FEBRUARY 2012

DATA CENTER OPTIMIZATION

Bringing efficiency and improved services to the infrastructure

800.800.4239 | CDW.com/datacenterguide

CDW REFERENCE GUIDE

A guide to the latest technology for people who get IT

PEOPLE WHO GET IT™

WHAT'S INSIDE:

800.800.4239 | CDW.com/datacenterguide

CHAPTER 1: Driving Data Center Efficiency 3
- Strategic Improvement
- Optimization Gains
- Optimization via Orchestration

CHAPTER 2: Optimization Strategies 6
- Updating the Existing Data Center
- Data Center in a Box
- Reference Architectures
- VM Mobility

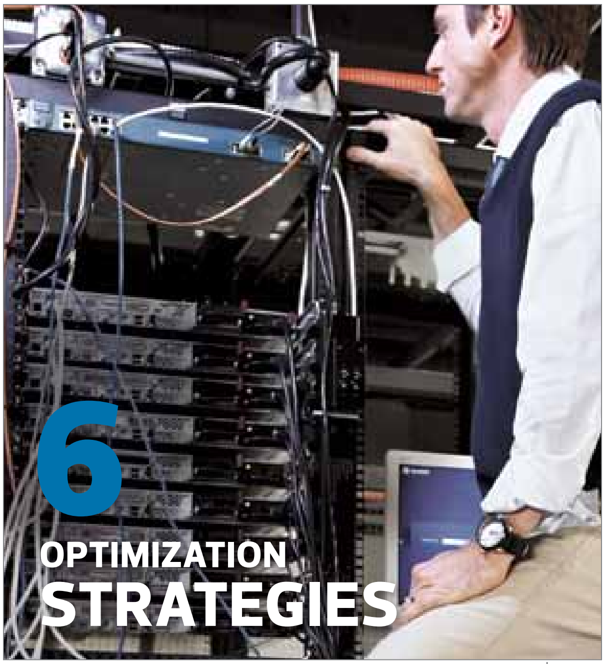
CHAPTER 3: Hardware Optimization Essentials 10
- The Network
- Servers
- Storage
- Power & Cooling

CHAPTER 4: Software Essentials 27
- Virtualization Basics
- Orchestration and Automation Tools
- Infrastructure Management Tools

CHAPTER 5: Management and Process Improvements 29
- Management Processes
- Management Tools
- Capacity Planning
- Chargeback
- Vendor Management

GLOSSARY 33

INDEX 35



VISIT CDW.com/datacenter FOR MORE INFORMATION ON OPTIMIZING THE DATA CENTER.

What is a CDW Reference Guide?

At CDW, we're committed to getting you everything you need to make the right purchasing decisions — from products and services to information about the latest technology.

Our Reference Guides are designed to provide you with an in-depth look at topics that relate directly to the IT challenges you face. Consider them an extension of your account manager's knowledge and expertise. We hope you find this guide to be a useful resource.

Strategic Improvement
Optimization Gains
Optimization via Orchestration

Driving Data Center Efficiency

What's driving data center optimization and why organizations need to plan for it.

Powerful operations, technology and economic forces have converged to drive change in enterprise data centers.

From an *operational* standpoint, organizational leadership has come to view the data center as more than a factory-like utility that collects and processes information. Top management understands the value of data that's available in real time to help inform and shape decision-making. They also expect the data center to be the enabler of new, rapidly deployed, public-facing and internal user applications.

From a *technology* standpoint, today's data center must support mobility, provisioning on demand, scalability, virtualization and the flexibility to respond to fast-changing operational situations.

From an *economic* standpoint, a few years of strained fiscal conditions have imposed tight budgets on IT organizations in both the public and private sectors. That's helped bring new thinking to data center design and management. Organizations must deliver maximum output for every dollar invested in IT. They also face pressure to reduce power use as a component of overall organizational strategies for reducing their carbon footprint.

Strategic Improvement

These three converging forces foster an exciting but challenging approach to data centers. In the mainframe days, IT groups added processors, storage or operating system upgrades as demand required.

But in reality, operational needs fluctuate. Projects come and go. The cycle of application development quickens. The processing loads now must scale up and down much more quickly than in the batch processing era. The result? CIOs must find ways to support mobile users, partners and customers with speed and agility.

And they've got to do it all within tight budget constraints and, in many cases, even static budgets.

How can organizations dial up their data center operations and fuel

KEY ELEMENTS OF AN OPTIMIZATION STRATEGY

The efficient use of physical space	This often takes the form of consolidation. Consolidation encompasses not just large, glass-house facilities, but also small server clusters and wiring closets. If there are too many data centers (no matter their size) supporting the organization, they add unnecessary cost, chip away at manageability and lead to power inefficiencies. Plus small "vampire" facilities that suck power needlessly tie up inventory and cash that could be pumped back into the IT organization.
Maximum server efficiency through the virtualization of applications and related data resources	Virtualization is fast spreading to individual users and their personal resources. Its benefits extend beyond merely improving the efficiency and usage ratio for central servers and storage systems. They also include faster provisioning of new projects, applications or users, as well as greater reliability and business continuity thanks to fast replication of virtual machines as backups.
Utilizing external cloud computing (where it makes sense)	A growing number of organizations have begun migrating select applications (principally e-mail and other utilitarian functions) to cloud environments run by third parties. That frees up staff and infrastructure for more mission-focused work such as application development.
Deploying internal clouds (when it can provide flexibility)	To support a wide range of users and services with easy scalability and rapid provisioning, many organizations have launched their own cloud computing infrastructures. Internal cloud strategies force a rethinking of hardware architectures, utilizing consolidated, converged or data-center-in-a-box approaches.
Optimized power use	New form factors, principally blade servers and converged infrastructures, need up-to-date cooling strategies. Rather than simply chilling the interior of an entire facility, the latest cooling technologies now focus concentrated cooling where and when it is specifically needed within racks and aisles to radically reduce cooling costs.
Maximize the use of storage and networking	End-to-end enterprise data architectures require rethinking the long-held strategy of continuously adding expensive disk storage to take on new loads. Instead, IT organizations now must update networking protocols and architectures to make the most use of bandwidth and capacity. This will more tightly weave together the organization's data infrastructure.

reinvestment in their backbone IT assets? By pursuing a strategy of optimization. More than a fixed end state, optimization requires continuous assessment and planning for each element in the data center: hardware, software and supporting infrastructure.

The goal of optimization is the creation of a data utility that fully supports the organization's operational goals at the lowest possible cost. The process of optimization therefore is holistic and continuous.

Optimization Gains

Although the paths to data center optimization are many, the benefits that organizations gain fall into three chief categories:

1. Lower costs and greater return on investment (ROI) for IT dollars
2. Improved infrastructure performance
3. More operational value from the technology in the form of agility, flexibility and fast response to change

Simply put, the more space and time that data centers take up, the more dollars they consume. With careful planning, optimization leads to consolidation, a smaller energy footprint and more efficient use of the computing infrastructure.

For example, by consolidating storage from dedicated, stand-alone subsystems to a single storage area network (SAN), the organization gains improved manageability and faster performance — and it saves space. The same holds true for server virtualization, which can reduce the number of needed physical servers by up to 80 percent.

Performance and business continuity also flow from data center optimization. The two are related: Applications optimized for consolidated hardware platforms yield more uptime, faster responses and greater predictability through load balancing, data caching and Transmission Control Protocol (TCP) processing.

Better business continuity accrues not simply from more reliable failover and cybersecurity, but also from constant and predictable performance available to users and customers.

Data center optimization also helps organizations realize a better return on their IT investments. For example, internal clouds offer constituents (departments and users within the organization) the kind of agility they need to respond to fluctuating demand. The cloud owner gains efficiency through higher and more consistent utilization rates and ease of centralized management.

The organization also optimizes its power costs through virtualization. The marginal increase in power consumed by a given server running at a high use rate comes in lower than the power used by two low-use-rate servers — then multiply those savings by hundreds, sometimes even thousands, of virtual machines.

