SRDF/ S	$10/GB
MirrorView/A	$7/GB
RecoverPoint - CRR	$7/GB

Figure 7: An example of an IT service delivery charge rate report to an organization with varied IT services.

Consideration 9: Continuously Strive to Optimize IT Operations

An ongoing review of IT requirements and services with functional departments and business units will ensure continuous alignment with business requirements. A quarterly review of delivered services over the previous term and discussion of requirements for the future will build confidence and provide an opportunity to educate other organizations about how IT requirements are delivered. Simultaneously, IT teams will better understand how and why requirements are generated. Engagement with business owners can fine-tune IT. Collaboratively, the business and IT stakeholders can identify areas where higher service levels are required, or areas where the existing quality of service can be dialed down to increased IT efficiency.

Consideration 10: Make Assessments a Routine Activity

Perform asset utilization assessments on a regular basis to ensure that all assets are appropriately and fully utilized. Server assessments spanning a period of time provide practical and actionable information on average, peak, and low utilization. This helps to determine where adjustments can be made. Similarly, a file system assessment for an application can identify certain patterns of usage by primary storage such as the type of data, number of duplicate copies of files, when a given file was created, modified, and last accessed.

Implementing an Efficient, Green Data Center

In this section, we will discuss designs and architectures for the most prominent aspects of a functional data center. We will discuss data center power and cooling and the use of virtualization and other data center technologies that form and operate an efficient, green data center. This will yield financial and operational benefits for the business and environmental benefits for the Earth.

We will first address those aspects of a data center that apply across the entire operation, namely power and cooling requirements. From there, we will work our way into the granular functional elements of a classic operational data center. Our definition of a classic data center is one that hosts a complete infrastructure to deliver day-to-day services to businesses of any size including small, medium, and large enterprises. Our classic data center is one that has requirements of efficiency, availability, performance, scalability, manageability, and business continuity.

Data Center Environmental Optimization

Power and Cooling Assessment and Planning

Data centers require power and cooling. Consider the total power available to the data center when designing it. Measure the total power available from the power grid to the data center. It is important to note that the total power available is not a measurement of the actual power available to operate IT equipment. A significant portion of data center power is consumed in the operation of secondary support equipment such as uninterrupted power supplies, power distribution units, lighting, cooling, and so on. Hence, calculate power utilization in operating these secondary support functions and equipment. Once this data is available, we can calculate the power available for IT equipment consumption.

There are a number of metrics to calculate power efficiency. One method, recommended by the Environmental Protection Agency, is called Data Center Infrastructure Efficiency (DCiE). DCiE calculates total power available to operate IT load and is represented in percentages.

The formula for calculating DCiE is:

DCiE = Power to IT Load / Total Power to Data Center

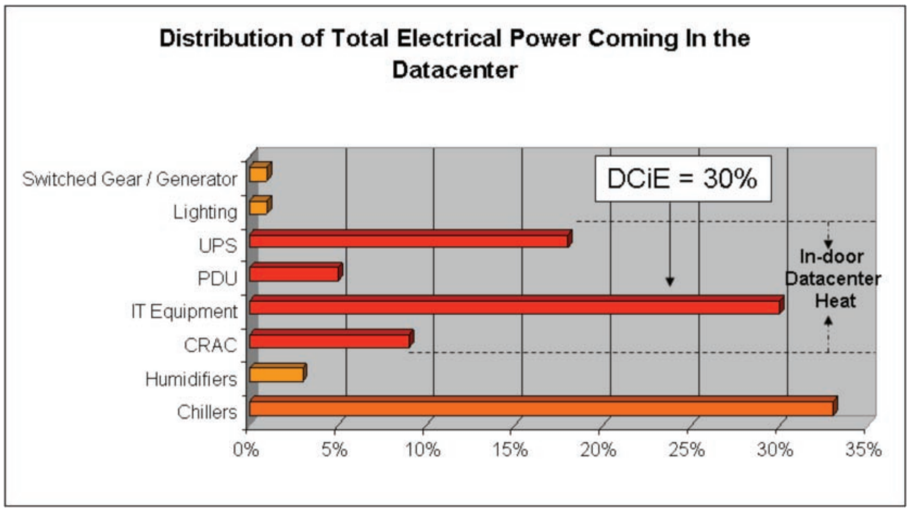


Figure 8: An example of power consumption in a data center operation. Using the DCiE metric for this data center will conclude that the data center is about 30 percent efficient. Using appropriate tools and processes, we can improve the efficiency in the other 70 percent of power consumption and can better use the data center's 30 percent efficiency.

Once data reflecting total power for IT equipment operations is available, the next step is to calculate the equipment's current power utilization. The power consumption of any hardware such as enterprise or blade servers, or network and storage devices, may vary significantly from the vendor specifications based on the deployed configuration. Thus, relying solely on a hardware manufacturer's power specifications will not yield accurate results. As an example, to accurately measure a storage array's power consumption, it is critical to measure component-level power consumption for the deployed configuration. In this example, the measurement for the array must be based on the number and type of storage processors; the speed, class, capacity, and total count of disk drives; count of battery backup units; and any other array hardware.

This cumulative study of power utilization across the data center will identify opportunities for power efficiency in the operation of secondary support and primary IT equipment. This information will also help to accurately project near and long-term power requirements to plan for and support business growth.

Equipment Management

Proper, proactive equipment management plays a vital role in decreasing cooling requirements and impacting the data center's overall power and cooling demands. The goal is to have a plan to generate efficient airflow circulation for the equipment.

Floor and Cable Management

First, consider the type of floor that will support the equipment. In a raised-floor data center with sub-floor cabling, a proper cable management system must be implemented in the sub-floor space as well as in the back of equipment rack to ensure that cool air intake and hot air exhaust flow is not obstructed by cables. In a non-raised floor with overhead cabling, maintain a clear path for hot air to exit the rear of the equipment rack and move along the ceiling to the hot air outlet without obstruction.

Rack Management

Manage open rack space to ensure proper airflow in the equipment aisles. Close any unused space in the racks with blank panels. Missing blank panels allow hot air exiting the rack to mix with the cool air coming into the rack, thus raising the overall level of cooling required.

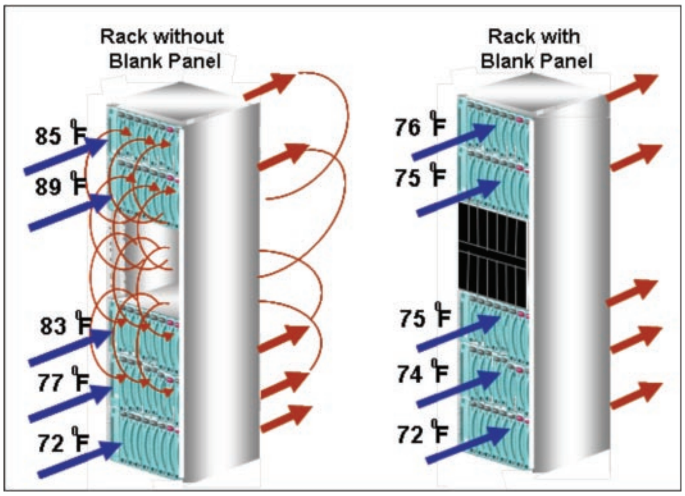


Figure 9: Above left, an illustration of inefficient air circulation due to missing blank panels in the rack requiring lower ambient temperature for cooling and increased cooling power. Above right, an illustration of a rack with blank panels installed leading to proper air circulation.

Create Hot Aisles/Cold Aisles

Create a hot and cold aisle configuration to support an efficient cooling architecture. Cooling requirements can be minimized to a measurable degree by eliminating the mix of hot and cold air in the same aisle. Place equipment so that the front of the equipment on the opposite sides of an aisle face each other, while the back sides of the equipment face each other in the next aisle. This creates alternating cold air aisles and hot air aisles. Ideally, equipment placement architecture needs to be considered during the data center's design phase since CRAC system placement must be aligned with the aisle configuration to maximize cooling efficiency.

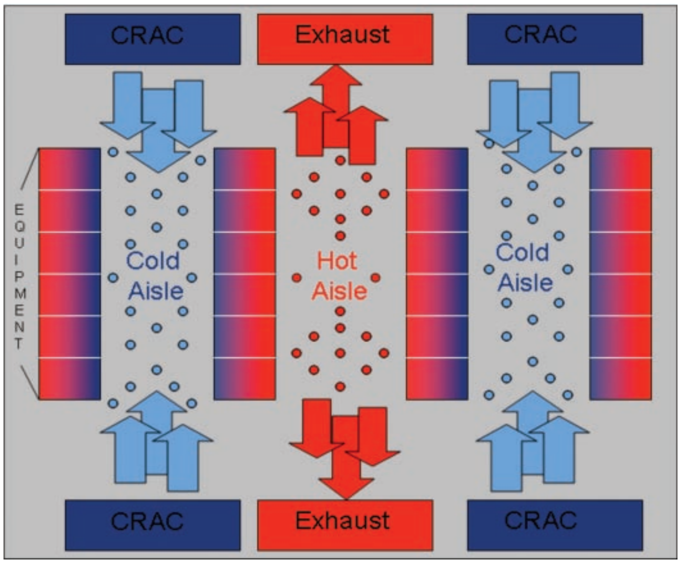


Figure 10: An overhead illustration of a properly planned data center with optimal equipment placement, air cooling, and hot air exhaust system.

Separation of High-Density Racks

The placement of high-density racks is another important design element of an efficient power and cooling architecture. Separating and distributing high-density racks across the data center avoids the build-up of hot spots and keeps the data center's relative temperature consistent. Consistent room temperature reduces overall cooling demand thus reducing cooling power usage and overall heat generation.

Once the environmental aspects of a data center have been appropriately addressed, the next phase is to build a physical IT infrastructure that will yield operational, financial, and environmental efficiencies.

Optimize Core IT Infrastructure

The core physical IT infrastructure includes servers, networks, and storage. In many existing data centers, addressing all elements of the core infrastructure at once is unmanageable. It is best to address one element at a time to better manage and ensure project success, while planning for other elements in the process. The objective of infrastructure optimization is to:

- Increase the utilization of core resources and reduce hardware footprint in the data center. This will yield financial savings in floor space, power and cooling, and operations and management costs.
- Achieve environmental efficiency by reducing the data center's carbon footprint.

