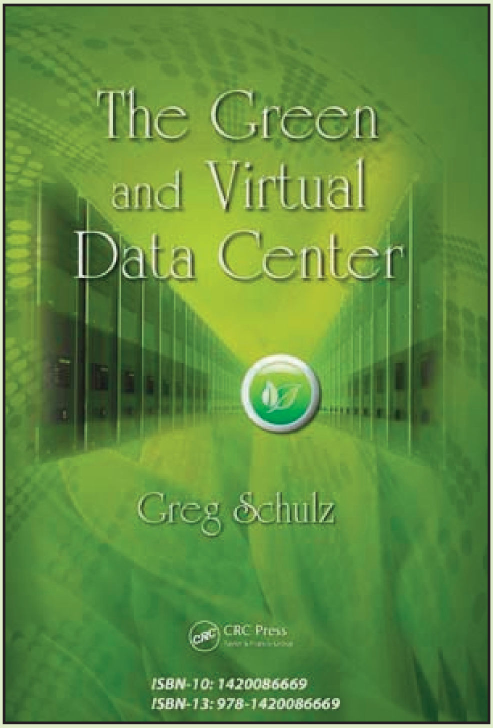
SearchStorage.com Chapter Download

THE GREEN AND VIRTUAL DATA CENTER

Chapter 4: IT Infrastructure Resource Management



ISBN-10: 1420086669
ISBN-13: 978-1420086669

By Greg Schulz

TechTarget
The Technology Media ROI Experts

Chapter 4

IT Infrastructure Resource Management

You can't go forward if you cannot go back.—Jerry Graham, IT professional

In this chapter you will learn:

- What infrastructure resource management (IRM) is
- Why IRM is an important part of a virtual data center
- How IRM differs from data and information life-cycle management
- Data protection options for virtualized environments

Leveraging various tools, technologies, and techniques to address various pain points and business challenges is key to enabling a green and virtual data center. Best practices, people, processes, and procedures combine with technology tools, hardware, software, networks, services, and facilities to enable a virtual data center. The importance of this chapter is to understand how all these elements coupled with existing and emerging technologies can be applied to improve IT service delivery in a cost-effective and energy-efficient manner. All of this together allows the data center to meet service-level requirements while sustaining business growth.

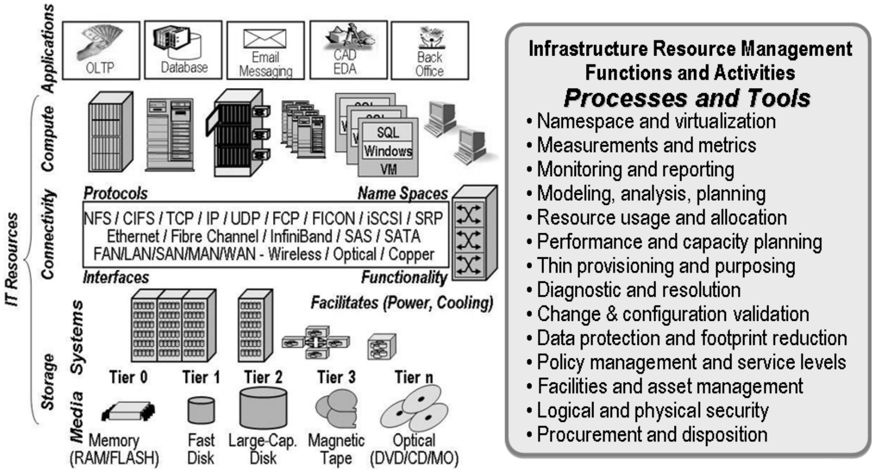
Tenets of a green and environmentally friendly virtual data center include improved productivity to enable more work to be processed, more information to be stored and accessed, while using less energy to boost productivity and business agility. A green and virtualized data center is:

- Flexible, scalable, stable, agile, and highly resilient or self-healing
- Able to adapt and leverage technology improvements quickly
- Application and data transparent from physical resources

- Efficient and effective without loss of performance or increased cost complexity
- Environmentally friendly and energy efficient yet economical to maintain
- Highly automated and seen as an information factories rather than a cost center
- Measurable with metrics and reporting to gauge relative effectiveness
- Secure from various threat risks without impeding productivity

Infrastructure resource management (IRM) is the collective term that describes the best practices, processes, procedures, and technology tools to manage IT data center resources. IRM has a focus across multiple technology domains (applications, servers, networking, storage, and facilities) to address effective and maximum resource usage to deliver a given level of application service or functionality. IRM focuses on processes, procedures, hardware, and software tools that facilitate application and data management tasks. Although there can be areas of overlap, the aim of IRM is to deliver application services and information to meet business service requirement objectives while addressing performance, availability, capacity, and energy (PACE) and power, cooling, floor space, and environmental (PCFE) requirements in a cost-effective manner. Examples of IRM encompassing functions and IT disciplines across different technology domains are shown in Figure 4.1.

Part of the process of implementing a virtual data center is to remove barriers and change traditional thinking such as hardware vs. software, servers vs. storage, storage vs. networking, applications vs. operating systems and IT equipment vs. facilities. A reality is that hardware cannot exist without software, and software cannot exist or function without hardware. Servers need networks; networks need storage. Collectively, all these IT resources need a habitat with adequate PCFE capabilities to be functional. As a result, a virtual data center looks at bridging gaps between different functional groups and technology domain areas to improve management and agility to support growth and improve application service delivery.



Infrastructure Resource Management Functions and Activities

Processes and Tools

- Namespace and virtualization
- Measurements and metrics
- Monitoring and reporting
- Modeling, analysis, planning
- Resource usage and allocation
- Performance and capacity planning
- Thin provisioning and purposing
- Diagnostic and resolution
- Change & configuration validation
- Data protection and footprint reduction
- Policy management and service levels
- Facilities and asset management
- Logical and physical security
- Procurement and disposition

Figure 4.1 IRM Functions across Different Technology Domains

4.1 Common IRM Activities

As shown in Figure 4.1, there are many different tasks and activities along with various tools to facilitate managing IT resources across different technology domains. In a virtual data center, many of these tools and technologies take on increased interdependencies because the reliance on abstracting physical resources to applications and IT services.

Consequently, while specific technology domains may focus on specific areas, interdependencies across IT resource areas are a matter of fact for efficient virtual data centers. For example, provisioning a virtual server requires configuration and security of the virtual environment, physical servers, storage, and networks as well as associated software and facility-related resources. Backing up or protecting data for an application may involve multiple servers running different portions of an application requiring coordination of servers, storage, networks, software, and data protection tasks.