In short, optimization frees resources from infrastructure operations and maintenance and makes them available to help drive innovation and operational value.

Optimization via Orchestration

Organizations under pressure because of IT infrastructure costs discover that the consolidation of data centers occurs naturally as a result of optimization, not the other way around.

Several tools and techniques have made their way into optimization efforts in recent years. But fundamentally, the state of the art in optimization encompasses not just tools and techniques, but also a primary goal of providing excellent IT services to the organization's functional components.

Server consolidation via virtualization has become the de facto technology in optimization efforts. Client virtualization comes in as the next logical step behind server virtualization. Lessons learned from hardware architectures, cooling and utility management, and automatic server provisioning apply as well to blade infrastructures that support individual users on whatever thin client or mobile devices the organization chooses to support.

But optimizing the output of the data center to meet specific service levels requires a technique called orchestration. Orchestration tools act as master controllers of subordinate tools that, for their part, operate and optimize various pieces of the technology infrastructure. For example, one product might control the allocation of memory and storage among applications. Another might manage network bandwidth.

Orchestration automates and speeds up the configuration of resources to meet service level needs defined by the organization. ■

IS YOUR BUILDING READY?

Data centers built in the 1970s on up to the 1990s might be big enough in square footage to support today's optimized data centers. But facilities for mainframe and client-server farms might not be right in other respects for the virtualized, agile, flexible and service-oriented data centers of today.

An organization might need to rebuild (or build) a whole new housing for its optimized data center.

Experts say a modern facility would do well to have in place (or plan to acquire) the following characteristics:

- Scalable electrical services that can change to match the blade and storage pool densities in the data center
- Revamped raised flooring that accommodates not only traditional wiring, but also works as a plenum for air exchange in cooling systems (Plus, flooring must be able to support the concentrated weight of multiblade racks. Older data center designs planned for more distributed weight.)
- Targeted lighting systems support the ability to read information on components from both sides of a rack and can automatically dim and then switch on and off to maximize power usage (The standard fluorescent panels in the dropped ceilings of yesteryear are inadequate in an optimized data center.)

Updating the Existing Data Center
Data Center in a Box
Reference Architectures
VM Mobility

Optimization Strategies

How to get the data center to its desired state

Keeping data center optimization aligned with the organization's mission and goals means that the management team and the IT group must plan strategically.

Optimization planning can be thought of as a three-legged stool that balances capacity, cost and service-level needs:

- For **basic capacity**, there must be a path toward sufficient storage, bandwidth, servers, real estate and utilities — not to mention people and skills.
- For **rational costs,** there must be a strategy for automation and management tools that allow IT to productively control the infrastructure.
- To meet **service-level requirements,** the organization needs an overall plan for continuously optimizing its infrastructure, applications and tools.

Optimization really begins before an IT engineer pulls a single cable. It starts with analysis. That's a top-down, bottom-up team effort. Data center planning starts with an inventory of the current situation, including not merely physical assets and speeds-and-feeds, but also the utilization rate of each and every component.

What follows are three basic strategies for moving the organization toward a service-oriented data center.

Updating the Existing Data Center

Optimization of an existing data infrastructure means identifying each component and analyzing how it can be made more efficient. For many organizations, optimization involves consolidating two, three or more data centers to reduce costs and simplify management.

It's possible to accomplish basic consolidation with a simplified merger approach. But simply squeezing two inefficient facilities into one limits savings and fails to achieve the agility and flexibility needed for a service-centered approach.

That's why consolidation works best in tandem with virtualization. Server virtualization not only reduces physical servers but also enables dynamic reallocation of server capacity. Organizations pursuing cloud computing environments need capacity on demand. Virtualization, the amassing of multiple systems in virtual containers on single physical servers, is a cost-effective way to achieve this.

Obviously, consolidation and virtualization require the IT team to rethink hardware. Within a given processor architecture, organizations that want to physically consolidate their virtualized servers often turn to blade setups. Blade servers fit in full or half-width, rack-mounted cages. Each blade contains the components of a server.

A grouping of blades shares a common chassis and networking backplane and sometimes a local storage pool.

Blade servers individually consume less power than stand-alone servers, but they mount in multiples, close together. That's why, in optimizing server hardware, it's necessary to factor in power and cooling too.

Rethinking Power & Cooling

The current thinking around cooling aims to conserve power, often as part of a green data center strategy. Rather than chill an entire room to ensure that hot spots around blade or rack-server chassis don't overheat, a variety of new techniques are available that concentrate cooling precisely where it's needed. Some systems use convection to effect cooling, which saves electricity.

It's helpful to remember that optimization is not solely about cutting costs, but also about ensuring IT service levels. That can't happen without maintaining reliability. So power management remains an important element in optimization. Power and cooling needs should be addressed in tandem.

Don't overlook floor layout. It affects how narrowly the facility can direct cool air and how efficiently the layout swaps warmed air for cool. For that matter, optimization can lead to reconsidering the building shell that will house the power-concentrated, virtualized equipment.

Many organizations originally outfitted their data centers for mainframe or minicomputer operations. Power distribution, cabling, flooring and lighting requirements have changed. For that reason, it's worth considering and pricing out architectural and construction services as part of an optimization project.

Storage Virtualization

Server virtualization spurs the need for storage virtualization, the second key component in optimizing. Storage virtualization results from consolidating two or more heterogeneous storage systems into a single consolidated pool of storage.

Thin client provisioning allocates the minimum (instead of the maximum) storage needed by an application or user. It dynamically allocates storage on the assumption that all applications or users are unlikely to call on all available capacity simultaneously.

Data deduplication helps reduce storage requirements, whether the storage is virtual or fully physical. Deduplication tools analyze data, searching out and eliminating repeat instances of information among users and databases. Vendors' specific techniques vary, but broadly speaking, they fall into two categories:

- **Inline deduplication:** This technique applies the deduping algorithms as data blocks enter a device and prevents second instances of data from being committed to storage.
- **Post-process deduplication:** This approach is implemented after data has been stored, rooting out and eliminating occurrences of identical data.

Each dedupe approach has advantages and disadvantages in terms of system speed and primary storage requirements. But either way, deduped data requires lower overall disk capacity for an organization, which reduces costs and speeds processing and throughput.

Additionally, to lessen the storage load on the infrastructure, a growing number of organizations are moving to offsite cloud services for backup storage. Depending on the terms, this can cost less than having internal backup because an organization can pay as its storage requirements grow rather than paying up front for anticipated capacity.

Data Center in a Box

IT planners have long pursued lights-out data centers — locked, darkened rooms of servers that operate automatically and require no service from administrators. That goal is not entirely achievable, but organizations can build data centers that are nearly hands-free.

The data-center-in-a-box approach embodies cloud principles, viewing the data center not so much as a system but rather as a service provider. Another term for this approach is "converged infrastructure."

Converged (or "turnkey") data centers integrate all the elements of the facility in an internal cloud. Basically, the organization rolls this single data hub into the existing data center.

This approach yields an infrastructure-as-a-service model. An internal cloud condenses and virtualizes scattered physical parts, but lets the IT team maintain the necessary logical isolation among operational functions such as human resources, finance and personally identifiable user information.

HOW TO ACHIEVE WORKLOAD MOBILITY

Almost since the inception of current-day server virtualization, vendors have been working toward the ability to move virtual machines from physical server to physical server, across a network, without an interruption in operation. This process is known as hot migration, and it fulfills several needs:

- Frees up a given physical server for maintenance without downtime for users;
- Dynamically balances workloads among physical servers so that they all run at optimal levels;
- Prevents a facility's underperforming servers (or those on the brink of failure) from interrupting operations.

The technical hurdles involve moving virtual machines (VMs) among servers that don't share the same virtual LAN, memory and storage resources. First, the bandwidth required to move VMs over distances can be prohibitive. The advent of 10 Gigabit Ethernet, mechanisms for bit-mapping memory and shipping it along with the VM, and virtualizing the network so that IP and Media Access Control (MAC) addresses travel along with the VM have mitigated these problems.