Consolidation of servers and storage are the two most appropriate opportunities for physical infrastructure optimization. Servers and storage consume the most footprint and power with relatively low utilization rates among other resources. While there is no particular sequence to consolidate these resources, there is a major dependency on storage infrastructure in the case of server consolidation. We will discuss this dependency in the section "Virtualize & Consolidate the Back-end Infrastructure – The Storage." We will first discuss server consolidation followed by storage consolidation to logically organize the thought process from the top of the operational IT infrastructure stack to the bottom. We will discuss network infrastructure in the context of storage consolidation.

Perform Infrastructure and Resource Assessment

Current state infrastructure and resource assessment is required to understand and manage the scope of the project. This includes project goals and objectives, effort estimation and investment, and expected returns. The assessment will include information on existing server infrastructure, including resource utilization rates, deployment, operational, and maintenance costs including hardware, power and cooling, administrative, training, and services costs. The assessment must include information on future (virtualized) states as well. However, a phased adoption plan with classified workload is required to accurately project future efficiency targets for the consolidated environment.

Classify and Prioritize Applications

The first and most important step is to inventory all workloads/applications in a data center, then classify and prioritize them by function and criticality to the business. This will determine workloads that can be consolidated onto a single server and build out a complete virtual server. Then, requirements for physical hardware to host these applications as virtual servers can be derived based on the application performance and availability requirements. This will also help to identify workloads that are not good candidates for virtual servers, but must coexist alongside the virtual infrastructure.

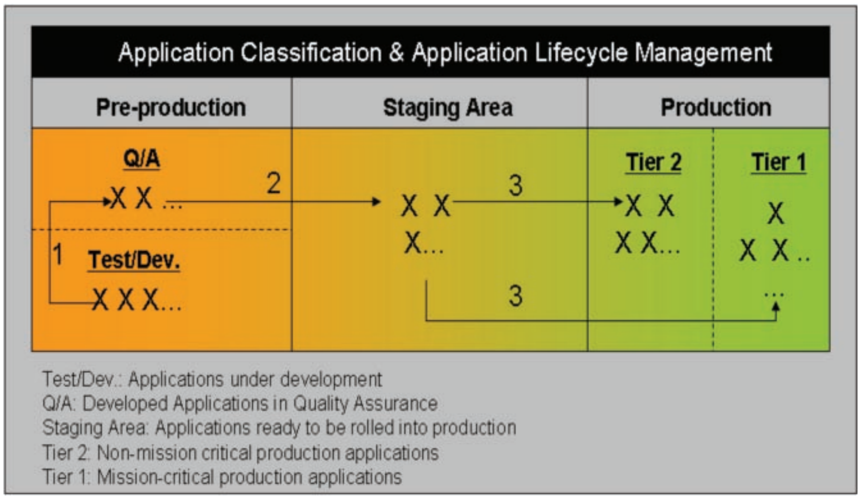


Figure 11: An example of 15 applications (represented by "X") in an enterprise classified by application lifecycle phases and business priorities.

Plan a Phased Approach to Infrastructure Virtualization and Consolidation

Physical infrastructure consolidation to virtual infrastructure must be completed in four planned phases:

- Phase 1 is an assessment. Then, especially for organizations new to server virtualization, there should be an evaluation or feasibility study that includes virtualizing a small number of test and development workloads to a few pre-production servers.
- Phase 2 deployment brings server virtualization mainstream into a pre-production environment. It deploys most test/dev and quality assurance workloads onto a virtual infrastructure. There may still be a requirement to test certain products under development on a physical infrastructure, so we should also account for these special requirements. However, physical infrastructure should become an exception after successful completion of phase 2. The primary focus of server virtualization evaluation and adoption in phase 1 and phase 2 needs to be hardware acquisition and data center footprint cost containment.
- Phase 3 introduces server virtualization for non-critical production workloads. The primary objective of phase 3 is to further build on cost containment achieved in phase 2 by streamlining core infrastructure management and taking preliminary steps toward business process automation. In phase 3, plan for broader efficiencies, including operations management automation and some level of business process automation.
- The focus and objective of Phase 4 deployment is to bring the virtual infrastructure mainstream in the production environment with the physical infrastructure for special-case requirements. Business process automation and optimization via end-to-end infrastructure mapping and management also need to be the centerpieces of the phase 4 virtualization initiative and planning.

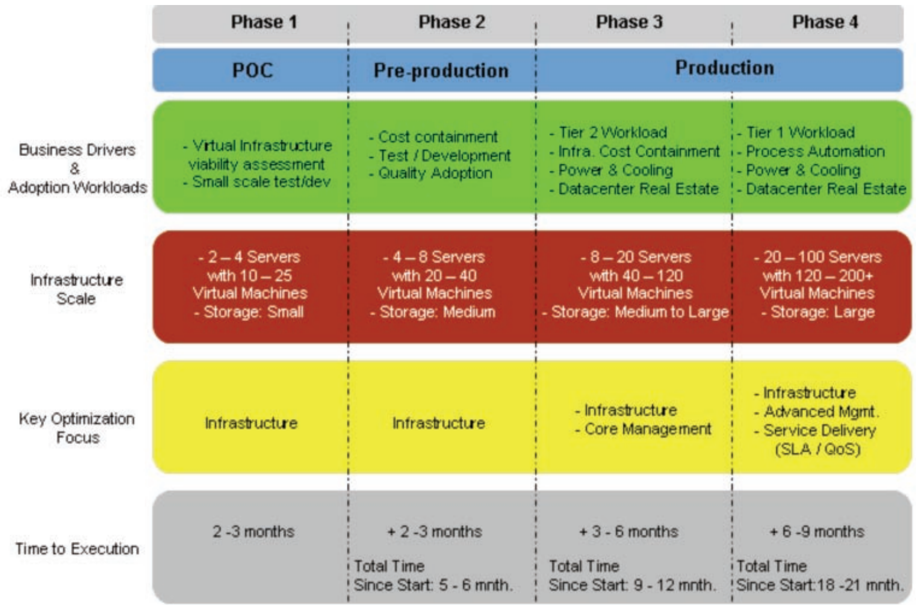


Figure 12: A phased approach to the adoption of virtualization illustrating key business drivers and workloads, infrastructure scale, key initiative focus, and timeline for each of the four adoption phases.

Virtualize & Consolidate the Front-end Infrastructure—Servers & Applications

The resource utilization of an actively accessed x86 server in production is estimated to be well below 20 percent. A typical server consumes between 30 to 40 percent of its maximum power in idle state. Servers also consume a large portion of data center floor space, power resources, and operational budgets. Server virtualization provides the fundamental platform upon which an entire, efficient data center with enterprise-class services for robustness, availability, and performance are built.

Server consolidation is accomplished by increasing the number of applications and workloads on a physical server. In a typical enterprise, applications run on heterogeneous server and operating system platforms. Since operating systems directly interact with server hardware, this limits the ability to effectively consolidate server and application infrastructure and reduce operational and management costs. Server virtualization breaks the barrier by introducing a layer of abstraction, called the Hypervisor, between the operating systems and the server hardware. This allows multiple instances of different operating systems to run simultaneously on any single x86 32-bit and 64-bit server.

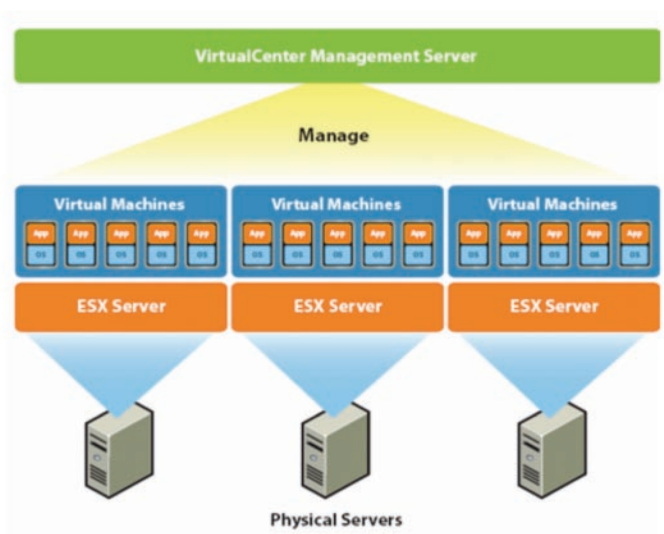


Figure 13: Three physical servers running VMware server virtualization software to create multiple virtual servers and host multiple heterogeneous operating system environments with each running heterogeneous applications with complete isolation and security.

This results in server utilization increasing from the typical 8 to 15 percent to 70 to 80 percent. Reducing the number of physical servers also reduces power and cooling costs and provides more computing power in less space. As a result, energy consumption can also decrease by as much as 70 to 80 percent. A consolidation ratio of 8-to-10 server workloads on a single physical server is typical; however, based on the processing capabilities of servers, some organizations have consolidated as many as 30 to 40 workloads onto a single server.