Common IRM activities involved with provisioning and managing IT resources include change control and configuration management, such as updating business continuity and disaster recovery (BC/DR) plans and documents or validatng configuration settings to avoid errors. Other tasks related to change control and configuration management include notification of changes and interdependencies to various systems or IT resources, fallback and contingency planning, maintaining templates, blueprints, run-books, and guides for configuration, as well as change and configuration testing and coverage analysis. Another IRM task involves configuring physical resources, including server and operating system or virtual machine setup, networking and input/output (I/O) configuration, storage system RAID (Redundant Array of Independent Disks), data protection and security, along with storage allocation including logical unit number (LUN) creation, mapping, and masking. Other configuration and resource allocation IRM activities include ongoing physical and virtual software patch and update management, high-availability and failover configuration, network zoning, routing, and related security tasks, as well as cabling and cable management.

IRM activities also entail creating and configuring virtual resources from physical resources—for example, establishing virtual servers and virtual machines, virtual storage (disk and tape based) and file systems, and virtual networks and I/O interfaces. Once virtual resources are created from configured physical resources, data and applications need to be moved and migrated to the available servers, storage, and network entities. For example, as part of routine IRM activities, migration or conversion of applications, servers, software, and data to virtual machiness involves converting guest operating systems and applications and moving them from physical to virtual servers. Another example is the movement of data and applications from an existing storage system to a new or upgraded storage solution. Other IRM activities involve physical-to-virtual, virtual-to-virtual, and virtual-to-physical resource movement and migration for routine maintenance, load-balancing, technology upgrades, or in support of business continuity and disaster recovery.

Data protection and security are important IRM tasks to ensure that data is available when needed but safely secured from various threat risks. Protecting data involves logical and physical security, including encryption and authentication as well as ensuring that copies exist, such as using snapshots, replication, and backup of data to meet service objectives. Another dimension of IRM activities across servers, storage, networks, and facilities is monitoring, managing, analyzing, and planning. These tasks involve resource usage monitoring, accounting, event notification, and reporting, as well as determining what resources can be consolidated and which ones need scaling to boost performance, capacity, or availability. Other tasks involve balancing to various service levels for different applications performance, availability, capacity, and energy, along with applicable reporting. Diagnostics, troubleshooting, event analysis, proactive resource management, and interdependency analysis between business functions and IT resources including asset and facilities management are also central IRM tasks.

4.2 Data Security (Logical and Physical)

There are many different threat risks for IT data centers, systems, applications, and the data they support. These range from acts of man to acts of nature and include technology failure, accidental, and intended threats. Many organizations feel that, other than malicious threats, their real threats are internal, and that how and who they hire and retain as employees is a differentiator. Most organizations agree that threats vary by application, business unit, and visibility. A common belief is that most threat risks are external; in reality, however, most threats (except for acts of nature) are internal. Some organizations believe that their firewalls and other barriers are strong enough to thwart attacks from outside. Some feel that firewalls and similar technology provide a false sense of security with little protection from internal threats.

Threats can be physical or logical, such as a data breach or virus. Different threat risks require multiple rings or layers of defenses for various applications, data, and IT resources, including physical security. Virtual data centers rely on both logical and physical security. Logical security includes access controls or user permissions for files, objects, documents, servers, and storage systems as well as authentication, authorization, and encryption of data. Another facet of logical security is the virtual or physical destruction of digital information known as digital shredding. For example, when a disk storage system, removable disk or tape cartridge, laptop or workstation is disposed of, digital shredding ensures that all recorded information has been securely removed. Logical security also includes how storage is allocated and mapped or masked to different servers, along with network security including zoning, routing, and firewalls.

Physical data protection includes securing facilities and equipment and access to management interfaces or workstations. Another dimension of physical security includes ensuring that data being moved or transported electronically over a network or physically is logically secured with encryption and physical safeguards including audit trails and tracking technology. For example, solutions are available today to retrofit existing magnetic tape and removable hard disk drives with external physical barcode labels that include embedded radio frequency identification (RFID) chips. The RFID chips can be used for rapid inventory of media being shipped to facilitate tracking and eliminate falsely reported lost media. Other enhancements include shipping canisters with global positioning system and other technologies to facilitate tracking during shipping.

In general, the overwhelming theme is that there is a perception that encryption key management is complex and this complexity is a barrier to implementation. Not protecting data, particularly data "in flight," with encryption due to fears of losing keys is similar to not locking your car or home because you might lose your keys. Key management solutions are available from different sources, with some solutions supporting multiple vendors' key formats and technologies.

Some organizations are exploring virtual desktop solutions as a means of moving away from potential desktop data exposure and vulnerabilities. Many organizations are racing to encrypt laptops as well as desktops. Some organizations limit USB ports to printer use only. Some organizations are also beefing up audit trails and logs to track what data was moved and copied where, when, and by whom. USB devices are seen as valuable tools, even given all of their risks, for moving and distributing data where networks don't exist or are not practical.

An evolving dimension to protecting data and securing virtual data centers is distributed remote offices along with traveling or telecommuting workers who occupy virtual offices. The threat risks can be the same as for a primary traditional data center and include others such as loss or theft of laptops, workstations, personal digital assistants (PDAs) or USB thumb drives containing sensitive information. When it comes to security, virtual data centers require multiple levels of logical and physical security across different technology domains.

4.3 Data Protection and Availability for Virtual Environments

If data is important enough to be backed-up or replicated, or if it needs to be archived for possible future use, then it is important enough to make multiple copies (including on different media types) at different locations.

There are many challenges but also options related to protecting data and applications in a virtual server environment. For example, in a non-virtualized server environment, the loss of a physical server will affect applications running on that server. In a highly aggregated or consolidated environment, the loss of a physical server supporting many virtual machines (VMs) will have a much more significant impact, affecting all the applications supported by the virtual servers. Consequently, in a virtual server environment, a sound data protection strategy is particularly important.

Popular approaches and technologies to implement server virtualization include Citrix/Xen, Microsoft, Virtual Iron, and VMware, as well as vendor-specific containers and partitions. Many data protection issues are consistent across different environments with specific terminology or nomenclature. Virtual server environments often provide tools to facilitate maintenance and basic data protection but lack tools for complete data protection, business continuity, or disaster recovery. Instead, virtual server vendors provide **application programming interfaces (APIs),** other tools, or solution/software development kits (SDKs) so that their partners can develop solutions for virtual and physical environments. For example, solutions from VMware and Virtual Iron include SDKs and APIs to support pre- and postprocessing actions for customization and integration with VMware Consolidated Backups (VCBs), VMotion, or LiveMigration from Virtual Iron.