Organizations are also employing storage virtualization combined with the *active-active* data center. As opposed to active-backup (also known as active-standby), constant production-mirroring supports not only planned motion of workloads but also allows for high-assurance failover in the event of a catastrophic failure at a site — for example, a fire or other physical damage to a cabinet.

A chief benefit of such a setup is a more predictable IT cost structure. An organization will be able to reduce the number of suppliers it has relationships with.

Although the component products tend to be dominated by the prime vendor's hardware, these offerings tend to involve multiple partner vendors. Ultimately, no single manufacturer monopolizes the parts. But the solutions integrate third-party products so that the organization doesn't have to take on that work.

The speed and simplicity of this approach allows for rapid provisioning of new users, departments, offices and projects. A fully optimized data center must offer the agility required to meet fast-changing operational requirements.

One promising aspect of cloud computing is instant provisioning. Data-center-in-a-box offerings do just that, using master operating systems that orchestrate computing resources and balance loads in a way that maintains policy-based service levels.

Here are four questions CIOs will want to consider when evaluating a turnkey data center solution:

1. **Is it brand agnostic?** Management tools vary greatly in capability. But differences among hardware tend to be narrower. So even a product set such as the HP CloudSystem Matrix, while using primarily HP hardware, can incorporate other manufacturers' blade servers and storage subsystems.
2. **Can it provide cybersecurity at the individual application and virtual machine layer?** For these highly virtualized environments, security at only the system firewall level is inadequate. Beyond the protection of data assets, assuring this level of security often proves essential in gaining acceptance by the facility's user departments, who will be concerned as much about insider threats as external security threats. The data center management must be able to demonstrate security of data in motion on the network and in the compute layer.
3. **What vendor support is provided?** Given the all-in-one nature of this approach, it's essential that 24x7 support for both hardware and software be part of any agreement with a vendor.

4. **Are there clear upgrade pathways?** As requirements change and technology advances, the "box" solution should accept updated or upgraded swap-outs so that the organization can take advantage of the latest processor or storage technologies, or even evolutionary changes in computing.

Reference Architectures

Organizations that decide to implement a build-your-own strategy may refer to data center reference architectures for guidance.

An architecture outlines the basic design of a data center optimized for greatest flexibility. Typically, published architectures help an organization integrate industry-standard software products, such as directory services, databases and collaboration tools. Then they guide the organization through the interconnections of the supporting hardware.

Published data center reference architectures tend to be multivendor but not always vendor-agnostic. This is because they are often the result of collaboration among complementary software manufacturers, or both software and hardware manufacturers.

Examples include the VMware Reference Architecture (tuned for Cisco networking and EMC storage products) and the FlexPod architecture (incorporating Cisco, VMware and NetApp products). Microsoft publishes what it calls the Infrastructure Planning and Design Guide series, which is centered on its own enterprise products.

The use of a reference architecture can potentially save dollars that an organization might have allocated to architecture planning, either internally or by consultants and integrators. Another benefit of applying an existing reference architecture? The IT team can safely presume that the vendors collaborating on the architecture tested the configurations they reference and have a high degree of confidence that the result will be parts that can plug and play.

VM Mobility

Server virtualization has vastly increased the utilization rates of hardware. But it can tax the physical data center infrastructure as the number of virtual machines grows into the thousands. Virtual LANs can accommodate a maximum of 4,094 virtual machines (VMs) — and many organizations create more than that.

In large organizations, moreover, it's likely that there will need to be separate applications serving different departments or operational areas — for reasons such as data privacy or financial compliance, for instance. These applications and data may be separated logically but running on common hardware. After all, separating virtual workloads physically defeats the advantages of virtualization.

VMs often need to be moved around, sometimes across multiple data centers, to achieve balanced server loads, ensure disaster recovery or maintain service levels. Solutions such as VMware vMotion address this need for real-time and automated VM mobility.

These types of tools take into account that, for fully agile VMs, the network infrastructure must be virtualization-aware. This ensures that any system addresses, quality-of-service parameters and access control lists (ACLs) associated with a particular VM move with it. Otherwise, VM mobility requires manual follow-up.

A recent development from Cisco known as Locator/ID Separation Protocol (LISP) decouples an IP identity from a physical location. This promises to ease the movement of VMs among data centers across the Layer 3 transport, without these VMs becoming "lost." ■

The Network
Servers
Storage
Power & Cooling

Hardware Optimization Essentials

Up-to-date technology can transform your data center and your organization.

Optimization requires maintaining an eye on all the physical components in the data center, both the network and the server hardware. Inventory, track, analyze, optimize — repeat. Providing top-tier services and support requires that the IT and management teams within the organization live this cycle.

The Network

Nothing happens in a data center without the network. As a result, a couple of elements are fundamental to any successful ongoing optimization strategy. The organization needs to carefully manage the backbone network services connecting offices and the data center (or centers), as well as network services supporting local storage and application networks. This is accomplished through a number of initiatives.

Strengthening Redundancy

Redundant network capabilities can optimize throughput when everything operates properly and prevent stoppages when a component in the network chain fails. Therefore, any optimization strategy must include an initiative for redundancy.

A basic component of redundancy calls for two network interface cards (NICs) at the physical layer of the server network. The principal virtualization products, such as VMware, provide a software-configurable, virtual dual-switch interface on top of the paired physical cards.

One approach is having a physical NIC assigned to each virtual server to prevent a hiccup with one VM from interfering with another on the same physical server. An alternative approach calls for connecting each of the NICs to their own switches to ensure every server can reach the backbone network in the event of a NIC or switch failure.

In both instances, a network or application administrator could also reach a problematic VM given the same circumstances.

For the wide area network and backbone connections between data centers or between an organization-owned data center and a cloud provider, optimization requires finding the best approach for maximizing uptime without the expense and complexity of creating identical copies of each and every element.

While a simple, nonredundant WAN might require less time to repair than a highly complex redundant one, it's also less likely to provide the high availability expected by users today.

It's worth considering a more fine-grained strategy for WAN redundancy: Assess the average failure times of components and then provide redundancy for only those pieces likely to fail often. Typically, those would be power supplies in routers and switches, as well as power supplies or disk drives in servers.

Building in Resiliency

Network resiliency, though related to redundancy, is not the same thing. Redundancy is an avenue to resiliency — resiliency being the ability to respond to and overcome interruptions regardless of cause. While it may be counter-intuitive, a totally redundant network may not be the most resilient.

Resiliency stems from network elements that can detect one another and give fast failover as needed. To achieve resiliency, a network must embody the least practical complexity and remove all single points of failure (SPOFs).

Network design, more than any particular product, yields resiliency. An organization likely won't need the same degree of uptime across every segment of the network. The enterprise backbone might require uptime of five nines, or 99.999 percent. But a branch office LAN or connections to the backbone might meet users' needs sufficiently at a lesser uptime.

Ensuring resiliency often depends on a combination of the right protocols coupled with the right level of hardware redundancy. For example, the Virtual Router Redundancy (VRRP) or Spanning Tree (STP) protocols can support network-level recovery, but the switches and routers must house redundant or hot-swappable components.

Looking at resiliency from an enterprise services perspective (and taking cloud provisioning, virtualization and mobility into account) will allow the organization to define its data center dependence. The IT team can then evaluate the carriers and utilities entering the data center, the mean time before failure of network components, the power backup plan, and the design of the network and network services. All of these play a role in optimizing for resiliency.

Maximizing Flexibility

An agile or flexible network is one designed to automatically respond to variable demands. Sources of that demand include the creation of new virtual machines, the addition and then configuration of hardware, shifts in application workload and growth in users.

Organizations sometimes need flexibility in the form of speed, such as when provisioning new users or devices. At other times, flexibility might be tied to capacity and scale changes, such as a large file transfer among technical or scientific staff or a seasonal processing load.

Virtualization again has proven a useful technology for achieving flexibility. For example, in disaster recovery applications, conventional servers typically sit ready but unavailable until needed. By contrast, virtualized servers are easily replicated to backup sites, or the IT group can stage them there for testing.