The foundational components required to build a virtual server infrastructure include:

- Virtual Infrastructure Operating System
- Virtual Infrastructure Management Server
- Virtual Infrastructure SDK
- Storage Infrastructure

Virtual Infrastructure Operating System: Virtual Infrastructure Operating System, referred to as Hypervisor, is a robust virtualization layer that runs on physical servers and abstracts processor, memory, storage, and networking resources into multiple virtual machines (virtual servers). The virtual operating system provides all essential functions for the virtual infrastructure. Some of these functions include:

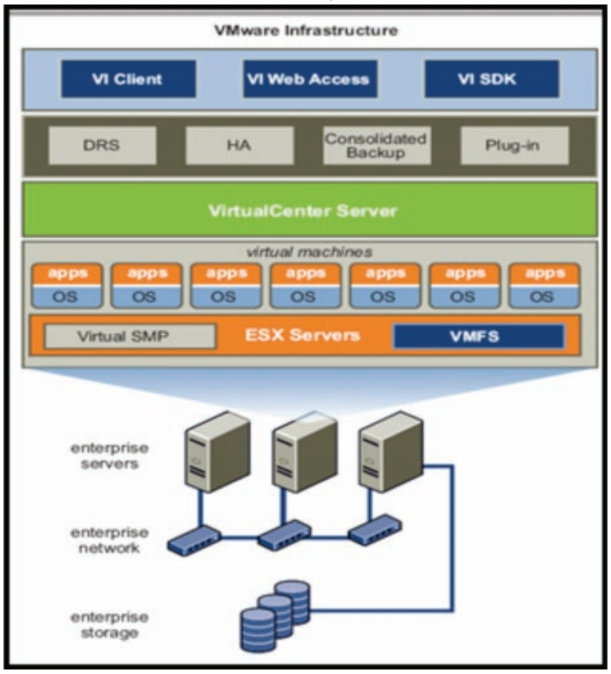
- Virtual SMP: Virtual Symmetric Multi-Processing enables a single virtual machine to use multiple physical processors simultaneously.
- Virtual Machine File System: An enterprise-class, high-performance cluster file system to host virtual machines.
- Virtual Networking: Provides all of the functions and capabilities of a physical switch inside the virtual operating systems of all virtual machines.
- Virtual Infrastructure Client: An interface that allows users to connect remotely to the Virtual Infrastructure Management Server or individual Virtual Infrastructure Servers from any Windows PC.

Virtual Infrastructure Management Server is the central point for configuring, provisioning, and managing virtualized IT environments. All essential and advanced functions such as online migration of running servers, online power and resource management, high availability, and disaster recovery functions are configured and managed through the VI Management Server.

Virtual Infrastructure SDK provides a third-party solution to integrate with the core virtual infrastructure.

Figure 14: A logical representation of VMware Virtual Infrastructure components.

Storage Infrastructure: Storage Infrastructure is required on the back-end to host (store) the entire



virtual infrastructure. Storage is a critical element. The realization of efficiencies in the virtual environment hinge on the correct and appropriate adoption of storage. Requirements and the role of storage infrastructure are covered in much greater detail in the following sections.

There are tools available to assist in consolidating and converting existing physical servers into virtual servers. They are wizard-driven and presented in a tutorial-like fashion. These server consolidation process automation tools use capacity-planning functions to discover physical systems in the environment, analyze them, and make recommendations for server consolidation based on workload requirements. Integrated conversion functionality in these tools transforms physical systems into virtual servers. Once applications have been consolidated to fewer physical servers, these and other service optimization tools help to migrate applications from one virtual server to another to maintain service quality and infrastructure efficiency.

It is also important to implement a process to transition applications from one stage of application lifecycle to another. There are tools available to assist IT administrators visualize and define application services, system configurations, and release processes to accelerate the completion of change requests to production. Some of the use cases for these tools include patch revision of existing production applications and introduction of new applications, enforcement of IT policies, and compliance with industry regulations.

Virtualize & Consolidate the Back-end Infrastructure—The Storage

As organizations work their way through the deployment phases of the virtual infrastructure (see figure 12), it becomes extremely important to pay close attention to the selection and purchase of appropriate storage. Adoption of an incorrect storage platform can very quickly lead to inefficiencies in the virtual infrastructure with penalties in the form of storage infrastructure acquisition, operations and management, data center footprint, power and cooling costs, and environmental costs.

Storage Considerations and Requirements for Phase 1 and Phase 2

Requirements for storage are relatively simple during the initial project feasibility, evaluation, and small-scale test/development deployment phase of an infrastructure consolidation project. These requirements include simplicity and ease of storage use, basic storage connectivity support for server virtualization technology, support for the iSCSI protocol for block storage, low cost of storage, and some level of maximum storage capacity. While functional storage requirements during the evaluation phase are simple, organizations must understand that requirements evolve very quickly as they move from initial to pre-production phase, and then from infrastructure virtualization into production. The use of multi-protocol unified storage is a fundamental requirement in building an efficient and consolidated storage infrastructure. During this evolution, functions such as point-in-time data copy, data cloning, support for NFS (NAS) protocol, and time-to-deployment become additional storage requirements. Investment in multi-protocol unified storage keeps the environment consolidated as storage requirements can be met and serviced out of the same storage infrastructure. Unified storage also contains and streamlines power and cooling, floor-space, and management requirements.

Storage Considerations and Requirements for Phase 3

As organizations move infrastructure virtualization into production, begin with tier 2 applications that are not mission-critical. Deploying tier 2 applications gives organizations an opportunity to feel their way into production with a virtual infrastructure and work out the operational and management issues before deploying business-critical applications.

Phase 3 introduces a new set of storage requirements. For instance, the level of availability and performance of a production IT infrastructure has a direct impact on the business. Hence, redundancy of storage and support for the Fibre Channel protocol, in addition to iSCSI and NFS, are important additional storage requirements. In phase 3, for tier 2 production applications, the use of low-cost Fibre Channel or Serial-ATA storage may be required to serve certain operational requirements such as data staging for backup or staging of applications on class-B Fibre Channel storage prior to promoting applications to a Class-A infrastructure. Hence, in addition to multi-protocol unified storage, tiered storage is another major requirement when building a compact, agile, and efficient storage infrastructure. Once again, investment in multi-protocol unified storage that features tiering functions results in infrastructure optimization.

Storage Considerations and Requirements for Phase 4

Once experience has been acquired in deploying some production applications on a virtual infrastructure, Tier 1 workloads such as sales, finance, marketing, and other mission-critical applications and databases that involve online transaction processing can be consolidated onto fewer physical servers. Since tier 1 applications carry the most financial impact to the business and even minimal unavailability or poor application response time may result in millions of dollars in lost revenue, there are stringent storage availability and performance requirements. These environments can benefit from both multi-protocol unified tiered storage and purpose-built tiered storage designed specifically to meet the highest performance requirements and efficient data management through information lifecycle management.

Organizations must deploy a storage platform that provides five-nines availability (99.999 percent) and deliver complete protection from any type of system outage to address phase 4 storage infrastructure availability requirements. Deployed equipment must be fully redundant and protected at all levels including everything from power distribution units and power rails to storage processing controllers and storage memory. Additional redundancy must be available at a data layout and disk drive level inside a storage enclosure. In case of an outage, the storage system must be able to fully protect in-flight transactions in the storage system memory. Similarly, the SAN or NAS infrastructure outside of the storage subsystem must be configured for full availability with multiple access paths through redundant SAN and Ethernet networks. Furthermore, multiple application I/O processing paths must be configured in the host to reroute client (system users) transactions in case of a path failure in the end-to-end infrastructure between the user interface of the application to the location of information on any given disk or disk set in the storage subsystem.

The same storage platform must leverage and support all existing and future industry-leading data processing protocols, such as 4 Gb and 8 Gb Fibre Channel and 10 Gb Ethernet protocols to address phase 4 storage infrastructure performance requirements. To keep the data center consolidated and efficient, these storage subsystems must support all classes of storage devices in the same unit ranging from enterprise-class solid state (Flash) disk drives to all speeds of FC drives to Serial-ATA drives.

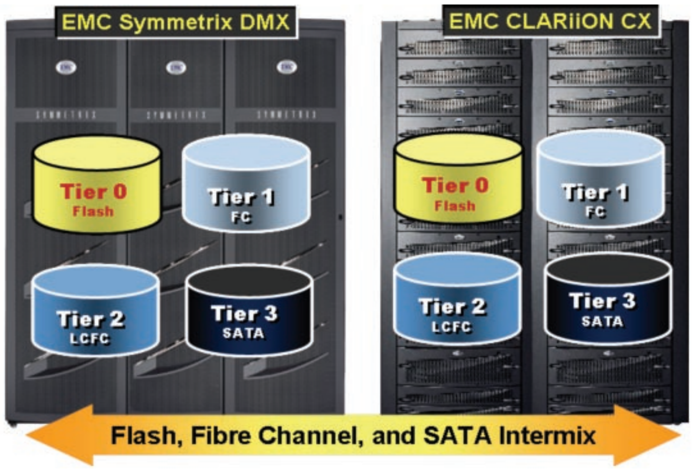


Figure 15: EMC Symmetrix® and CLARiiON® storage platforms with multi-protocol, tiered storage, quality-of-service capabilities to support availability, performance, and salability requirements for an organization of any size.

To optimize application performance, these storage subsystems must contain the flexibility and intelligence to dynamically fine-tune and right-size the use of system processing resources according to the application's performance requirements. For example, to consolidate three workloads with varying performance requirements onto a storage subsystem, the memory and storage processor resources must be assigned according to their performance requirements. To achieve this objective, storage subsystem intelligence is required to monitor and act against a set of established priorities for system memory, CPU, and I/O bandwidth utilization. Upon meeting the set priorities, either adjust and assign resource allocation to applications in a self-tuning automated fashion or provide the storage administrators with recommendations about the correct resource allocation to maintain the quality of storage service and meet overall infrastructure and application service levels.

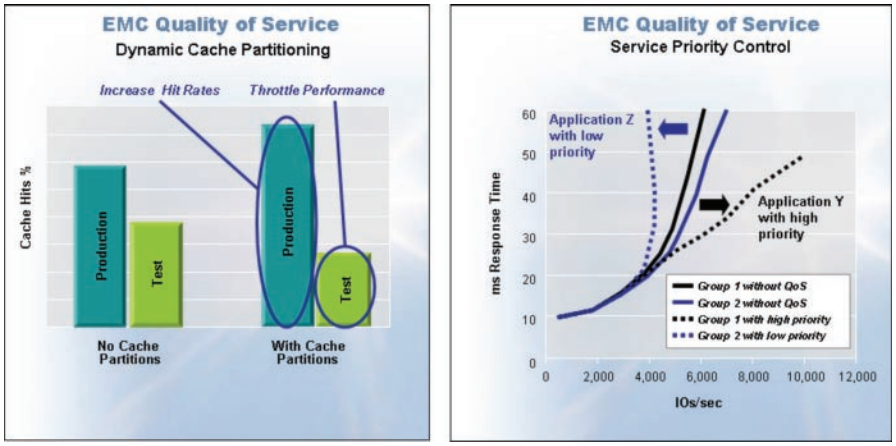


Figure 16: Examples of how EMC storage platforms optimize quality of service through intelligent optimization of storage cache and I/O bandwidth.