4.3.1 Time to Re-Architect and Upgrade Data Protection

A good time to rethink data protection and archiving strategies of applications and systems data is when server consolidation is undertaken. Instead of simply moving the operating system and associated applications from a "tin"-wrapped physical server to a "software"-wrapped virtual server, consider how new techniques and technologies can be leveraged to improve performance, availability, and data protection. For example, an existing server with agent-based backup software installed sends data to a backup server over the LAN for data protection. However, when it is moved to a virtual server, the backup can be transitioned to a LAN-free and server-free backup. Thus LAN and other performance bottlenecks can be avoided.

From a historical data protection perspective, magnetic tape has been popular, cost effective, and the preferred data storage medium for retaining data. Recently, many organizations have been leveraging storage virtualization

Table 4.1 Data Protection Options for Virtual Server Environments

Capability	■ Characteristics	■ Description and Examples
Virtual machine (VM) migration	■ Move active or static VMs ■ Facilitate load balancing ■ Provides pro-active failover or movement as opposed to data recovery and protection	■ Vmotion, Xenmotion, LiveMigration ■ May be physical processor architecture dependent ■ Moves running VMs from server to server ■ Shared-access storage required along with another form of data protection to prevent data loss
Failover high availability (HA)	■ Proactive movement of VMs ■ Automatic failover for HA ■ Local or remote HA ■ Fault containment/isolation ■ Redundant Array of Independent Disks (RAID)-protected disk storage	■ Proactive move of a VM to a different server ■ Requires additional tools for data movement ■ Low-latency network bandwidth needed ■ Replication of VM and application-septic data ■ Isolate from device failure for data availability
Snapshots	■ Point-in-time (PIT) copies ■ Copies of current VM state ■ May be application aware	■ Facilitate rapid restart from crash or other incident ■ Guest operating system, VM, appliance, or storage system based ■ Combine with other forms of data protection

Table 4.1 Data Protection Options for Virtual Server Environments (continued)

Backup and restore	▪ Application based ▪ VM or guest operating system based ▪ Console subsystem based ▪ Proxy server based ▪ Backup server or target resides as guest in a VM	▪ Full image, incremental, differential, or file level ▪ Operating system- and application-specific support ▪ Agent or agent-less backup in different locations ▪ Backup over LAN to server or backup device ▪ Backup to local or SAN attached device ▪ Proxy-based for LAN and server-free backup
Local and remote replication	▪ Application based ▪ VM or guest operating system based ▪ Console subsystem based ▪ External appliance based ▪ Storage array based	▪ Application or operating system based ▪ Network, fabric, or storage system based ▪ Application-aware snapshot integration ▪ Synchronous for real-time low latency ▪ Asynchronous for long distance, high latency
Archiving	▪ Document management ▪ Application based ▪ File system based ▪ Compliance or preservation	▪ Structured (database), semistructured (email), and unstructured (files, PDFs, images, video) ▪ Compliance or regulatory based ▪ Noncompliance for long-term data retention

in the form of transparent access of disk-based backup and recovery solutions. These solutions emulate various tape devices and tape libraries, and coexist with installed backup software and procedures. Magnetic tape remains one of, if not the most, efficient data storage medium for inactive or archived data. Disk-to-disk (D2D) snapshots; backups, and replication have become popular options for near-term and real-time data protection.

With a continuing industry trend toward using D2D for more frequent and timely data protection, tape is finding a renewed role in larger, more infrequent backups for large-scale data protection in support of long-term archiving and data preservation. For example, D2D, combined with compression and de-duplication disk-based solutions, is used for local, daily, and recurring backups. Meanwhile, weekly or monthly full backups are sent to tape to free disk space as well as address PCFE concerns.

4.3.2 Technologies and Techniques—Virtual Server Data Protection Options

Just as there are many approaches and technologies to achieve server virtualization, there are many approaches for addressing data protection in a virtualized server environment. Table 4.1 provides an overview of data protection capabilities and characteristics in a virtualized server environment.

Complete and comprehensive data protection architectures should combine multiple techniques and technologies to meet various **recovery time objectives (RTOs)** and **recovery point objectives (RPO)**. For example, VM movement or migration tools such as VMware VMotion or Virtual Iron LiveMigration provide proactive movement for maintenance or other operational functions. These tools can be combined with third-party data movers, including replication solutions, to enable VM crash restart and recovery or basic availability. Such combinations assume that there are no issues with dissimilar physical hardware architectures in the virtualized environment.

Data protection factors to consider include:

- RTO and RPO requirements per application, VM/guest or physical server
- How much data changes per day, along with fine-grained application-aware data protection
- Performance and application service-level objectives per application and VM
- The distance over which the data and applications need to be protected
- The granularity of recovery needed (file, application, VM/guest, server, site)

- Data retention as well as short-term and longer-term preservation (archive) needs

Another consideration when comparing data protection techniques, technologies, and implementations is application-aware data protection. Application-aware data protection ensures that all data associated with an application, including software, configuration settings, data, and current state of the data or transactions, is preserved.

To achieve application-aware and comprehensive data protection, all data, including memory-resident buffers and caches pertaining to the current state of the application, needs to be written to disk. At a minimum, application-aware data protection involves quiescing (suspending) file systems and open files data to be written to disk prior to a snapshot, backup, or replication operation. Most VM environments provide tools and APIs to integrate with data protection tasks.

4.3.3 Virtual Machine Movement and Migration

Often mistaken, or perhaps even positioned, as data protection tools and facilities, virtual machine movement or migratory tools are targeted and designed for maintenance and proactive management. The primary focus of live VM migration tools is to be able to proactively move a running or active VM to a different physical server without disrupting service.

For example, VMotion can be used to maintain availability during planned server maintenance or upgrades or to shift workload to different servers based on expected activity or other events. The caveat with such migration facilities is that, while a running VM can be moved, the VM still needs to be able to access its virtual and physical data stores. This means that data files must also be relocated. It is important to consider how a VM movement or migration facility interacts with other data protection tools including snapshots, backup, and replication, as well as with other data movers.

In general, considerations in live movement facilities for VMs include:

- How does a VM mover support dissimilar hardware architectures (e.g., Intel and AMD)?

- Is there a conversion tool (e.g., physical to virtual), or does it perform live movement?
- Can the migratory or movement tool work on both a local and wide area basis?
- How do tools interact with other data protection tools when data is moved with the VM?
- What are the ramifications of moving a VM and changes to Fibre Channel addressing?
- How many concurrent moves or migrations can take place at the same time?
- Is the movement limited to virtual file system-based VMs, or does it include raw devices?