Either way, VMs enable a backup site to be used continuously, both for load-balancing production processing and for handling any development or testing requirements. In this scenario, an organization experiences near-zero wind-up or transfer latency when its backup site comes online.

Organizations also find that taking a service-oriented architecture (SOA) approach to the design of data centers, whether conventional servers or virtualized, enhances flexibility.

Because SOA separates code for operations process applications from operating systems and other support functions, processes can run on any of several servers. In some ways, the cloud computing movement has rejuvenated the SOA movement in that it frees up software objects to entire applications from hardware stovepipes.

Flexibility benefits any function or operational process that IT supports, plus it drives down infrastructure costs while increasing confidence in the IT shop.

Embracing 10 Gigabit Ethernet

It's easy to visualize how a bullet train can't realize its potential on tracks designed for a steam engine. It's the same with virtualized servers and high-performance storage subsystems, which need fast networking fabric to realize the potential of an optimized data center.

Although the 10 Gig-E standard was established in 2002, adoption has quickened over the past three years, driven in part by the growth in virtualization. In fact, in 2010, the IEEE ratified the 802.3ba specification for 100 Gigabit Ethernet, and Terabit Ethernet is on the horizon.

Shipments of 10 Gig-E adapters now number in the millions per year. This ramp-up, combined with the advent of faster standards, has led to a drop in pricing comparable to Gigabit Ethernet just a few years ago.

Most vital: 10 Gigabit Ethernet can affordably support comprehensive optimization because it is compatible with an installed base of literally millions of switches.

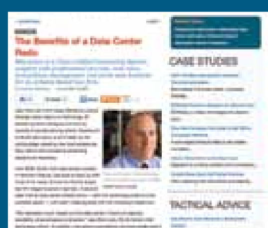
Simply upgrading to 10 Gig-E generally speeds up application performance. Plus, it eases the organization's move to a unified multiprotocol network strategy encompassing WAN, LAN and storage I/O components.

CASE STUDY:

The Benefits of a Data Center Redo

Read about how an Atlanta-based law firm approached its data center upgrade at:

CDW.com/dccasestudy



In fact, moving to speedier Ethernet can help uncover bottlenecks that lower bandwidth service may have masked, thereby identifying potential application and configuration optimization targets.

Interlacing Application and WAN Optimization

The dual personality of today's organizations is most apparent in how extensive WAN services have become. As data center resources have grown more centralized, mobility has allowed users to become more widely dispersed.

In addition, the emergence of Voice over IP (VoIP) and video have added traffic that competes with enterprise applications for WAN throughput. The challenge for IT teams lies in providing LAN-like response over the backbone network despite the wide diversity of applications it carries.

Optimizing the WAN, therefore, requires addressing the network and applications as separate but interrelated activities.

Modern applications mainly use Common Internet File System (CIFS), Messaging Application Programming Interface (MAPI) or Hypertext Transfer Protocol (HTTP). Two techniques can lessen the effect that these "chatty" protocols have on WAN response times.

The first, byte caching, maintains common packet elements at the endpoints, which reduces network traffic. The second, protocol optimization, slowly but surely converts serial traffic to parallel traffic as it "memorizes" typical user requests and prepositions likely required data.

Additional approaches to speeding WAN traffic include object caching (which works like byte caching but in a more application-specific way), file compression and endpoint compression.

Still another approach involves detecting and then blocking or limiting the bandwidth of traffic not related to the organization's function or operations. For example, iTunes, Skype and many shopping and gaming applications use the same ports as enterprise traffic. Some of them hop ports to avoid filters guarding commonly used entryways, such as ports 80 and 443.

Developing "Any Device, Anywhere" Access

A popular CEO once famously remarked that his company's world headquarters was wherever he happened to be at a given moment. Today, an organization's workplace is wherever its staff happen to be, which could be at home, on organization-related travel, in the field or in a facility designed for staff to use temporarily. In other words, the resources of the data center and WAN must deliver applications and data anywhere.

Moreover, consumer technologies in the form of smartphones and tablet devices are moving into the enterprise. In fact, many CIOs are adopting bring-your-own-device (BYOD) programs. That approach can result in lower costs for the organization and greater flexibility for users.

It's also driving client virtualization, an added requirement for data centers when application servers are already virtualized. Today's approach to user mobility typically follows a rather specific model: A user's applications and related files exist as virtual machines accessible wirelessly from OS-X, iOS, Android and Windows devices. In return for this accessibility, users agree that the IT staff may wipe the device remotely the moment it is lost, stolen or otherwise misused.

What this model implies is that an organization's data center has the capability for rapid provisioning (including IP address assignment), monitoring (including location services) and digital destruction from a central console.

Using Network Extensions

Extensions build on existing protocols to overcome any limitations tied to those protocols. Current efforts focus mainly on Ethernet. Engineers see it as the basis for a kind of universal fabric for data transport within and among data centers. The problem has been that the best-effort approach in Ethernet can result in lost packets.

Tunneling Fibre Channel frames over Ethernet MAC frames is one way to get lossless transmission at the speed of Ethernet. But data center bridging (DCB) has emerged as a set of extensions to Ethernet that maintain speed but overcome packet losses during congested periods.

The DCB extensions use several technologies based on a subset of the IEEE 802.1 standards. Manufacturers have incorporated DCB into switches. The individual technologies are complex, but their benefit is clear: Enabling a converged network in which a single wiring infrastructure can carry all applications and protocols.

Other network extensions help the local benefits of virtualization extend across the enterprise. Within a local domain, VMs routinely move among servers to improve workload balance and avoid server failures. But moving workloads among data centers across Layer-3 networks has proven a bigger challenge, in part because of distance and in part because MAC addresses get lost along the way.

New products exist to overcome this limitation. One example, Cisco Overlay Transport Virtualization, is a switch enhancement. It maintains MAC addresses for every device on the domain and encapsulates workloads so they move as frames over the IP WAN from data center to data center. It works in conjunction with LISP technology *(see Chapter 2)*.

Servers

Servers are the heart of the data center, so optimizing them can quickly yield the most benefits. Servers continue to exhibit the classic principle of Moore's Law: Productivity doubles every 18 to 24 months for a given dollar of investment — mainly because of advances in chip and bus technology.

CIOs and data center managers don't usually think in terms of what's inside a microprocessor. But that's where advances in data center capability ultimately start. As manufacturers continue to shrink transistors and silicon interconnects, microprocessors and in turn servers (along with every other kind of digital device) continue to gain capacity. For example, in late 2011 Intel began delivering chips using a 22-nanometer process.

Such advances are more than technical curiosities. Chip changes lead to more server power per dollar and less power consumption per computing unit. In a large data center, that adds up to legitimate savings that can be reinvested in the infrastructure.

Server selection comes down to two basic items: form factor and processor. Because every environment has its own unique factors, organizations upgrading their data centers should ask themselves several questions before making a server purchase.

Is the organization static or growing — in either data or users? Is it about to embark on a virtualization effort, or has it already virtualized? Is it consolidating, or adding, data centers? The answers to these questions will affect both the choice of form factor and processors.

Current server technologies encompass form factors that range from floor-standing towers to half-width, 1U blades. Racks may have one, two or several processor sockets, each equipped with single, dual, and even four- or eight-core processors.

The processor choice depends on whether applications have been optimized for multiprocessor environments, application licensing arrangements and operating systems (based on the number of processors or cores within a single chip), and cooling and power considerations.

Aside from computer power itself, up-to-date servers support so-called green computing or power efficiency, fine-grained manageability and virtualization.

In theory, any server can support virtualization by definition because VMs interact with the virtualization package's hypervisor and not the hardware. But in reality, the processor, memory, networking and storage interact in subtly different ways for otherwise identically equipped servers.

A new benchmark known as SPECvirt_sc2010 gives comparable test results for servers by simulating workloads for multiple VMs. The latest version of the benchmark came out in May 2011.

Increasingly, server strategy must take into account the possibility of client (or user state) virtualization. As the next wave in virtualization, client virtualization is driven by several factors.

First, the BYOD movement increasingly finds organizations accommodating popular consumer platforms, principally iOS and Android smartphones and tablets.