Similar to storage subsystem resource allocation, server resources must be monitored and right-sized using either an automated or a guided policy-based resource allocation system. Once the server infrastructure has been right-sized on the front end, application path management tools on the server optimize application response time by distributing the I/O load across all available paths. Address and optimize performance at all levels of the infrastructure including applications and databases, servers, networks, and storage to generate and deliver best possible application response time.

Secure the Information Repository in the Consolidated Infrastructure

Controlling Access through Service Credentials

Storage security must be considered along with the other storage requirements previously discussed. The relevance of infrastructure security to operational efficiency may not be readily visible; however, infrastructure security is necessary. An optimized and efficient infrastructure must not be vulnerable to external or internal security threats. Deployed storage must include strong data security functions to completely eliminate unauthorized access to the system and information. This requires storage security applied at multiple levels.

First, the subsystem must be safe; no unauthorized individual should be able to gain access to or control of the storage subsystem and perform service functions. This requires implementation of role-based access control, validation, and authentication though secure service credentials in the storage subsystem. For instance, a rule can be established that allows staff members to gain access to storage control functions from specific servers via complex service credential authentication while denying all other attempts to access. Similarly, a storage staff member may be allowed system access to generate reports, but may make no configuration changes; or a team member may add new storage, but make no changes to the existing configuration. Other rules may include staff members' access to the system during a certain time of the day or for a specific period of time, etc.

Implement Activity Log and Incident Management System

Implementing system access logging across the infrastructure enforces and enhances enterprise security. Maintaining access log information may also be required for compliance with various industry regulations. Log management systems incorporate logging of all vital elements of the infrastructure. They capture and store hundreds of thousands of events per second. To fully secure a storage infrastructure, incorporate storage access logs into an enterprise log and incident management system to perform security functions such as storage asset identification, access reports and alert generation, and incident management. Ensure compliance with the organization's security policies.

Encrypt Data

Once the storage subsystem perimeter has been secured and access activity can be fully tracked and reported, the next step is to secure the data residing inside storage subsystems. This is accomplished by encrypting the data in-flight and the data at-rest. Data in-flight is traveling over the wire between the points of origination and destination, while data at-rest is stored and retained on the storage device. Data encryption protects against electronic and physical security breaches such as unauthorized access to the system; and protects at lease rollover or storage upgrades that may result in physical movement of data outside the secure perimeters. Applying host-based encryption ensures that data remains completely secure before it leaves the point of origination, during its flight, and after it arrives and is stored in the storage repository.

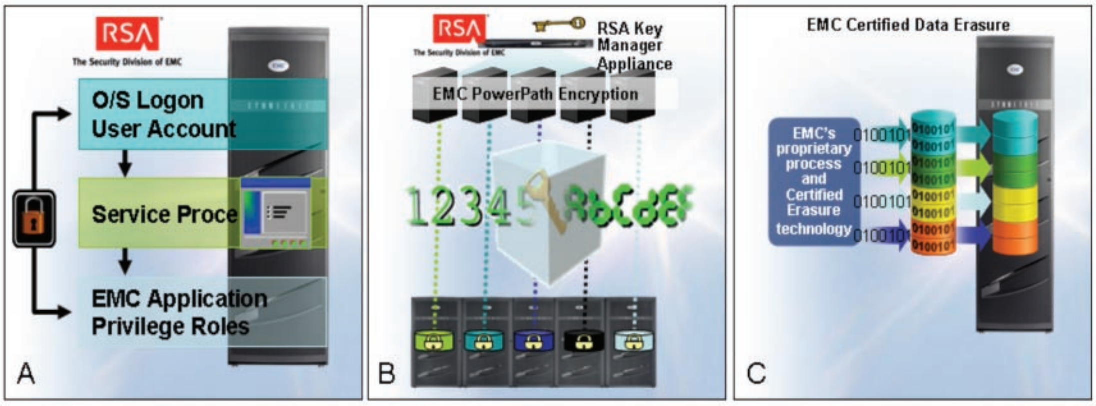


Figure 17 (A): EMC storage with secure role-based administrative access with full access logging. (B): EMC PowerPath® featuring security of data at rest with strong data encryption architecture. (C): Uncompromised protection of data through EMC Certified Data Erasure Service.

Use Data Erasure Services

Ensure that the security of data remains uncompromised during a lease roll-over, technology upgrade, or when performing bad disk replacements. Use certified data erasure service for open systems and mainframe storage. These services must adhere to stringent data erasure specifications such as DoD 5220.22-M and provide auditable records of data erasure.

Automate and Enhance IT Service Delivery

Adoption of advanced automated IT service orchestration and operations functions are essential to ensure the high quality of IT service and to enhance efficiencies derived from virtualizing and consolidating the core infrastructure.

Keep the Infrastructure Optimized with Minimal to No Disruption

A slow or poorly performing service/application means loss of potential revenue. An application that cannot fulfill user workload requests means that clients will turn to alternatives. The Internet is an example. If an online shopping service takes too long, users will turn to competitors. Notice that this is due to the poor quality of service and not a result of service unavailability. Hence, optimizing end-to-end system performance where the infrastructure can remain fine-tuned under heavy stress is extremely important. Virtualization addresses this level of performance optimization far more effectively than the physical environment. For virtual servers, solutions that can move workloads between virtual and physical servers, based on requirements, in a nondisruptive and automated manner are very useful. Using this technology, live, running applications with active in-flight transactions are moved from one server with insufficient resources to another server with necessary resources. IT staff has the option to either entirely automate this process or set alerts. If they opt to set alerts, they will be informed either prior to or upon a violation; they can then make an appropriate decision.

This technology also helps to eliminate maintenance windows as maintenance can be performed around the clock. For example, suppose we want to apply critical patches to a production application in the middle of the day. System administrators can seamlessly move live applications running on a maintenance target server to another server. Once the target server has been appropriately updated, the live/running applications can be moved back to the original server. This can all happen with complete transparency to users with no disruption.

Similarly, the quality of service issue may not be due to or limited to server resources. It can result from the storage subsystem's suboptimal configuration and resource utilization. Hence, it is important to use storage devices with built- in service quality management and optimization features. Some of these include support for multiple tiers of storage such as enterprise-class Flash disk drives, high-speed and low-cost Fibre Channel disk drives, and Serial ATA disk drives. Additionally, dynamic alignment of storage processor CPU, cache based on tiers, and characteristics of applications such as read or write patterns and workload fluctuations are required. Seamless, non-disruptive migration of application data from one class of storage to another, or one RAID configuration to another, is needed to ensure optimal business operations.

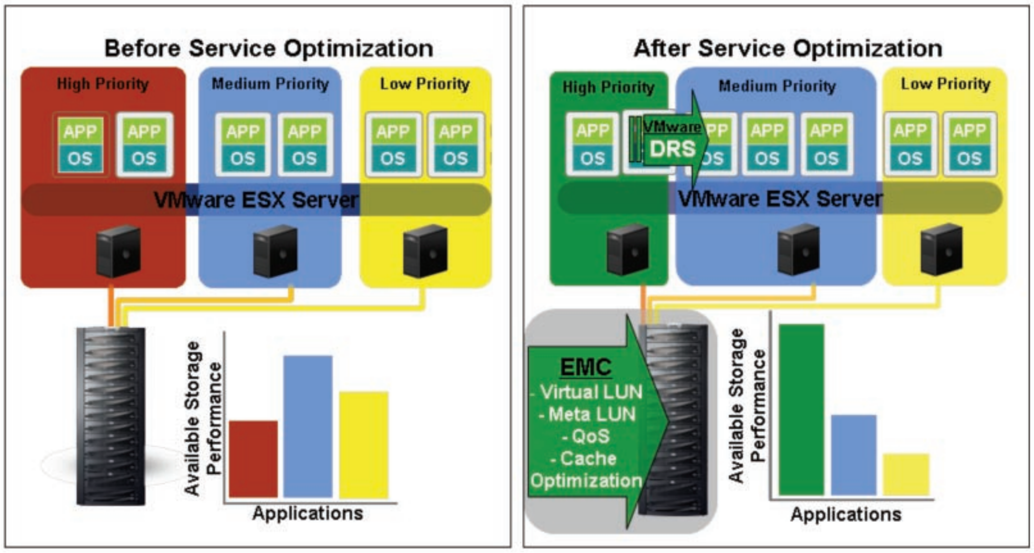


Figure 18: Above left, a logical illustration of non-optimized use of server and storage resources impacting the performance of applications of different classes. On the right, an illustration of the use of VMware DRS and various EMC storage resource optimization technologies which align the use of server and storage resources with application priorities and classes.

Add Seamless Data Mobility across Storage Subsystems

In addition to seamless optimization and data mobility within storage subsystems, it is equally important to address seamless data retention, movement, and migration across physical storage subsystems. A number of situations require this level of mobility. One example is an initiative to increase storage utilization across physical storage subsystems by making storage space from multiple physical storage arrays available to any given application. Others include upgrading storage subsystems to newer ones, expiring and rolling over leased storage from one vendor to a different vendor, and consolidating data centers that requires data migration between sites. Similarly, data mobility across heterogeneous storage may also be required to make a copy of production data available to support business functions. These may include data mining for decision-support systems, research and development functions, and data backup functions.

There is a tremendous opportunity to optimize the efficiency of block and file storage. For block storage, organizations can either deploy host-based data mobility solutions or leverage intelligent SAN fabric-based solutions that allow the use and movement of data across heterogeneous storage infrastructure for any or all application workloads.

Similarly, in NAS environments, there are solutions that leverage both in-band and out-of-band optimization technologies. This makes NAS storage efficient, simple, and easy to manage. It is important to ensure that these solutions do not introduce data vulnerability and do not operate in a manner that results in data loss due to equipment unavailability, especially for in-flight transactions that have not yet been committed to disk. Hence, these data mobility solutions must be stateless and act only as a pass-though service for requested functions. These devices must not store in-flight or committed data in their cache prior to committing data to disk, otherwise there will be an additional point of potential data loss.