4.3.4 High Availability

Virtual machine environments differ in their specific supported features for **high availability (HA),** ranging from the ability to failover or restart a VM on a different physical server to the ability to move a running VM from one physical server to another physical server (as discussed in the previous section). Other elements of HA for physical and virtual environments include eliminating single points of failure to isolate and contain faults, for example, by using multiple network adapters, redundant storage I/O host bus adapters, and clustered servers.

A common approach for HA data accessibility is RAID-enabled disk storage to protect against data loss in the event of a disk drive failure. For added data protection, RAID data protection can be complemented with local and remote data mirroring or replication to protect against loss of data access due to a device, storage system, or disk drive failure. RAID and mirroring, however, are not a substitute for backup, snapshots, or other point-in-time discrete copy operations that establish a recovery point.

RAID provides protection in the event of disk drive failures; RAID does not by itself protect data in the event that an entire storage system is damaged. While replication and mirroring can protect data if a storage system is destroyed or lost at one location, if data is deleted or corrupted at one location, that action will be replicated or mirrored to the alternate copy. Consequently, some form of time interval-based data protection, such as a snapshot or backup, needs to be combined with RAID and replication for a comprehensive and complete data protection solution.

4.3.5 Snapshots

There are a number of reasons why snapshots, also known as **point-in-time (PIT)** copies and associated technologies might be utilized. Snapshots create a virtual backup window to enable data protection when a physical backup window is shrinking or no longer exists. Snapshots provide a way of creating virtual time to get essential data protection completed while minimizing impacts to applications and boosting productivity. Different applications have varying data protection requirements, including RTO, RPO, and data retention needs. Other reasons for making snapshots include making copies of data for test purposes, including software development, regression testing, and disaster recovery testing; making copies of data for application processing, including data warehouse, data marts, reporting, and data mining; and making copies to facilitate non-disruptive backups and data migration.

Snapshots can reduce downtime or disruptions associated with traditional data protection approaches such as backup. Snapshots vary in their implementation and location, with some being full copies while others are "delta-based." For example, an initial full copy is made with deltas or changes recorded, similar to a transaction or redo log, with each snapshot being a new delta or point-in-time view of the data being protected. Snapshot implementations can also vary in where and how the snapshot data is stored.

Because snapshots can take place very quickly, an application, operating system, or VM can be quiecesed, a quick snapshot taken of the current state at that point in time, and then resume with normal processing. Snapshots work well for reducing downtime as well as speeding up backups. Snapshots reduce the performance impact of traditional backups by only copying changed data, similar to an incremental or differential backup but on a much more granular basis. Snapshots can be made available to other servers in a shared storage environment to further off-load data protection. An example is using a proxy or backup server to mount and read the snapshots to construct an offline backup.

For virtual environments, snapshots can be taken at the VM or operating system layer, with specific features and functionalities varying by vendor implementation. Snapshots can also be taken in storage systems that are integrated with a guest operating system, applications, or VM. Snapshots can also be taken in network- or fabric-based appliances that intercept I/O data streams between servers and storage devices.

One of the key points in utilizing snapshots is to make sure that when a snapshot is taken, the data that is captured is the data that was expected to be recorded. For example, if data is still in memory or buffers, that data may not be flushed to disk files and captured. Thus, with fine-grained snapshots, also known as near or coarse continuous data protection (CDP), as well as with real-time fine-grained CDP and replication, 100% of the data on disk may be captured. But if a key piece of information is still in memory and not yet written to disk, critical data to ensure and maintain application state coherency and transaction integrity is not preserved. While snapshots enable rapid backup of data as of a point in time, snapshots do not by themselves provide protection in the event of a storage system failure; thus, snapshots need to be backed up to another device.

4.3.6 Agent-Based and Agent-Less Data Protection

Agent-based backup, also known as LAN-based backup, is a common means of backing up physical servers over a LAN. The term comes from the fact that a backup agent (backup software) is installed on a server, with the backup data being sent over a LAN to a backup server or to a locally attached tape or disk backup device.

Given the familiarity with and established existing procedures for using LAN- and agent-based backup, a first step for data protection in a virtual server environment may be to simply leverage agent-based backup while re-architecting virtual server data protection.

Agent-based backups as shown in Figure 4.2 are relatively easy to deploy, as they may be in use for backing up the servers being migrated to a virtual environment. Their main drawback is that they consume physical memory, CPU, and I/O resources, causing contention for LAN traffic and impacting other VMs and guests on the same virtualized server.

Backup client or agent software can also have extensions to support specific applications such as Exchange, Oracle, SQL, or other structured data applications as well as being able to handle open files or synchronize with snapshots. One of the considerations in using agent-based backups, however, is what support exists for the backup devices or targets. For example, are locally attached devices supported from an agent, and how can data be moved to a backup server over a network in a LAN-friendly and efficient manner?

Physical servers, when running backups, have to stay within prescribed backup windows while avoiding performance contention with

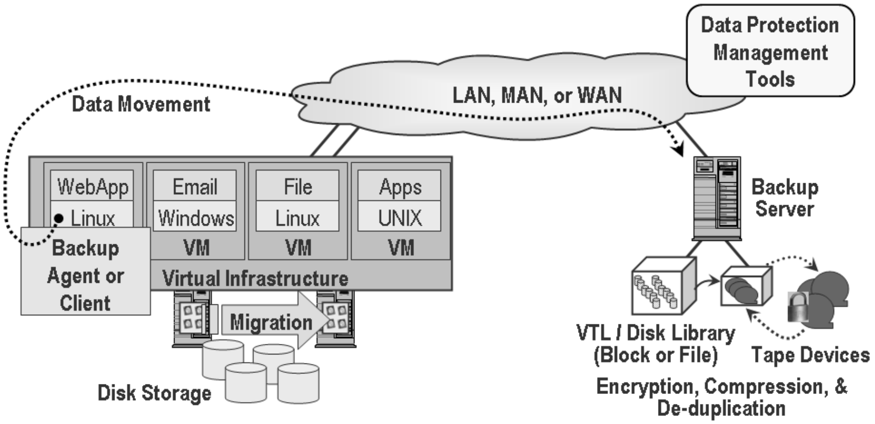


Figure 4.2 Agent-Based Backup over a LAN

other applications and also avoiding LAN traffic contention. In a consolidated virtual server environment, multiple competing backup jobs may also vie for the same backup window and server resources. Care needs to be exercised when consolidating servers into a virtual environment to avoid performance conflicts and bottlenecks.