Next, anywhere, anytime access is leading many organizations to specifically identify reducing demand for office real estate as part of their broader consolidation strategies. Even in "hoteling" scenarios, in which workers come in from the field temporarily, staff prefer to use their own devices, relying on ubiquitous Wi-Fi for access to enterprise computing resources.

Finally, there are easy cost savings to be had from migrating to thin clients. Compared to full PCs, thin clients offer lower total cost of ownership because they're less pricey, boast longer lives and use less power.

In short, a best-practices approach to servers hinges on the organization's upfront planning — taking into account anticipated workloads and devising the most efficient approach for delivering sufficient computing power and scalability.

Configuration, cabling and consolidation concerns all play a role in strategic server provisioning.

Configuration Factors

Servers are definitely hardware, but they are essentially useless husks without software. At least 100 variable attributes of any given server exist as software definitions.

So when it comes to replacing or adding servers, the IT shop can expect to spend considerable time ensuring their proper configuration. Often in large enterprises, the settings don't exist in one place. Some may reside on network switches or external storage devices, even other servers.

Stateless computing can remedy this by gathering all of the variables that together form a server's identity and saving them as a package that systems administrators can assign to any physical machine. A stateless server is one with no permanent configurable settings.

The essential values that form the profile capable of moving from one machine to another include MAC addresses and World Wide Name (WWN) or World Wide Identifier (WWI) values, universally unique identifiers (UUID), firmware and basic input-output systems (BIOS).

Stateless servers reduce costs by lowering the administrative overhead of setting and resetting configurations. In stateless environments, the IT team handles server configurations remotely from an administrative console.

Reduced Cabling

A rat's nest of wiring has no place in an optimized data center. Often legacy, point-to-point cabling is overlooked when an organization replaces computing machinery.

Given the agility and flexibility that organizations seek, why maintain hard-to-remove, dedicated cables? But before optimizing cabling, an organization needs a thorough understanding of its existing wiring, and it must document the existing cable plant before revamping it.

Two sets of considerations guide cable planning: protocols and the types of cables, and topology, which dictates the most efficient physical layout.

On the protocol front, Fibre Channel running on copper Ethernet wiring is one technique for reducing cabling costs. This approach puts SAN connections on a common fabric with the transport network.

Copper cable for 10 Gig-E is thicker than optical cable and thicker than copper used for 1 Gig-E. Cost comparisons have grown more complicated because of rising prices for metal and falling prices for fiber. Plus, cabling is only one component in the data movement chain, which also includes the hardware interfaces for the copper and fiber.

For 10 Gig-E and beyond, data centers likely still need a combination of fiber and copper. But multiprotocol fabric technologies can reduce overall cabling, reduce costs and speed throughput.

On the topology front, planning should take into account the size and shape of the data center and the building in which it resides. Best practices dictate a series of wiring consolidation points that eliminate point-to-point wiring, reduce required replacement cable inventories, and enhance labeling and traceability of cabling.

A structured and carefully planned approach to wiring will also help reduce electricity demands. By routing cabling on top of cabinets or in overhead raceways, it's possible to remove it from the under-floor plenum, which will increase heat exchange efficiency.

Server Consolidation

Collapsing systems into a smaller footprint can occur within the data center as well as across multiple data centers. The obvious benefits are lower acquisition costs, real estate requirements and power consumption.

Today, server virtualization almost always accompanies consolidation. Even so, there's more to server consolidation than buying a rack of blades and then seeing how many virtual machines the IT team can jam into that rack.

Determining the best consolidation ratio for an organization requires a calculus of several factors: user patterns; service-level agreements (SLAs) with units served by the data center; anticipated growth in databases and network traffic; and status of applications (those that should remain on their own servers versus those that should be virtualized).

It's also critical to consider the maximum practical number of production VMs per server that the organization can sustain. Then it also must determine the same information for its test and development environments.

Once the organization has decided which apps are virtualization candidates, it must then determine how to group them as VMs on the physical servers. Some systems engineers advise mixing heavy- and light-load applications on servers to get efficient utilization. The thinking is that this approach avoids

WHITE PAPER:

Data Storage Optimization
Learn more about what options are available for optimizing storage at:

CDW.com/dcwhitepaper

taxing a server with too many processing-intensive apps or wasting resources by amassing too many lightweight systems.

Keep in mind that an optimized data center must avoid being over-consolidated. In some ways, virtualization tools are almost too successful. In unskilled hands or where the organization lacks a strong framework for data center operations, VMs can be created basically at the touch of a button.

Although virtualization can address the problems of server sprawl and grossly low, inefficient load rates, unwisely planned VMs can create their own sprawl and overburden an organization's consolidated servers.

The upper limits of utilization can quickly usurp shared memory and harm application performance and stability. Failover mechanisms on a crashed server can tax network resources and cause cascading complications. Optimization means finding the right balance between maximum utilization and maximum reliability and manageability.

Storage

Organizations that optimize servers and networks but continue on with the same data storage practices won't achieve fully optimized data centers. Absolute storage needs consistently rise faster than unit storage prices fall. Growing numbers of physical storage subsystems, disk volumes and logical unit numbers (LUNs) increase both the complexity and the cost of managing the storage environment.

Storage optimization therefore has three goals: reducing storage costs, improving storage performance and simplifying storage management.

Capital spending requests, including those from IT, now travel far up the senior management chain as organizations try to rationalize or cut costs. The IT shop can no longer justify rubber-stamping storage growth of more than 50 percent a year. The following four best practices will help optimize storage in virtual environments.

Virtualize

Storage virtualization is a fast-moving technology. Manufacturers have solved the long-standing technical challenges of separating the physical location of stored data from its logical location.

Virtualization abstracts the functions of the storage systems and isolates them from applications and host servers. That in turn enables the pooling of disparate storage resources on consolidated physical drives.

Or, for faster performance, virtualization may disaggregate data across more drives. Properly configured, storage virtualization provides higher utilization rates for storage resources while helping the organization reduce or avoid storage sprawl. The third benefit? Easier management of the storage infrastructure.

Achieving storage virtualization involves making several decisions up front, some highly technical in nature. That's why it's particularly important to keep the intended benefits in mind when evaluating technologies and products.

Functionally, virtualization requires a software layer sitting between application I/O calls and data storage. This layer directs calls to the physical drives containing the data.

In practice, numerous paths exist to storage virtualization: at the network-attached storage (NAS) or SAN levels, at the storage device level, at the file level and at the host level. Some virtualization tools come integrated with storage systems, others work from dedicated switches.

An organization must also decide between in-band and out-of-band virtualization. In-band (or symmetric) virtualization occurs between the host server and the storage systems. It's easy to deploy but can create performance bottlenecks. Out-of-band virtualization occurs beyond the host-storage data path but requires extra host software. It is the more complex scheme, but it can scale more easily.

Another important consideration is the compatibility of storage subsystem brands with the chosen virtualization product. Universal interoperability among storage manufacturers doesn't exist yet, but it is possible to map data and move workloads across the disparate systems.

The server virtualization market is centered around a "big three" set of competing products, but storage virtualization is far more diverse. Nonetheless, storage virtualization can ultimately optimize a complex and expensive portion of the data center environment.

Its financial benefits include lower capital costs and more efficient use of data center administrative staff. Its technical benefits include ease of mobility for the organization's data blocks, which in turn enhances replication, backup and disaster recovery.

Data Dedup

Data deduplication reduces physical storage requirements by eliminating redundant instances of data even though the files or blocks in question have two or more references in the system. Traditionally, administrators have applied deduplication to backup and archive files, but deduping production or primary storage is also worthwhile.

Deduplication works in one of two ways, inline or post-process. Inline deduping blocks of data as they are written to disk prevents multiple copies from the get-go. It's worth noting that with applications calling to a structured database, inline dedupe algorithms can extract a performance penalty. But it doesn't require extra storage to write duplicative blocks of data for later removal.