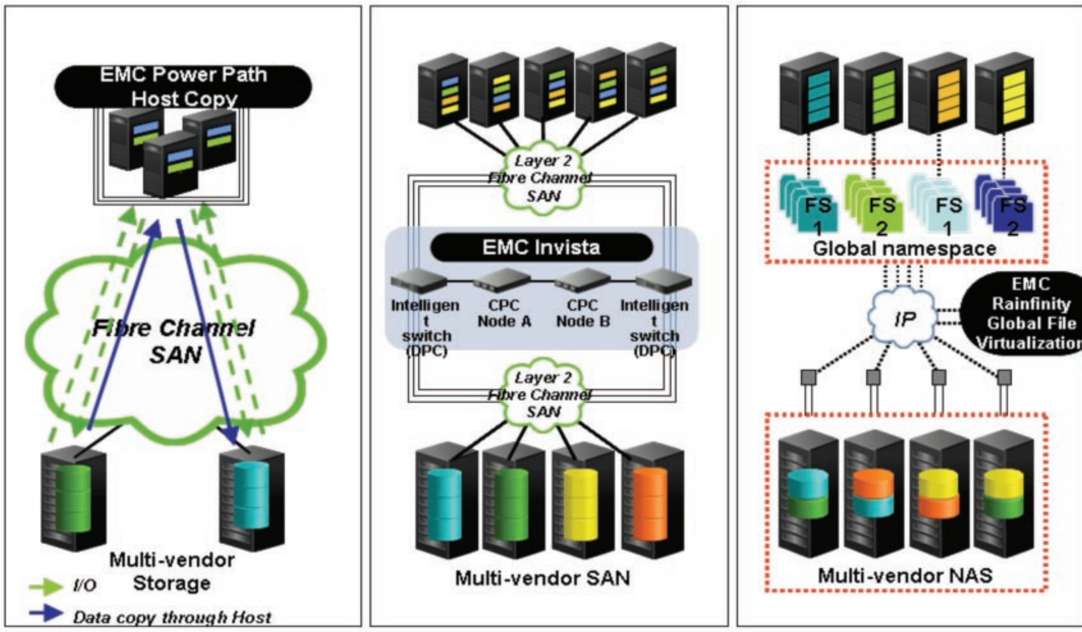


Figure 19: Examples of different options available for non-disruptive, seamless migration and management of data in SAN and NAS environments.

Apply High Availability for Continuous Business Operations

It is necessary to address infrastructure availability when optimizing the infrastructure for performance. Complete, end-to end high availability of the entire infrastructure is the only way to ensure ongoing business operations. Any outage from server to storage or anything in between means downtime resulting in lost revenue and productivity. Implement advanced, high-availability features from leading server virtualization vendors to ensure that virtual servers can be restarted automatically on an available server to protect a server from becoming unavailable in the virtualized environment. Clustering can be implemented at a virtual server level to ensure applications are started in the proper sequence with all required dependencies. This level of application protection further ensures that if only one virtual server becomes unavailable, the failing virtual server and related applications can be started on an alternative machine instead of failing over all virtual machines. There are tools available to orchestrate the entire process and workflow required for quick recovery. They further streamline availability in a virtual environment. High availability in a virtualized environment also means availability of shared storage to all nodes in a cluster. After virtual server failover, the server with workload ownership can acquire access to application data and keep the application running.

Business Continuity and Disaster Recovery for the Efficient Infrastructure

Business continuity and disaster recovery are related concepts that are often confused. While the net desired result for both is the same, availability of IT services to the business, business continuity means taking proactive measures to avoid an IT service outage. The outage could be due to any operational issue such as application or database problems, file system crashes due to storage space outage, equipment failures, or data corruption. In most cases, these issues cannot be categorized as disasters and do not involve outage of an entire data center site.

Disaster recovery means taking reactive measures to quickly restore service after a failure, mostly involving a site or partial site outage, due to natural or unnatural causes such as earthquakes, hurricanes, flooding, fire, and so on. This results in sizeable damage to the data center. Disaster recovery, typically, involves restarting IT services from a second data center. Technology is one major element in the implementation of disaster recovery scenario. Planning for disaster recovery involves people, process, and procedure with clearly defined and articulated roles, responsibilities, and steps. Implementing robust disaster recovery practices is critical since many aspects of business operations rely on technology.

In addition to the basic high-availability and clustering technique discussed in the previous section, there are a number of information protection techniques to provide business continuity and disaster recovery for data center protection. These techniques can be broadly defined as local protection and remote protection.

Local Protection

Local protection involves various degrees and methods of data backup and recovery. It protects from information loss scenarios such as user errors, application or database corruptions, and equipment failures. Building an efficient infrastructure means leveraging the right technology for every environment and scenario in a data center. There are a number of methods available to protect information; most environments use a mix of techniques. For non-critical applications, basic tape or disk-based backups might be sufficient, whereas for more important applications, online disk-based backups with various levels of recoverability can be implemented. The exact protection methodology will depend on the application's recovery-point and recovery-time requirements. When implementing a backup solution, it is important to consider varying requirements for different workloads and ensure that the implemented solution can handle all diverse local protection requirements.

Leverage Disk for Backup and Recovery

Backup disks have replaced traditional tape to a great degree in many organizations. However, tape continues to coexist with disk backup due to industry regulations and other requirements. There are a few reasons for the popularity of disk as a backup medium, including affordability, capacity, integration with backup applications and online snapshot software, online availability of information, quick data recovery, ease of use and management simplicity. Disk-based backup libraries, known as virtual tape libraries, emulate physical tape libraries making the replacement of tape libraries simple and easy. Disk libraries can be programmed to behave as tape libraries, thus requiring no changes in backup applications, scripts, or processes. The addition of data de-duplication functions in disk libraries further justifies the business case for the use of disk as the preferred backup and recovery medium.

Leverage Disk-based Snapshots for Online Backup of Data

While there are a number of ways to perform backup, a typical online backup of an important application or a database uses the backup application to initiate the job and create a of point-in-time copy of data. The backup application integrates with storage software to perform online snapshots of applications and databases. This data is typically stored on Serial-ATA disk for some period of time before it is moved to its final backup repository, which could either be another set of Serial-ATA drives in a disk array, a disk library, or a tape device. The intermediate repository, where data is held for some time, is called the disk staging area. Since most data recovery requests are made during the first few days or weeks after backing up a dataset, the purpose of disk staging is to retain recent data for quick recovery in case of a request for information. After a few weeks, data can be moved to an official backup repository and the staging area can be recycled.

Leverage De-duplication

Data de-duplication is one of the latest innovations in the backup space. It offers unparalleled levels of efficiency to streamline and minimize the total amount of unique data that requires protection. This technology revolutionizes how organizations approach backup and recovery. Data de-duplication, as the name suggests, identifies unique instances of data at a sub-file level, eliminating backup redundancy. Given that only a small portion of the overall dataset changes on a daily or a weekly basis, data de-duplication delivers tremendous efficiency to the environment. Efficiencies include shorter backup cycles, meeting backup windows, protecting more data in original backup window, and a major reduction in the total used storage capacity.

It is important to understand the environment when implementing data de-duplication technology. In a virtual infrastructure environment, there are multiple virtual servers running in a physical server. Each virtual server requires protection. In these environments, since there are finite available CPU, host bus adapter, and network resources that are shared among all virtual servers, the most efficient way to perform backup is to de-duplicate data at the source, prior to it leaving the physical server. Using this technique, finite server resources do not become bottlenecks and online production applications in these virtual servers do not suffer from lack of performance.

Provide Continuous Data Protection

Mission-critical applications with very aggressive recovery-point requirements such as financials, securities exchange, online auctions, and live media broadcast cannot afford to lose any data. Continuous data protection provides the ability to take ongoing snapshots of data. Every write committed to production storage is sent in parallel to a fully redundant CDP appliance that maintains a log and a full replica volume of production data. Data can be accessed from the CDP journal and replica volume to provide business continuity in the event of an outage of production storage.

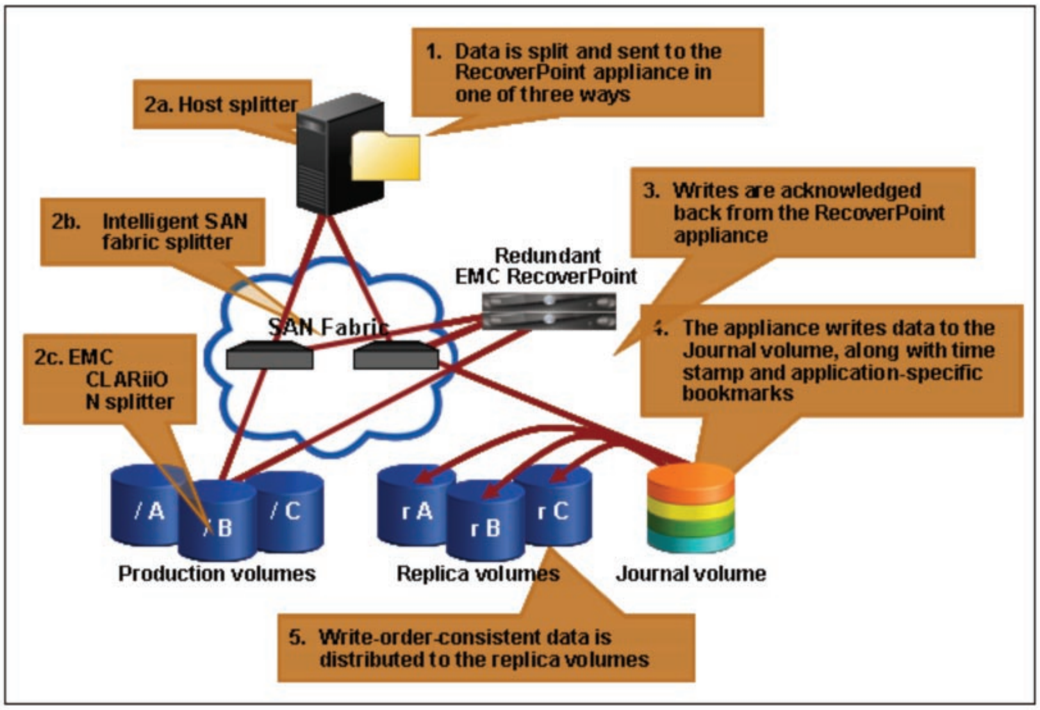


Figure 20: An illustration of how continuous data protection (CDP) for up-to-the-second local data protection is implemented in an EMC RecoverPoint solution.