4.3.7 Proxy-Based Backup

Agent- or client-based backups running on guest operating systems consume physical resources, including CPU, memory, and I/O, resulting in performance challenges for the server and the LAN (assuming a LAN backup). Similarly, an agent-based backup to a locally attached disk, tape, or virtual tape library (VTL) will still consume server resources, resulting in performance contention with other VMs or other concurrently running backups.

In a regular backup, the client or agent backup software, when requested, reads data to be backed up and transmits the data to the target backup server or storage device while also performing associated management and record keeping tasks. Similarly, during restore operations, the backup client or agent software works with the backup server to retrieve data based on the specific request. Consequently, the backup operation places a demand burden on the physical processor (CPU) of the server while consuming memory and I/O bandwidth. These competing demands can and need to be managed if multiple backups are running on the same guest operating system and VM or on different VMs.

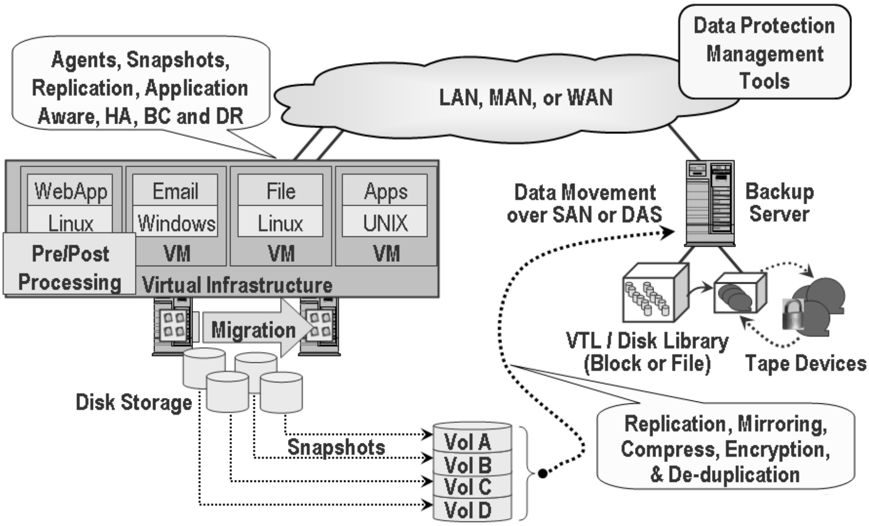

Agents, Snapshots, Replication, Application Aware, HA, BC and DR
LAN, MAN, or WAN
Data Protection Management Tools
WebApp
Linux
Email
Windows
File
Linux
Apps
UNIX
Pre/Post Processing
VM
VM
VM
Virtual Infrastructure
Migration
Disk Storage
Data Movement over SAN or DAS
Backup Server
VTL / Disk Library (Block or File)
Tape Devices
Snapshots
Vol A
Vol B
Vol C
Vol D
Replication, Mirroring, Compress, Encryption, & De-duplication

Figure 4.3 VMware VCB Proxy-Based Backup Example

One approach to addressing consolidated backup contention is to leverage a backup server and configure it as a proxy (see Figure 4.3) to perform the data movement and backup functions. Proxy backups work by integrating with snapshot, application, and guest operating system tools for pre- and postprocessing. As an example, VMware Consolidated Backup (VCB) is a set of tools and interfaces that enable a VM, its guest operating system, applications, and data to be backed up by a proxy while reducing the CPU, memory, and I/O resource consumption of the physical server compared to a traditional backup.

VCB is not a backup package. Rather, it is an interface to VMware tools and enables third-party backup and data protection products to work. To provide data protection using VCB, third-party backup tools are required to provide scheduling, media, and backup management. Third-party tools also manage the creation of data copies or redirect data to other storage devices, such as virtual tape libraries and disk libraries, equipped with compression and data de-duplication to reduce data footprint. VM virtual disk images are sparse or hollow, meaning that there is a large amount of empty or blank space with many similar files that lend themselves to being compressed and de-duplicated.

In addition to off-loading the physical server during the proxy backup, LAN traffic is not affected, as data can be moved or accessed via a

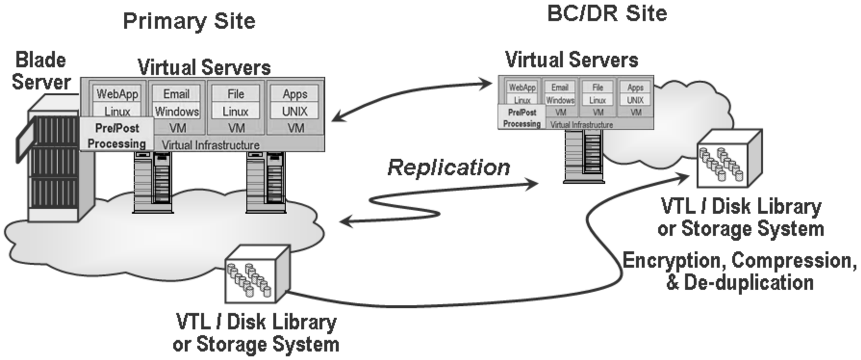


Figure 4.4 Data Replication for High Availability and Data Protection

shared storage interconnect. Third-party backup and data protection software on a proxy server can also perform other tasks, including replicating the data to another location, keeping a local copy of the backup on disk-based media with a copy at the remote site on disk, as well as on a remote offline tape if needed.

4.3.8 Local and Remote Data Replication

There are many approaches to data replication and mirroring, shown generically in Figure 4.4, for local and remote implementations to address different needs, requirements, and preferences.

An important consideration in data mirroring and replication is distance and latency, which may result in data delay and negative performance impacts. Distance is a concern, of course, but the real enemy of synchronous data movement and real-time data replication without performance compromise is latency. There is a common perception that distance is the enemy of synchronous data movement. Generally speaking, latency increases over distance; thus, the common thinking is that distance is the problem in synchronous data movement. The reality is that even over relatively short distances, latency can negatively impact synchronous real-time data replication and data movement.

Distance and latency bear on replication and data movement by affecting decisions as to whether to use synchronous or asynchronous data movement methods. Besides cost, the major factors are performance and data protection. Synchronous data transfer methods facilitate real-time data protection, enabling a recovery point objective (RPO) of near zero. However, the trade-off is that over distance, or, high-latency networks, application performance is negatively impacted while the system waits for remote I/O operations to be completed.

Another approach to the problem is to use asynchronous data transfer modes, in which a time delay is introduced along with buffering. By using a time delay and buffering, application performance is not affected, as I/O operations appear to applications as having completed. The trade-off with asynchronous data transfer modes is that although performance is not degraded over long-distance or high-latency networks, there is a larger RPO exposure potential for data loss while data is in buffers waiting to be written to remote sites.