Post-process deduping's pros and cons are the opposite of the inline approach: There is no application penalty, but post-process deduping requires extra storage. Applying post-process deduplication to unstructured data, such as Microsoft Office documents residing on anywhere from two to 2,000 users' disk partitions, is a good place to start because such data tends to stay put, using up disk space.

Backup data dedupe is still an important optimizing strategy. It lessens storage needs and allows shorter recovery time objectives (RTOs) because deduped data can be quickly backed up. Nearly all backup software applications have built-in dedupe functions, but they don't all work equally.

For agile, virtualized data centers, strive for global deduplication, which looks at data across storage systems and domains and results in thorough deduplication. This can be a performance advantage in environments where backup data centers run live to support workload balancing.

Storage Consolidation and Replication

Consolidation and replication best practices will allow an organization to avoid expensive, brute-force techniques for overcoming bottlenecks and capacity limitations. The goal is to achieve more-efficient storage — essentially higher utilization rates of existing capacity.

Storage consolidation begins with rigorous data tiering, storing data within a hierarchy of media depending on frequency of access. At the top tier are relatively expensive but very fast solid-state drives, followed by primary disks, secondary disks and finally archival media (optical or tape).

Tiering delivers two primary benefits. First, it improves the performance of applications interacting with live data and rationalizes storage costs by relegating as much data as possible to the lowest-cost medium. Tiering also makes better use of the bandwidth by which data is returned to applications.

The proper tiering scheme depends on defining the required access times (anywhere from milliseconds to days), the recovery point objective (RPO) and the RTO in the case of backup or replication, and cost.

Storage Management

Storage management generally means using a comprehensive tool to give data center managers a look into the performance and utilization rates of their facilities' storage systems.

DR-READY:
UP AND RUNNING, NO MATTER WHAT

When optimizing server, network, storage and supporting infrastructures, the result is a data center that users can count on no matter what. A sweeping program of optimization will ultimately result in enhanced disaster recovery and business continuity.

In the case of disaster recovery, the mobility of virtual machines, not only within a data center but also among multiple data centers, changes the notion of standby backup. Virtual resources, because of their mobility, turn backup sites into both backup and live production sites.

That gives the organization increased flexibility and supports easier recoverability. No matter what befalls the organization's IT environment — whether it's a single corrupted VM or a major electrical outage — a virtual and optimized environment will allow a near-instantaneous recovery.

Because mission delivery has become inseparable from IT services, a loss of data and application services can cause operations to grind to a halt. Like silence on a radio station, even partial loss of online presence can quickly translate into lost business, harmed reputation and mission failure. In the end, avoiding such an occurrence makes data center optimization worthwhile.

These tools typically operate from a web management console. They poll the data center (or centers) to maintain an inventory of storage assets, and gather information about how these assets are configured and the degree to which they are used by servers, applications and virtual machines.

Manageability extends to planning for and provisioning storage by making best use of a data center's particular topology and SLAs. A full-range tool should simplify installing, configuring, tuning and upgrading storage by distilling these tasks on a single control screen.

Power & Cooling

Many of the tools driving data center modernization are also thoroughly transforming cooling requirements.

Two converging trends raise the importance of good power and cooling practices for data centers.

Growing national awareness and concern about electricity demand and generation has caused major power consumers to explore ways to conserve. One new scholarly estimate states that data centers in 2010 consumed as much as 2.2 percent of all power in the United States — and data center power consumption is up 56 percent since 2005.

Yet the total consumption figure is lower than estimates made midway through the past decade. Those earlier predictions didn't anticipate two factors: the economic slowdown and the decrease in server deployment due to virtualization.

Still, IT departments face continuous pressure to trim costs, and power is an easy target. These trends have propelled the concept of the green data center. Of course, there is no hard and fast definition of "green." People generally agree it means using less power per computing unit, using it more efficiently and operating in a continuous power improvement mode.

Reducing Power Consumption

A surefire way to reduce power consumption is to replace aging servers. The federal government's Energy Star program continuously lowers power allowances for machines to achieve its Energy Star rating.

For example, previously a 1U rack server could get by using 100 watts while idling. Under today's Energy Star guidelines, it must draw only 55 watts in idle. Therefore, replacing servers with newer models using about half the power would quickly pay back the capital costs and then produce cost avoidance for the life of the servers.

Another fairly simple way to bring some green to the data center is to raise the maximum temperature setting. The old rule of a chilly 64 degrees no longer applies because new computer cooling techniques concentrate cooling precisely where it's needed, rather than keeping the entire facility unnecessarily cool.

Consider moving some disk storage to solid-state drives because they use less power natively and are easier to cool than rotational drives.

Closer management of the infrastructure can also yield power savings with the following actions:

- Weed out unused or poorly utilized servers and other equipment.
- Concentrate low-wattage lighting where it's precisely needed and reduce ambient lighting.
- Replace lead-acid batteries as interim backup power between the time of a utility failure and when the uninterruptible power supply (UPS) kicks in. Fuel cells, flywheel generators and capacitor arrays might provide useful, if limited, alternatives.

Data Center Cooling

Data center cooling requires a more detailed approach with the evolution of virtualization and the expanded use of rack-mount and blade servers. These formats create differently shaped and more concentrated heat patterns than stand-alone servers.

The best practice for cooling these rack environments starts with positioning them in such a way that the hot sides of adjacent racks face toward one another to create hot and cold aisles. Then, it's possible to direct the output of computer room air-conditioners (CRACs) into the hot aisles while keeping the ambient temperature relatively high.

Another technique involves isolating the hot aisles and then using tenting arrangements and a combination of convection and blowers to exchange cool and warm air. This second approach may involve a dropped ceiling or subfloor plenum area.

In some parts of the country, data center heat exchange can also benefit from the use of a Kyoto wheel, which is a large, slow-turning fan that works as a heat exchanger. It's an old technology enhanced with new materials. One drawback is the large footprint this apparatus requires.

Data center operators also have begun using special tanks of inert coolant into which they immerse arrays of rack or blade servers, but that requires removal of fans and optical drives.

While this is a technology area that is still evolving, clearly every effort at greening can incrementally drive up optimization and aid in cost avoidance. ■

Virtualization Basics
Orchestration and Automation Tools
Infrastructure Management Tools

Software Essentials

Strategic software deployments are key to data center optimization.

Virtualization in many ways is integral to data center optimization. For decades, IT had been exploding into a sprawl of decentralized infrastructure, brought about by the advent of the minicomputer and later the PC.

Today, virtualization has matured to the point where it can support personal user IT "accounts," productive devices and mobility. It also returns some measure of control over server proliferation, data center real estate growth and their concomitant costs to the organization.

Virtualization Basics

Virtualization, or the abstraction of software and supporting hardware resources into logical units (virtual machines) that share hardware, dates to the earliest days of computing. It hearkens back in some ways to the time-sharing approach to computing in the 1970s. This focus on sharing hardware resources was eclipsed by the advent of inexpensive x86 hardware, which masked the inherent inefficiency of this one-application-per-server model.

Modern virtualization tools instantly create VMs. They rely on a layer of software called the hypervisor that resides between the server hardware, or "bare metal," and the VMs (which contain their own operating systems). The hypervisor manages the interaction between VMs, the network and the storage back end.

Although a technologically interesting solution, virtualization has taken hold because of the benefits it can deliver. It ensures maximum utilization of servers, which reduces hardware proliferation.

And because VMs can move easily across networks to other servers, virtualization increases availability of IT services by automating server failover and allowing it to occur in near real time. Because less equipment is needed, maintenance is reduced and acquisition dollars can be dedicated to innovation or improving the organization's bottom line.

Orchestration and Automation Tools

The latest wave of software developed to take advantage of virtualization, orchestration and automation tools speed up the provisioning of IT services for new applications, business or mission services, or workgroups. In recent years, the web has fostered a self-service model for actions that traditionally required help desks, service tickets, phone centers and similar analog-era processes.

Today's virtualized data centers are similar to internal cloud service providers. Commercial cloud providers deploy tools to let customers, in essence, set up an account and click for the services they want. Services might range from storage and a few tools to manage a software development project to full-blown data center replication services.