Protect NAS environments

NAS is a popular architecture to store data for a variety of workloads such as engineering data for research and development, file sharing, storing templates, ISO images, and print services. Since NAS deployments leverage the existing IP network, the classic method for protecting NAS is to send backup data over the IP network. This is not the most feasible and efficient solution in many environments where network bandwidth comes at a premium and administrative use of a production network means taking resources away from production workloads. While backup traffic can be separated from production network traffic via creation of separate vLANS, carving out vLANs for backup jobs still means reducing the aggregate available bandwidth for production use. The Network Data Management Protocol (NDMP) is designed specifically for NAS data backups. NDMP is supported by and comes as part of major backup software utilities. It streams data directly from the NAS device to the backup unit that could either be a tape, disk library or disk storage unit. Using NDMP, only control frames from the backup server traverse over the IP network enabling LAN-less and server-less backups. The net effect is sustained infrastructure efficiency by optimizing the use of the IP network for production traffic.

Automate the backup process

A complete backup process involves a number of steps and is not limited to scheduling the backup job in the backup utility. Some of the process steps include creating disk-based snapshots, integrating with application frameworks and databases such as Microsoft VSS and Oracle RMAN to synchronize data and perform application-level data integrity checks, and handing off valid and verified snapshot data to the backup utility. The deployed backup software must be comprehensive and feature-rich to ensure the efficiency of backup operations. It must handle all different types of backups and must be able to manage the entire backup workflow and process. It must also automate all of the required backup process functions.

In addition to backup software, there are also tools to handle the entire disk-based data replication process at an application and database level with full integration with Microsoft, Oracle, SAP, and other application and database vendors' frameworks and utilities. These tools can be leveraged on an as-needed basis to optimize backup operations.

Stay Informed of Backup Environment and Manage through Canned and Custom Reporting
The ability to generate and receive on-going reports on the status of your backup environment is one of the most critical requirements for backup administrators. Reporting is the only way to manage backup environments in the majority of medium to large-size organizations. Awareness is the first step to take proactive or reactive corrective action. Using automatically generated reports, administrators can study backup reports in the morning to get a picture of the previous day's and/or night's backup activity. There are backup reporting tools that tightly integrate with backup software and provide business-level and detailed reports. Some of the reporting functions include status of backup jobs, inventory of resources, capacity utilization rates and trend analyses, bandwidth utilization, duration and length of backup jobs, and many other valuable reports.

Remote Protection
Remote protection involves replicating data between physical sites as close as the building next door or as far as another city, state, country, or continent. Businesses have various reasons to perform remote data replication. These include the need to service a geography or region from a secondary site or to quickly recovery from a disaster to provide business continuity. There are a number of ways to replicate data between sites. It can be replicated from within the physical or virtual servers or the networks, or the storage devices through local or wide area networks.

As with any other IT implementation, the goals and objectives of performing remote replication must be clearly defined and well understood to ensure efficient and successful deployment. Applications must be classified and prioritized based on business requirements, internal and intra-application dependencies must be identified, and recovery-point and recovery-time requirements must also be defined. This information is necessary to select and implement a solution that is efficient and meets business objectives.

Server-based Remote Replication Protection
Data can be replicated in a synchronous or asynchronous manner using server-based replication software. Synchronous replication ensures that data is best protected with the potential of data loss approaching zero. Every transaction is committed on both the primary and the secondary site in real time. While synchronous replication has the least exposure to data loss, this method requires sufficient network bandwidth for practical operations and has distance limitations. The distance limitation is dictated by the use of network technology and the application characteristics. Attempts to replicate application data over extended distances using insufficient network bandwidth may result in network congestion and application timeout leading to service unavailability.

Asynchronous replication offers the flexibility of replicating data over extended distances without limitation. However, since data is transmitted on a scheduled basis, there is a time lag in the states of data between sites so the degree of data loss can be minutes to hours depending on replication frequency. While server-based replication may be the most affordable solution from a cost-of-implementation perspective, this method has drawbacks. These include server resource utilization for replication functions, and limitations in scalability relative to network- or storage-based replication. Server-based replication may be a good remote protection solution for some small to medium-size organizations. For larger organizations, network- and storage-based remote replication offers the highest scalability and performance.

Network-based Remote Replication Protection
Network-based remote replication leverages an out-of-band appliance architecture that complements the existing IP network. The redundant network appliance offloads the replication function from the host or the storage and offers the flexibility to replicate data between heterogeneous storage devices on two sites. This method eliminates the need for special data protocol converters such as an FC-to-IP converter. The appliance performs important functions such as optimizing network bandwidth utilization through automatic network throttling and data compression with compression ratios as high as three to 15 times the original payload.

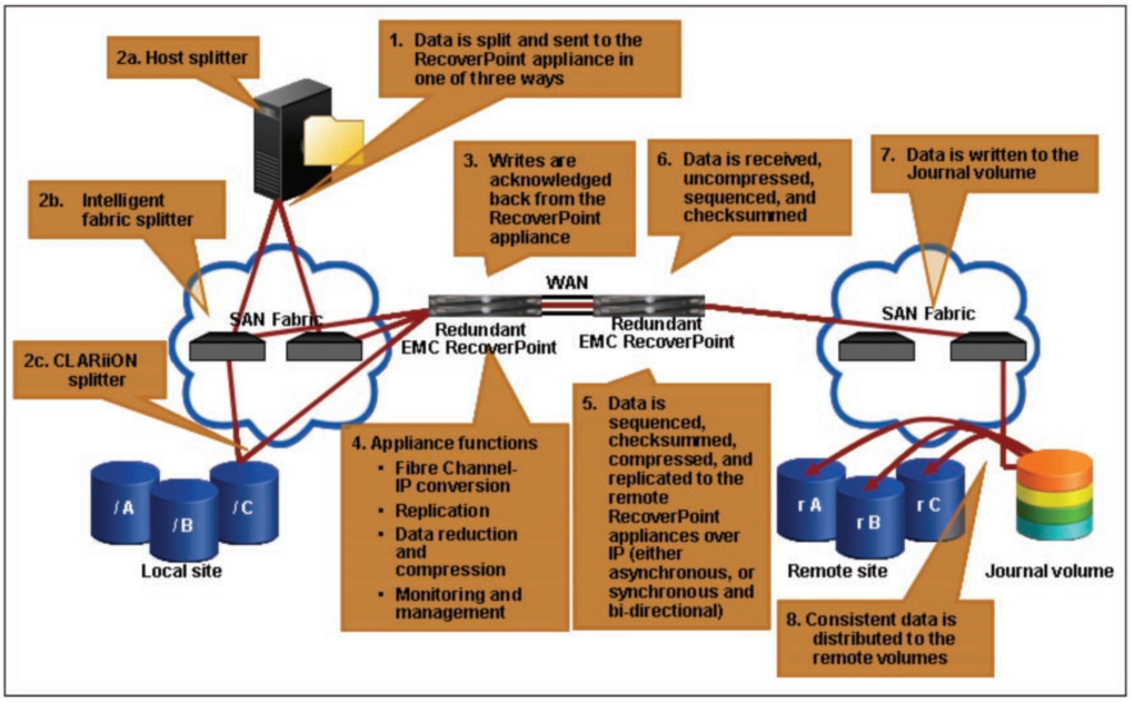


Figure 21: A functional illustration of the use of EMC continuous remote replication solution for business continuity and disaster recovery.

Network-based remote replication also features "consistency groups" in which multiple individual storage LUNs are logically treated as one unit. Replication policy is applied to all member LUNs in the consistency group simultaneously. This is especially important for applications and databases that operate across heterogeneous server and storage environments and have data dependency on multiple storage LUNs. Consistency groups keep all member storage components fully synchronized. In a disaster restart scenario, applications and databases can be brought back online with minimal downtime. In network-based replication, data can be transmitted in real time or on a scheduled basis.

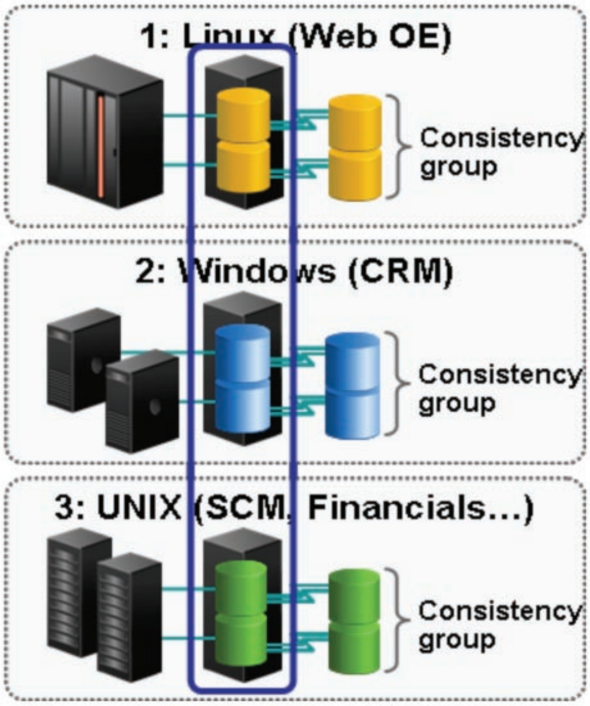


Figure 22: Illustration above shows the use of consistency groups in a heterogeneous/cross-platform environment to ensure data protection of a multi-application business process across a heterogeneous storage infrastructure.