A combination of synchronous and asynchronous data transfer may be used, providing a tiered data protection approach—for example, using synchronous data transfer for time-critical data to a reasonably nearby facility over a low-latency network, and replicating less critical data asynchronously to a primary or alternative location farther away. A hybrid approach is to perform synchronous data replication to a nearby facility, then perform a second, asynchronous replication to another site farther away.

A general caveat is that replication by itself does not provide complete data protection; replication is primarily for data availability and accessibility in the event of a component, device, system, or site loss. Replication should be combined with snapshots and other point-in-time discrete backup data protection to ensure that data can be recovered or restored to a specific RPO. For example, if data is corrupted or deleted on a primary storage device, replication will replicate the corruption or deletion to alternate sites, hence the importance of being able to recover to specific time intervals for rollback.

4.3.9 Archiving and Data Preservation

Data preservation or archiving of structured (database), semistructured (email and attachments), and unstructured (file-oriented) data is an effective means to reduce data footprint and associated PCFE, backup/recovery, business continuity, disaster recovery, and compliance issues. Given the current focus on addressing PCFE-associated issues and the growing awareness of the need to preserve data offline or near-line to meet regulatory requirements, magnetic tape is an effective complementary technology to D2D backups. Magnetic tape continues to be a strong solution for its long-term cost and performance and its effective offline data preservation.

4.3.10 Complete Data Protection

Figure 4.5 shows a combination of data protection techniques including a storage area network (SAN) or network attached storage (NAS) system with RAID for data availability, local and remote data replication, snapshots to facilitate high-speed data backups, D2D, and tape-based backups with encryption. In addition, compression and de-duplication can be incorporated to help reduce data footprint and the physical storage space required, or to store more data in the same space.

Virtual backups, that is, backups to a remote managed service provider (MSP), software or storage as a service (SaaS), or cloud-based service can also be part of a complete data protection approach. For example, remote office and branch offices (ROBO) or home-based and traveling workers (virtual offices) can be backed up, replicated, and have their data protected to either an internal managed service (e.g., a virtual data center) or to a third-party service. For smaller organizations, third-party backup and data protection MSPs and cloud-based services or SaaS-based solutions are an effective means for enhancing timely and affordable data protection.

Things to consider when evaluating data protection techniques include:

- Potential threats to applications and data
- RTO and RPO requirements for applications and associated data
- Who will perform data recovery
- How transparent the data protection and recovery scheme need to be
- Technologies currently in place (hardware, software, etc.)
- Alternative techniques and technologies for data protection
- Budget, timeframe, tolerance to disruption, and risk aversion
- Which solutions are best for different applications
- Availability of experienced help for assessment, validation, or implementation

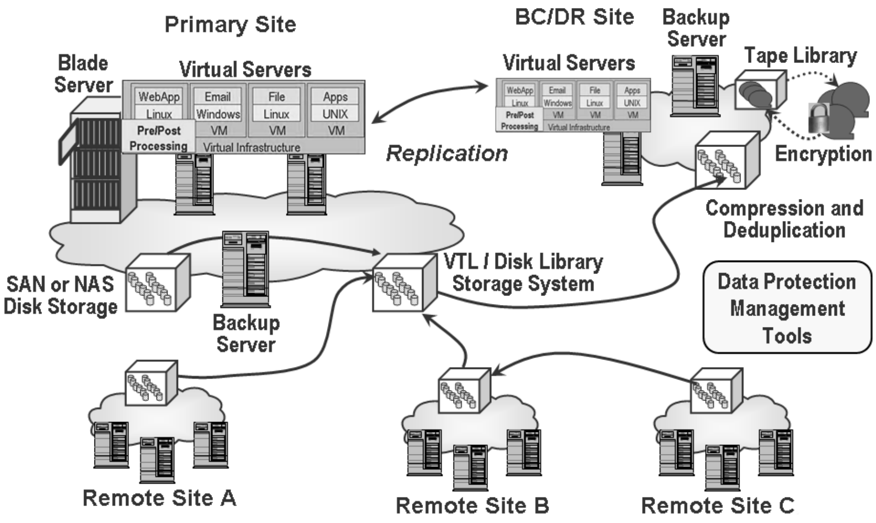


Figure 4.5 Local and Remote Data Protection

4.4 Data Protection Management and Event Correlation

Data protection management (DPM) has evolved from first-generation backup reporting technology to incorporate multivendor and cross-technology domain capabilities. In addition, DPM tools are evolving to manage multiple aspects of data protection along with event correlation.

Some DPM tools are essentially reporting, status, or event monitoring facilities that provide passive insight into what is happening in one or more areas of infrastructure resource management (IRM). Other DPM tools provide passive reporting along with active analysis and event correlation, providing a level of automation for larger environments. Cross-technology domain event correlation connects reports from various IT resources to transform fragments of event activity into useful information on how, where, why, and by whom resources are being used. In virtualized environments, given the many different interdependencies, cross-technology domain event correlation is even more valuable for looking at end-to-end IRM activities

Increasing regulatory requirements combined with pressure to meet service levels and 24-7 data availability has resulted in data protection interdependencies across different business, application, and IT entities.

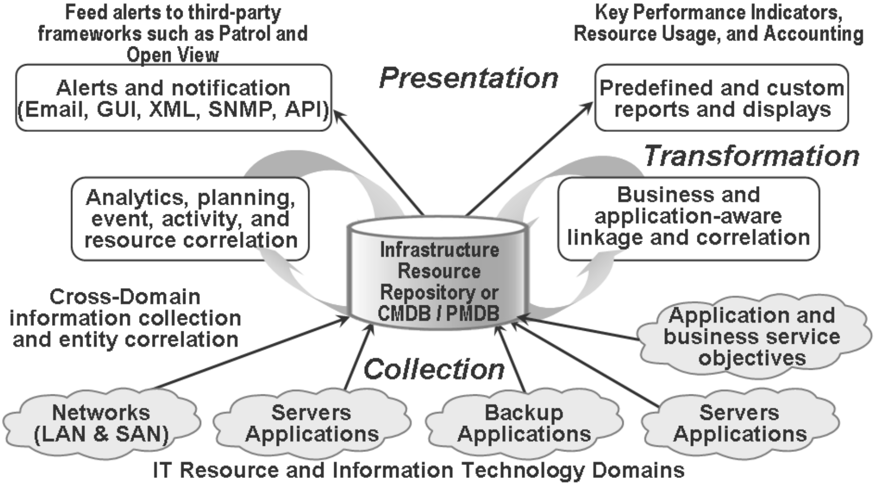

Feed alerts to third-party frameworks such as Patrol and Open View
Key Performance Indicators, Resource Usage, and Accounting
Alerts and notification (Email, GUI, XML, SNMP, API)
Presentation
Predefined and custom reports and displays
Transformation
Analytics, planning, event, activity, and resource correlation
Business and application-aware linkage and correlation
Infrastructure Resource Repository or CMDB / PMDB
Cross-Domain information collection and entity correlation
Application and business service objectives
Collection
Networks (LAN & SAN)
Servers Applications
Backup Applications
Servers Applications
IT Resource and Information Technology Domains

Figure 4.6 Data Protection Management Across Technology Domains

Consequently, timely and effective DPM requires business and application awareness to correlate and analyze events that affect service and IT resource usage. Business awareness is the ability to collect and correlate IT assets to application interdependencies and resource usage with specific business owners or functions for reporting and analysis. Application awareness is the ability to relate IT resources to specific applications within the data protection environment to enable analysis and reporting.