Many organizations, having virtualized and otherwise optimized their data centers, are looking at orchestration and automation tools for internal customers and to speed up tasks for the data center staff itself.

Automation implies a single task, whereas orchestration automates a series of tasks and results in a suite of integrated services. The objective is to reduce the manual work required to provide IT services. Orchestration manages the provisioning of servers, storage and networks, plus associated services such as cybersecurity, bug tracking (during software development) and other productivity tools.

The major virtualization vendors, including VMware, Microsoft and Citrix, offer orchestration tools, as do several third-party vendors. Although these programs function primarily as workflow automators, they are hyperaware of data center resources, employing existing capacity first and helping the organization avoid premature or unnecessary purchasing of new servers or storage capacity.

Infrastructure Management Tools

To gain a clear view into data center assets and how they connect and interact with one another, many IT shops are turning to data center infrastructure management (DCIM) tools. Deployed properly, they support the management of current processing loads, deployment and provisioning of new capacity, and information needed for planning.

These tools are only beginning to penetrate data center operations. But analysts expect fast adoption over the next few years. Driving this adoption is the increasing density of assets and the ensuing consequent complexity.

Data center consolidation, virtualization and the provisioning of cloud services within the data center have, ironically, increased the intricacy and diversity of equipment in data centers even as they have reduced the overall number of servers and facilities for many organizations.

DCIM tools incorporate asset discovery and tracking, data collection, and report generators. They enable the automation of repetitive functions such as the provisioning of new gear and other operational tasks. These tools monitor operational health, measuring parameters such as capacity utilization and energy consumption of data center components. And they can provide predictive analysis of the resources, including staff time, necessary to complete future tasks.

DCIM tools can improve data center management by integrating the roles of an organization's facilities, network and systems staff. Plus, they can aid in the design of specific processes and then the enforcement of associated policies and use rules. It's helpful to evaluate several tools closely before selecting a program for the organization's IT team.

Some tools have originated in the power monitoring or facilities management domains, others in the IT domain. The original orientation might remain after they integrate other functions, and they might not appeal to all of an organization's functional managers. Additionally, a given DCIM tool might have the power to control some components of the data center, such as those related to utilities, but not the technology itself. ■

DESKTOP: THE NEXT “AS A SERVICE” OFFERING

If client virtualization is the next logical wave of development following server and storage virtualization, can the self-service Desktop as a Service (DaaS) be far behind?

Large organizations tend to use standard user images within a given workgroup, office or operations function. Virtualizing users' desktops, whether or not they are standardized, brings the benefits of faster provisioning and updating — and ultimately lower management costs.

A growing number of organizations are letting staff bring their own devices to work. A few are even requiring it, which greatly diminishes hardware costs but can increase administration and cybersecurity exposure from the data center point of view.

One approach to these issues could be self-provisioning by users for applications and services they might need or want beyond an organization's standard image or set of images. By keeping these additional resources within the data center, the IT department would ensure that software remains up to date and could monitor who downloads what.

But moving in this direction would require chargeback mechanisms for resources that don't carry an unlimited site license. Procurement and buying mechanisms, rather than technology, are other potential hurdles.

Management Processes
Management Tools
Capacity Planning
Chargeback
Vendor Management

Management and Process Improvements

Updating data center practices and skills

In many ways, the data center is like an engine. It has a complex infrastructure, and it can be tinkered with endlessly. But an engine running on a test block offers only potential. Once under the hood of a car, there are challenges to running it efficiently.

Similarly, an optimized data center is a resource of untapped possibility unless the organization also optimizes the framework within which the facility operates. This requires the right governance and operating processes supported by industry best practices.

Many IT organizations have successfully adopted the framework embodied in the IT Infrastructure Library. ITIL gathers best practices aimed at ensuring that the big-picture operational needs of the organization are fully supported by its system. ITIL literature describes not only the processes themselves but also the competencies, skills and metrics necessary for the organization to be able to measure progress.

Management Processes

Among the core IT processes (whether you work with ITIL or not) are incident, problem and change management.

Incident management: This management focus includes all processes and activities related to responding to one-time events that can disrupt service in some way. Basically, it falls to the IT team to "scramble the jets" to get things up and running again. Some incidents will likely be known only to the operational staff — for example, a virtual machine or a communication link fails — because the affected system immediately cuts over to a backup.

That type of incident might interrupt service or violate a service level agreement. But because it invokes backup, it's critical to resolve it and restore redundancy. (Note: Organizational resilience is needed to maintain redundancy for the data center and reliability for users.)

TACTICAL ADVICE:
Capacity Planning

Discover tips on determining future capacity needs in the data center at:

CDW.com/dccapacity

Problem management: This management focus can help address recurring issues that might not be easy to identify. Problems can lead to incidents, but fixing an incident might not necessarily solve the problem. For that reason, managing problems often requires a different approach, with the staff performing deep diagnoses of the data center infrastructure to see what might be causing disruptions, slowdowns or error messages.

It's essential to find root causes and avoid workarounds. A thorough approach to problem management entails gathering data that can provide anticipatory information about possible problems on the horizon, much as disk drives self-diagnose and give warnings of impending failure.

Change management: This management focus ensures that any alterations to the data center occur in a consistent, documented way. A change management strategy should not only avoid interruptions in service but also allow for any changes to be made as efficiently as possible.

Management Tools

Given the complexity of today's data centers, maintaining optimal conditions requires the aid of tools. No IT staff can manage incidents, problems and changes with a clipboard and pencil. Numerous software makers offer suites of help desk, service desk and change management automation packages. These tools often incorporate ITIL best practices for IT service management.

The tools available range from highly niched products that might cover only one or two functions to fully integrated suites offered either as customer-installed products or online through software as a service (SaaS) programs.

As a class, the integrated toolsets automate many functions, such as identifying the IT group responsible for responding to particular repair, change or configuration requests. These programs typically include reporting capabilities that can monitor the time taken to resolve issues, identify recurring issues in the data center and survey users on their level of satisfaction with service offerings.

Over time, audit reports about incidents, problems and change management can provide information to aid in design and capacity planning. For that reason, these tools contribute to ongoing data center optimization. Plus, they also help keep the data center aligned closely with users' and organizational needs.

Capacity Planning

Capacity planning is the process of ensuring that the data center can meet future service-level agreements and the operational requirements of the organization. Such planning eventually leads to capital investments in servers, storage, networking and software, but it starts with capturing knowledge about workload and identifying anticipated workload.

For private sector organizations, this requires being alert to marketing and sales plans, product development and personnel trends. For all organizations, capacity planning requires knowledge unique to the IT function, such as vendor and technology trends and new products.

The most critical phase of capacity planning for the IT shop involves translating the day-to-day work of the organization, the service levels required by various departments and any legal or regulatory compliance requirements into computer quantifiable workloads and storage needs.

The wide range of demands on a data center tend to complicate capacity planning. For example, the engineering department uses different applications at different times of the day than the finance group, which might close and transfer large data sets overnight. For this reason, capacity use rates provide a rough measure of workloads over time.

But the data center staff will need much more detailed views into the nature of the work — the applications and processes running, to what extent and at what time. This "workload characterization" allows for the creation of fine-grained future-capacity estimates.

But even a closely calculated characterization won't give a data center all the information it needs unless the IT team can complete the circle by translating server and other resource use rates back into units of real work. "Getting 10 users up and running with 'x' applications in a new remote office" is an example of a unit of work.

For capacity planning to improve the data center in a budget-rational way, it helps to deploy capacity monitoring tools. Virtualized environments, in which individual servers likely run at high utilization rates, especially need close monitoring.

SKILL SETS FOR THE OPTIMIZED DATA CENTER

Workforce trends in the second decade of the 21st century don't stop at the door to the data center. After all, the application-oriented, social media–devouring mobile generation entering the marketing, finance, product development, education and public service areas are coming to work for the CIO.

Users are calling on service-oriented data centers for fast development and deployment of apps, creative use of organizational data and support for work trends such as mobility and collaboration. So it stands to reason that a workforce in tune with those requirements will help optimize the data center around it.