Network-based replication is semi-synchronous. Once data is acknowledged by the remote appliance, acknowledgement is sent back to the primary appliance. This architecture has lower data loss exposure than the asynchronous architecture. While this type of replication has clear advantages over server-based replication, even network-based replication does not have the scale and performance of native storage-based replication. Network-based replication is a good solution for medium to some enterprise environments.

Storage-based Remote Replication Protection
Storage-based replication is the most robust, scalable, and effective way to perform remote data replication. Storage-based replication is ideal for large commercial and enterprise customers with aggressive remote replication performance, availability, and scalability requirements. In addition to advanced functions such as consistency groups and dynamic network bandwidth throttling, this architecture allows for many other enterprise-class advanced functions. These include multiple-site replication, simultaneous replication from one device to multiple devices at multiple remote sites and vice versa, and the ability to perform cascading replication between multiple sites using synchronous and asynchronous replication methods simultaneously. While replication at this level is typically performed between similar storage devices at both sites, there are solutions that allow special scenarios such as data migration to replicate data among different storage subsystem types.

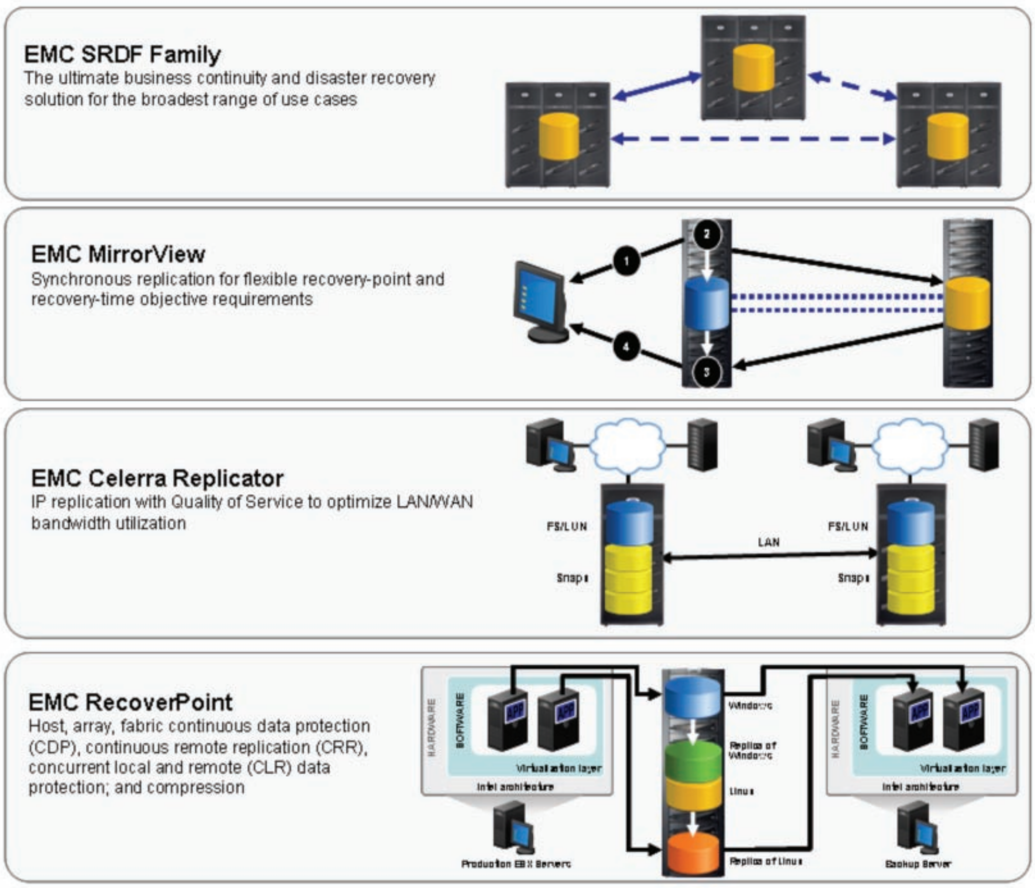


Figure 23: An illustration of a number of different storage-based SAN and NAS remote data protection solutions available to protect organizations and environments of all sizes and types.

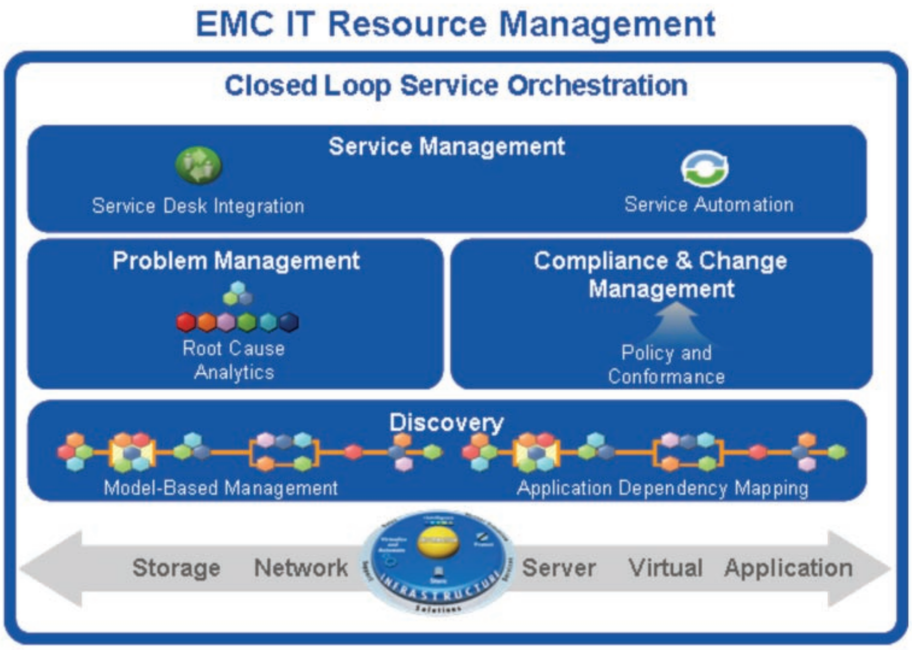
Automating Remote Replication and Disaster Restart Orchestration
Implementing and operating remote replication for disaster recovery is a complex mixture of art and science. It involves an in-depth understanding and orchestration of a number of independent functional elements that must converge in the proper sequence for the environment to work. Hence, once set up and operational, it is important to perform disaster restart drills on an ongoing basis to test the accuracy of disaster recovery plans. There are disaster recovery workflow process tools that define and configure a complete restart scenario. In the event of a disaster, IT equipment, applications, and services can be automatically restarted at the remote site based on the policies defined in the workflow process tool. In order for a tool of this capacity to be functional, it must integrate with applications as well as storage subsystems and their functions. In addition to disaster recovery

workflow orchestration tools, there are remote replication automation tools available for normal day-to-day remote replication functions. These tools integrate with user applications and databases such as Microsoft, Oracle, SAP and others and also have storage-level integration. Leveraging replication automation technology can greatly reduce the margin of error and improve recoverability success.

Resource Management in an Efficient, Green, Virtualized Data Center

Management of IT operations is estimated to be the largest data center cost at approximately 67 percent of the total operations budget. Building a virtualized data center to gain IT infrastructure and operations efficiencies introduces new requirements for resource management and demands increased automation of management functions. A closed-loop service orchestration model is required. Data for all infrastructure resources is collected in a federated manner and provided to a "data discovery" CMDB. Information from this CMDB is then used by other tools in this model to perform advanced management functions such as application dependency mapping, root-cause analysis, change management and compliance workflow orchestration, and service management.

Figure 24: A logical representation of IT resource management model that addresses end-to-end infra-



structure resource management and automation requirements and introduces new levels of efficiency and quality in IT service delivery.

A clear understanding and mapping of virtual and physical dependencies is required since physical and virtual resources are tightly coupled in a virtualized environment and have the flexibility to dynamically move around for service optimization. The next step is to overlay application dependencies on the virtual infrastructure, and finally map them to physical resources. This is particularly important for change management functions such as requests to bring new applications and services online, provision additional storage or servers, introduce new hardware or remove old equipment. Automated discovery and mapping of every infrastructure touch point between a physical resource and the user of a service is of significant value and importance in management operations. This level of automation enables important functions such as proactive analysis of change request impacts on the existing quality of service or on the segment, line of business, or group of users who have the potential of exposure and who may be impacted. Additionally, root-cause analyses of existing problems and identification of potential problems in the infrastructure, as well as compliance with IT governance and regulatory requirements, can be performed. Resource management automation also allows self-collection and analysis of infrastructure usage data. This improves the use of server and storage resources and quality of service to users. These tools provide details about an application's existing storage usage, or a database's future storage requirements.

Leveraging Infrastructure Virtualization Beyond the Core Data Center Infrastructure—The Desktop Virtualization

Once the core IT infrastructure has been optimized via virtualization of server and storage resources, the next step is to look for opportunities for efficiency in areas adjacent to the core application infrastructure. One such area is the user desktop. Desktops are the most widely distributed IT asset that live completely outside the confines of secure data centers, but operate off data center resources. Imaging, provisioning, and ongoing maintenance of desktops is a costly, time-consuming, and risk-prone operation. The cost of maintaining a desktop throughout its life is, in many cases, far greater than the initial hardware acquisition cost.

For example, an IT organization supports multiple desktop operating systems for employees and business partners who are provisioned desktop images and access to applications based on profile and job function. Keeping all combinations of desktop images and profiles updated and current with applications and security patches is a complex, time-consuming, and costly task. Even more expensive is the lapse in security when protecting an organization's confidential data. Stolen or lost laptops or unauthorized user access to company confidential data costs many organizations millions of dollars each year. Desktop virtualization takes care of these issues and drives major financial and operational efficiency in IT organizations.

Additionally, virtual desktop infrastructure improves energy efficiency by replacing underutilized PC desktop hardware with thin clients that consume far less energy and require less frequent replacement.