A challenge in business- and application-aware DPM has been the fact that many organizations maintain information about business units, applications, IT resources, or asset ownership and usage in disparate database and repository formats. For example, information is kept in configuration management databases (CMDB), performance management databases (PMDB), and metadata repositories, among other locations. To support business- and application-aware data protection, DPM tools need to support access of external sources.

Effective DPM includes knowing what, where, when, why, and by whom resources are being used to deliver service and how effectively that service is being delivered (Figure 4.6). To enable timely IRM across different technology domains and disciplines, including DPM, automated data collection and correlation of event and activity information is needed. Event correlation from different sources facilitates root-cause analysis so that service levels and compliance objectives can be meet. With a focus on reducing electrical energy consumption and associated environmental impacts, DPM can be used to ensure that data protection resources are being used optimally.

Although an environment may have multiple tools and technologies to support IRM activities, DPM tools are evolving to support or coexist with management of multiple data protection techniques, including backup (to disk or tape), local and remote mirroring or replication, snapshots, and file systems. Key to supporting multiple data protection approaches and technologies is the ability to scale and process in a timely manner large amounts of event and activity log information. At the heart of a new breed of IRM tools, including DPM solutions, are robust cross-technology resource analysis and correlation engines to sift disparate data protection activity and event logs for interrelated information.

Examples of products that have event correlation as their basis with different personalities (capacity management, compliance coverage, data protection management, configuration validation, replication management, network management) include Akorri, Continuity, EMC (Smarts), Onaro (now part of NetApp), and WysDM (now part of EMC). While the mentioned tools perform a specific or multiple IRM-related functions, they all support cross-technology domain event correlation.

4.5 Server, Storage, and Network Resource Management

Performance and capacity planning can be combined as complementary activities along with server or **storage resource management (SRM)** and utilization, or they may be handled as separate tasks. Performance tuning and optimization may initially be seen as reactionary tasks to respond to specific situations. A performance plan and ongoing performance tuning initiative can support a shift from reactionary to tactical and longer-term strategic management approaches. For example, shifting to a performance plan approach, in which performance and usage are analyzed and optimized as part of an overall growth plan, can help maximize and optimize spending.

IRM reporting and monitoring tools should allow an IT administrator to see across different technology domains and from virtual server to physical storage for the full IRM picture. In addition, capacity and resource usage tools add performance or activity reporting to traditional space or capacity utilization to provide a more holistic view of resource usage. Performance and utilization should be evaluated in tandem. It's bad policy to scale up utilization only to find that performance is suffering.

4.5.1 Search and eDiscovery

Data classification and search tools have several functions, including discovery, classification, and indexing, as well as searching, reporting, and taking action on discovered data—for example, identifying what files to migrate from active online storage to offline storage for archive purposes. For compliance-related data, taking action includes marking data for litigation hold to prevent tampering or taking action such as deleting data based on policies. In general, taking action refers to the ability to interface with various storage systems (online, near-line, and offline), including object-based and archiving systems, to enable management and migration of data.

Storage resource management and basic data discovery and classification tools include file path and file meta data discovery tools. SRM tools have a vertical focus on storage and file identification for storage management purposes, including allocation, performance, and reporting. Some tools provide basic SRM-like functionality along with more advanced capabilities including archiving, document management, email, and data migration capabilities. Deep content discovery, indexing, classification, and analysis tools support features such as word relativity, advanced language support, and search and discovery features for vertical markets.

When looking at data discovery and indexing tools, the intended and primary use of the technology should be kept in mind. For example, is the planned use of the tools to perform deep content discovery for compliance, legal litigation, and intellectual property search? Perhaps you are looking to identify what files exist, when they were last accessed, and what might be candidates for moving to different tiers of storage. By keeping primary objectives in focus, you may find that different tools work better for various tasks, and that more than one tool is needed.

Architectural considerations include performance, capacity, and depth of coverage, along with discovery, security and audit trails. Policy management should be considered, along with policy execution, interfaces with other policy mangers, and data migration tools. Some tools also support interfaces to different storage systems such as vendor-specific APIs for archiving and compliance storage. Consider whether the candidate tools have embedded or built-in support for processing different templates, lexicons, syntax, and taxonomies associated with different industries and regulations. For example, when dealing with financial documents, the tool should support processing of data in the context of various financial taxonomies such as banking, trading, benefits, and insurance, among others. If legal documents are being processed, then support for legal taxonomies will be needed.

Classifying data is complex, and for some services providers who merely "house" data, the actual value of the data may not be known. Although tools exist, they are limited in their extensiveness and scalability. Interaction with lines of business and those developing the applications are important to understand the value of data. Tiered security is needed, but a methodology also needs to exist and be tied to data value, location, and line of business.

Understanding target applications and needs for discovery tools will help to ensure a positive and successful solution. To understand what files exist on a system to help implement a tiered storage environment, start by looking at traditional SRM-type tools. If, on the other hand, deep data discovery is needed to support litigation, compliance, or other functions, then consider more advanced tools. Some tools can meet multiple objectives, but it is important to understand what other aspects of a system may be affected.

4.5.2 Rescuing Stranded or Orphaned Resources

Orphaned storage is any data, file, table space, object, file system, LUN, physical volume, or storage device that appears to be in use but has been abandoned or forgotten. Orphaned storage can result from application or system errors that have not been cleared after a restart, system maintenance, upgrade, or other activity. Orphaned storage can exist in different forms and in various locations in most environments, ranging from orphaned data in a database to orphaned storage in an email system or orphaned files in networked attached storage (NAS) or traditional file system. Orphaned storage can also exist in the form of unused or unallocated LUNs, physical volumes, or even individual disk drives.