Storage is the key enabler to deliver an efficient virtual desktop infrastructure. The premise of a virtual desktop is that a handful of physical operating system images with appropriate applications are created and maintained to support different desktop user communities, including employees and partners. Using the snapshot function of storage devices, hundreds of writable point-in-time snapshot copies of these golden images are created. Each snapshot copy represents a virtual desktop. Desktop users access their desktop images by contacting the connection broker server, which validates and authenticates users based on their active directory profile.

Once successfully authenticated, users are granted access to their snapshot image of the desktop. Since these are snapshot images, they consume practically no storage space. Instead, they point to and read data from master golden images to boot up into a desktop environment. Once booted, users are provided with a private user profile area to save their work. This allows the master boot image of the desktop to remain pristine since it is not impacted by a user's personal work. In the event of an application or operating system upgrade or patching of a desktop environment, the golden image can be updated and hundreds of new zero-space desktop snapshots can be created and provisioned in a matter of minutes. Users can re-log in to the updated desktop environment. Various GPO and RDP policies can be activated to keep certain users or groups from accessing, downloading, or printing sensitive information. This protects information confidentiality and prevents access to sensitive information. Desktop image access can be granted to individual users with private desktop images (one image to one user) or to a group of users (one image to many users).

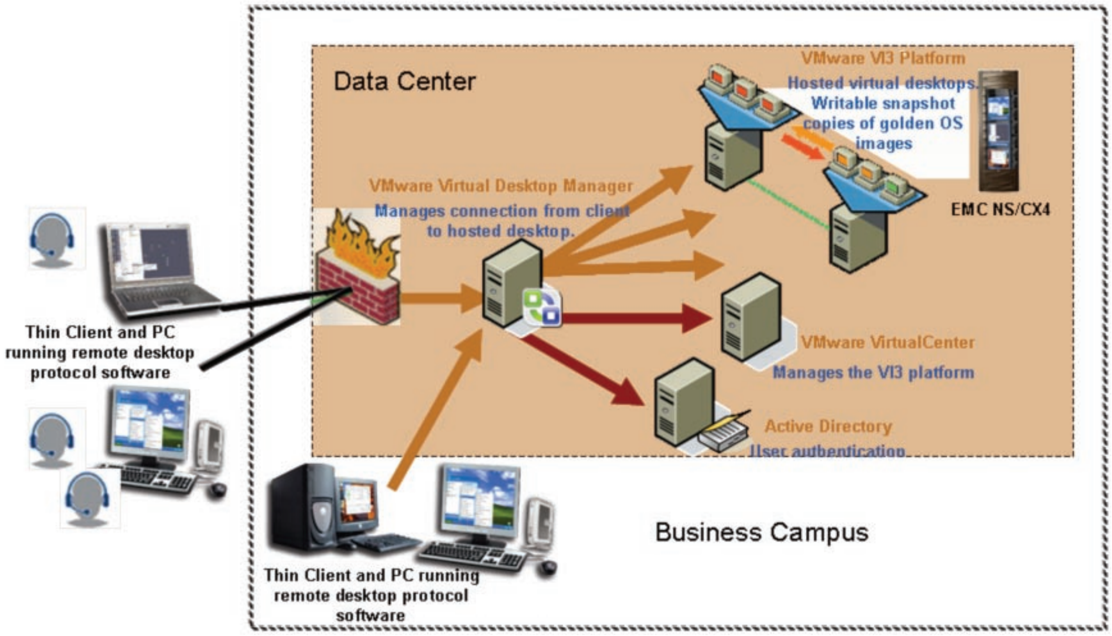


Figure 25: A conceptual architecture of virtual desktop infrastructure and its various components.

Storage Requirements for Virtual Desktop Infrastructure

Selection of the correct storage platform is critical for the implementation, scalability, and optimal use of a virtual desktop infrastructure. This architecture cannot exist without the storage subsystem's fundamental snapshot functionality. However, this is not the only dependency. There are a few distinct storage requirements including the type of storage used for hosting golden desktop images and user data. Storing golden desktop images requires the use of high-speed storage media and high-performance RAID configurations optimized for read operations. This is required because when users access their desktop image, they boot off of a handful of golden images. When a number of users attempt to log in to their desktop images at approximately the same time, the intensive read operations off the same diskset can bring desktop performance to a crawl. This can make the solution impractical and non-scalable. Alternately, allocation of separate storage is required for users to save their work. That can be well served by NAS file shares or iSCSI SAN. The use of a unified multi-protocol storage platform that supports FC, iSCSI, and NFS, all within one storage subsystem, is needed for optimal and efficient deployment of the solution. The configuration needs to be optimized for write processing to address these diverse storage requirements ensuring that the total infrastructure remains consolidated and efficient.

Ensuring Efficiency of the Non-Virtualized Server Infrastructure

While standardization on virtual server platform is the recommended strategy for data center servers, it would be impractical and unrealistic for organizations to fully migrate all server deployments to a virtual platform. There are a number of reasons for data centers to retain some portion of their operations on the physical server infrastructure. These include performance requirements that can only be met by dedicated physical servers for specialized workloads, size of server deployments for certain business processes, size of organization, and rate of project implementation. Infrastructure resource assessment helps to identify servers that must be retained on the physical infrastructure and, if eligible for migration to a virtual server platform, the most suitable path for migration.

In a non-virtualized server environment, all aspects of storage efficiency that have been discussed earlier apply the same way without distinction for server infrastructure implementation. No separate storage infrastructure is required. The same storage implementation (SAN, NAS, CAS) is leveraged to service both the virtual and physical server infrastructure. The strategy for deriving storage efficiency and optimization in a physical server environment remains the same as in a virtual server environment. Start with an assessment of the storage infrastructure to identify areas of opportunity for storage efficiency. Leverage multi-protocol, tiered storage to consolidate storage and ensure energy and power conservation, footprint reduction, increase in storage utilization rates, and improvement in overall management and operations of storage. Use backup and recovery, business continuity, and disaster recovery implementations equally for physical servers. Use purpose-built storage for specialized operations such as information archiving. Finally, implement a closed-loop resource service orchestration strategy to automate resource management with physical servers.

Conclusion

Organizations must take a holistic view of their data center operations, including people, process, and infrastructure to build efficient, green data centers. They need a staff with business process and technical expertise. Defining, creating, and implementing operational processes are fundamental for data center efficiency. Awareness, understanding, and an effort to optimize all aspects of data center infrastructure are required to build a fully functional and efficient green data center. This includes the facility, the data center support infrastructure such as electricity and cooling equipment, and the IT infrastructure including applications, servers, networks, and storage.

With electricity quickly becoming the second largest operational expense in the data center, and the risk of data centers running out of power for operations, organizations need to pay close attention to their power and cooling infrastructures and power consumption. Through formal assessments, organizations need to understand existing data center power and cooling designs, equipment power consumption and cooling requirements, and equipment resource utilization to identify areas for optimization.

Organizations must use virtualization as the core enabler to optimize the use of IT equipment and reduce the data center's power burden. Virtualization forms the basis for boarder efficiency realization across all functional aspects of IT including data mobility, backup and recovery, business continuity, disaster recovery, and management. Virtualization enables and complements important IT initiatives such as data migration, consolidation, and automation. With virtualization at the heart of the IT infrastructure, organizations build data centers that are financially efficient and environmentally safe.

Next Steps

Power, Cooling, and IT Asset Assessment with EMC and VMware Services

Organizations should perform power and cooling and server and storage utilization assessments. EMC and VMware offer a number of assessment services to help. Once existing and future power and cooling requirements have been identified and inefficiencies identified, the next step is to formulate a plan for infrastructure optimization. Organizations can create and implement a comprehensive infrastructure optimization plan to meet their own unique requirements by leveraging the expertise of EMC and VMware professionals and education services.

EMC and VMware Data Center Efficiency Solutions

To help organizations build efficient green data centers, EMC and VMware offer a broad set of hardware and software products and services. The use of EMC information lifecycle management and VMware server virtualization platforms results in improved and consistent processes and procedures for server and storage operations and management, optimization of power and cooling usage, and better utilization of server and storage resources. Organizations also achieve adequately protected information with recovery policies that are aligned with business requirements and compelling returns on IT infrastructure investments.

EMC Educational Services

EMC offers a comprehensive portfolio of education service programs. These programs are designed to help storage professionals (managers and administrators) develop storage competency to promote awareness and education to support building efficient, green data centers. This knowledge helps organizations improve day-to-day IT operations and staff productivity. Also, storage professionals develop the expertise required to successfully plan, design, implement, and manage strategic IT initiatives. Some EMC Educational Services programs include:

EMC Technology Curriculum: Suitable for EMC users, aligned with core technology areas and job roles/responsibilities.

Storage Administrator Certification: Suitable for EMC users, part of EMC Proven™ Professional framework, aligned with key technology specialties and job roles (storage administrator, storage manager, and storage architect).

EMC Storage Technologist Certification: Suitable for EMC users and overall IT industry, part of the EMC Proven Professional framework, this “open” curriculum focuses on concepts, principles, and core skills and not on products.

EMC Academic Alliance Program: “Open” curriculum about storage technology delivered by global colleges and universities; targeted to help build a highly skilled pool of future storage managers and professionals.

EMC Learning Partner Program: “Open” curriculum about storage technology delivered by leading, independent training companies; targeted at building or improving storage technology skills leading to better design and management of efficient storage infrastructures.

VMware Education Services

VMware Education Services provides a strong foundation and advanced training on the VMware infrastructure in instructor-led courses with hands-on exercises.

VMware Certified Professional: The VMware Certified Professional Program is designed for any technical individual — partners, end users, resellers, and consultants.

Additional Information

Contact your local EMC and VMware sales representative or visit:

http://www.EMC.com
http://www.vmware.com

References

1. (1): IDC Study
2. (2): EPA Study
3. APC white paper "Electrical Efficiency Measurement for Data Centers" By Neil Rasmussen" White paper # 154.
4. APC white paper "Ten Cooling Solutions to Support High Density Server Deployment" By Peter Hannaford. White Paper # 42
5. The Uptime Institute special report; March 2008
6. VMware whitepaper "The Roadmap to Virtual Infrastructure"

EMC Corporation
Hopkinton
Massachusetts
01748-9103
1-508-435-1000
In North America 1-866-464-7381
www.EMC.com