One challenge in finding orphaned storage and data is determining whether data is, in fact, orphaned. Some files may appear to be orphaned, but they may be in line to be archived, in which case they should be migrated to some other storage medium. Likewise, some data files or storage volumes may appear to be allocated, having not been released after some previous use. The data or storage may not have been de-allocated if someone forgot to tell someone else that the storage is no longer being used, or some documentation somewhere was not updated to indicate that the storage can be de-allocated and re-provisioned. Another cause of orphaned storage is system or application error. For example, over a period of time, inconsistencies can appear in databases or file systems that require a repair operation to free up unused, yet allocated, storage and index pointers.

Consider the following for finding and eliminating or adopting orphaned storage:

- Clean up temporary, scratch, and work space on a regular basis.
- Run database, application-specific, and file system consistency checks.
- Utilize vendor tools or have the vendor check for orphaned devices.
- Leverage discovery and SRM tools to verify how storage is being used.
- Use configuration analysis tools to validate storage configurations.
- Look for files that appeared around the time a system or application error occurred.
- Have database administrators (DBAa) check for duplicate data, orphaned rows or tables.
- Apply policies as part of a clean-up after upgrades to find orphaned storage.

4.5.3 Capacity, Availability, and Performance Planning

There may not appear to be a link between availability and performance and capacity planning, but there is a direct connection. If a resource is not available, performance is affected. And if a resource has poor performance or limited supply, availability and accessibility are affected.

Capacity planning and capacity management are used in a variety of businesses. In a manufacturing company, for example, they are used to manage inventory and raw goods. Airlines use capacity planning and capacity management to determine when to buy more aircraft. Electric companies use them to decide when to build power plants and transmission networks. In the same way, IT departments use capacity planning and capacity management to derive maximum value and use from servers, storage, networks, and facilities while meeting service-level objectives or requirements.

Consider some common questions and comments with regard to performance and capacity planning:

- Hardware is cheap; why tie someone up doing capacity planning and tuning?
    - While hardware is becoming less expensive, power, cooling, floor space, and environmental (PCFE) resources are increasing in cost and decreasing in availability.
- People are already busy if not overworked; why give them more to do?
    - With planning, resources (people, hardware, software, networks, and budget) can be utilized more effectively to address and maximize available PCFE resources.
- Why not buy more hardware and have a vendor manage it?
    - This may be an alternative if it is viable from cost and business perspectives. However, adding more resources adds to the expense to manage, protect, and utilize the resources as well as space to house the resources and power to operate and cool them.

Capacity and performance planning should address peak processing periods and degraded performance whether planned, scheduled, or unexpected. Performance and capacity planning activities have occurred in the enterprise environments of S/390 mainframes and open systems platforms for many years. Historically, capacity planning activities have for the most part focused on large (expensive) components, including processors, memory, network interconnects, and storage subsystems (disk and tape) for larger enterprise environments.

Capacity planning can be a one-time exercise to determine how much and what types of resources are needed to support a given application. A nontactical approach to resource needs assessment and sizing is simply to acquire some amount of resources (hardware, software, networks, and people) and buy more as needed. A strategic approach might evolve from the tactical to make more informed decisions and timed acquisitions. For example, knowing your resource needs ahead of time, you might be able to take advantage of special vendor incentives to acquire equipment that suits your needs on your terms. Similarly, if the terms are not favorable and resource usage is following the plan, you may choose to delay your purchase.

Virtual data centers help to abstract physical resources from applications and users. However, increased complexity needs to be offset with end-to-end diagnostics and assessment tools along with proactive event correlation and analysis tools. Having adequate resources when needed to sustain business growth and meet application service requirements is a balancing act. The balance is having enough server, storage, networking, and PCFE resources on hand without having too much, resulting in higher costs, or not enough, resulting in poor service.

Poor metrics and insight can lead to poor decisions and management. Look at servers from more than a percent utilization viewpoint, considering also response time and availability. Look at storage from an IOPS and bandwidth performance perspective along with response time or latency as well as available capacity. Look at networking from a latency standpoint in addition to cost per given bandwidth and percent utilization.

If you are new to capacity planning, check out the Computer Measurement Group (CMG), which is focused on cross-technology, vendor- and platform-neutral performance, and capacity planning management. In general, the recommendation is to start simple and build on existing or available experience and skill. Identify opportunities that will maximize positive results to gain buy-in and evolve to more advanced scenarios.

4.5.4 Energy Efficiency and PCFE Management Software

Power management and reporting software tools and appliances help with reporting of energy usage by various IT resources. Some solutions are able to interact and take proactive or reactive action to manage energy usage on a server, storage, network, or application basis. Solutions vary in robustness, with some able to apply business rules and polices on an application basis or technology basis (e.g. server, storage, or network).

For example, a power management solution based on set polices, including a weekly or monthly schedule, can instruct servers to go into a low-power mode or to interact with server virtualization software to shift virtual machines and applications to different servers to meet PCFE objectives. Similar to the electrical power generation and transmission environment discussed in Part I, which relies on command and control software and tools, active power management and intelligent power management tools continue to evolve to enhance command and control for energy optimization in IT data center environments.

4.6 Summary

There are many vendors with solutions to address various aspects of infrastructure resource management in a physical or virtual data center. Examples include BMC, Brocade, Cisco, Egenera, EMC, Emerson, HP, IBM, LSI, Microsoft, NetApp, Novell, Opalis, Racemi, Scalent, Sun, Symantec, Teamquest, Tek-Tools, Uptime, Vizoncore, and VMware, among others.

The benefits of server virtualization for consolidation as well as management transparency are becoming well understood, as are the issues associated with protecting data in virtualized server environments. There are many options to meet different recovery time objective and recovery point objective requirements. Virtualized server environments or infrastructures have varying functionalities and interfaces for application-aware integration to enable complete and comprehensive data protection with data and transactional integrity.

A combination of tape- and disk-based data protection, including archiving for data preservation, coupled with a data footprint reduction strategy, can help to address power, cooling, floor space, and environmental needs and issues. There is no time like the present to reassess, re-architect, and reconfigure your data protection environment, particularly if you are planning on or have already initiated a server virtualization initiative. Virtual server environments require real and physical data protection. After all, you cannot go forward from a disaster or loss of data if you cannot go back to a particular point in time and recover, restore, and restart, regardless of your business or organization size.

Resource from Fujifilm:

FUJIFILM

◀ **White Paper - Fujifilm: Optimizing Performance and Maximizing Investments in Tape Storage Systems**

Discover an assessment solution for tape storage systems. Learn how to maximize existing resources, avoid unnecessary expenditures and ensure regulation compliance.