The optimized data center uses virtualization, green techniques and the latest storage and communications technologies to achieve agility and resilience — all driven by an efficient approach. Whether totally provided in-house or through a cloud service, the effect is the same: the provision of IT services as a utility, configured and provisioned quickly for fast-moving organizations.

A transformed workforce will also increase the prestige and clout of IT within the greater organization. Recent surveys show that although corporate managers overwhelmingly (87 percent in one study) believe IT innovation is a key to success, a large percentage (54 percent) don't view the IT department itself as a source of innovative ideas.

To change that perception, the IT workforce will need to shift its skills and adjust its focus. Engineers trained extensively in mainframe and client-server environments will be in less demand. Instead, the optimized data center of the near future will need:

- Enterprise architects who understand both flexibility and scalability;
- Agile developers who can work with user groups to craft iterative but speedy deployments;
- In-house consultants who can translate user requirements into IT services;
- Cybersecurity experts who understand insider and behavioral-based threats;
- Systems engineers with expertise in the core technologies of the optimized data center: virtualization, wireless, high-speed networking, and scalable and flexible storage.

Storage capacity planning brings challenges of its own — again, especially for virtualized and cloud environments where storage itself is often virtualized too. The inefficiencies caused by a sprawl of data storage pools disappear. But the IT team still needs visibility into how VMs use storage and how available capacity shrinks over time. Otherwise, there's a danger of overcommitting virtualized storage.

Chargeback

Chargeback is the mechanism through which data centers recover costs for providing IT services. The fees that facilities charge users depend on workload capacity. Once again, automation can help track customer usage data.

Many organizations average their costs across departments and charge flat fees. Charging by the percentage of VMs used can cause wide fluctuations in billing, particularly if VMs are provisioned and decommissioned frequently. Charging by resource consumption (on the theory that not all VMs are equally resource-intensive) requires a lot of work, although there are tools that partially automate this approach.

Flat billing is the least data- and work-intensive approach for the IT group. But keep in mind that servers represent only a portion of data center costs. Storage, telecommunications services, power and cooling, and personnel costs all should be factored into the chargeback equation.

Regardless of the chargeback approach used, the function itself can aid data center optimization. Collecting usage data for billing provides detailed visibility into costs, utilization rates and capacity overhead. The information can be used to assure high capacity utilization rates and to inform budgeting.

Calculating Return on Investment

Detailed cost knowledge is also necessary for calculating the facility's return on investment. The financial ROI for the IT department takes the form of a payback period; that's the net present value or internal rate of return for cost-cutting investments.

But the intangible benefits of IT service are equally important. Improved internal or external customer/user service, the ability to roll out new services, and better supply-chain management (or equivalent) are but a few of the common intangibles that flow from an optimized IT environment.

Still harder to calculate are investments that prevent problems such as those related to cybersecurity or service disruptions from failed components. Such investments might be considered absolute requirements by the IT staff, but as "extra expenses" by the organization's management.

That's why gathering and reporting information from incident and problem management reports can help justify investments in data center optimization. The CIO may need to highlight the potential risks of forgoing backup, redundancy and resiliency in order to help secure budget for these initatives.

Vendor Management

Regardless of architecture, a data center typically relies on tools and services from many vendors. Vendor management, in turn, plays a critical role in maintaining continuous data center optimization. Supply-chain management has long been an important element in optimizing everything from manufacturing processes to war campaigns, so why not use vendor management to help continuous improvement in the data center?

Vendor management is essentially a set of practices designed to forge the strongest relationships possible with outside partners and to ensure the best performance from those partners. At a minimum, it can emphasize clear requirements, well-constructed contracts and detailed performance metrics.

A growing number of IT shops have launched vendor management offices that focus solely on these objectives. Some even include purchase and contracting officers in this organization rather than in separate acquisition offices.

Most important, as data center service increasingly enables an organization's innovation efforts, the facility will need to obtain top performance from vendors. Whether buying a management application from a software vendor or using a cloud provider for an infrastructure component, the careful management of vendors is key to preventing costly failures.

Simply put, products or services used to optimize any function in the data center stand a better chance of delivering on their potential when the relationship between the in-house and outside teams is strong and the expectations are clear. ■

This glossary serves as a quick reference to some of the essential terms touched on in this guide. Please note that acronyms are commonly used in the IT field and that variations exist.

Glossary

Agile development

Agile development is an approach to software development that emphasizes incremental steps and regular feedback from application users. It aims to incorporate learning and flexibility without adding excessively to requirements. In the data center, agility means being able to respond quickly to user and organizational changes.

Blade server

A blade server is a small form-factor computer, typically used in arrays mounted together in a frame that fits into a standard rack. Blades are narrower than standard 19-inch rack servers, but they contain most of the components of a complete computer.

Change management

Change management is a process for making sure changes to the data center infrastructure occur in a consistent, documented way. The focus should be on not only avoiding interruptions in service, but also effecting change as efficiently as possible.

Chargeback

Chargeback is a system by which the data center or the IT department in an organization bills users for the computing services they consume.

Cloud computing

Cloud computing is essentially delivering computing as a service. Cloud computing uses different chargeback plans than traditional data center approaches, typically billing flat fees per user per month.

Data center

A data center is a facility that houses enterprise computer resources and the supporting power and cooling infrastructure.

Data center infrastructure management (DCIM)

DCIM comprises software tools for discovering, monitoring and controlling assets forming a data center, including both power and computing resources.

Data deduplication

Data dedup is a method of minimizing data storage requirements by eliminating redundant instances of data. Various deduplication algorithms flag data at the file or block level.

Ethernet

Ethernet is a nearly ubiquitous network technology that divides data into packets or frames. First commercially available in 1980, it has become an industry standard. Throughput typically ranges from 1 gigabit per second to 10 Gigabit Ethernet, but IEEE has published standards for 40 and 100 Gig-E speeds.

Fibre Channel

Fibre Chanel is a high-speed data network technology commonly used for storage area networks within data centers. Originally developed for communications among super computers, it is capable of transfer rates of up to 10 gigabits per second.

Hypervisor

A hypervisor is the layer of software that manages communications between virtual machines and the hardware on which they are running at any given moment. It allows multiple operating systems to run simultaneously as guests on a host physical server.

Incident management

Incident management encompasses the processes and activities conducted in response to one-time events that disrupt data center service.

IT Infrastructure Library (ITIL)

ITIL is a set of service-oriented best practices to guide the data center in supporting the broader organization's needs. ITIL prescribes not only the processes for service management but also the competencies, skills and metrics that are needed.

Media Access Control (MAC) address

A MAC address is a unique identifier added to network interface cards by the manufacturer that is necessary for communicating on networks.

Network-attached storage (NAS)

NAS refers to data storage appliances, designed for use by two or more computers on a common network, that typically house redundant disk arrays. They appear as file servers to applications making calls across the network.

Orchestration

Orchestration refers to the automation of a series of tasks, from provisioning a user to invoking a series of services.

Power usage effectiveness (PUE)

PUE is a ratio measuring the power feeding into the data center to the power consumed by the IT equipment. A 1-to-1 ratio is a practical impossibility because lighting and air-conditioning also use power. A facility with a ratio of 1-to-5 is considered highly efficient.

Problem management

Problem management encompasses the processes and activities aimed at solving recurrent issues in the data center.

Resiliency

Resiliency is the ability of a data center to maintain service in spite of problems such as power outages, server failures or network link failures.

Service-level agreement (SLA)

An SLA is an agreement negotiated between the IT department and a user or a vendor that specifies how a service will be delivered in terms of response times, maximum allowable downtime and other performance parameters.

SPECvirt

SPECvirt is a benchmark published by the Standard Performance Evaluation Corp. to measure performance of servers in virtual environments. It simulates workloads under various server consolidation levels.

Storage area network (SAN)

A SAN is a storage environment that provides access to block-level data in disk arrays running on a dedicated network. SANs are distinguished from NAS, which provide file-level data, and use an interface standard that makes them appear as local storage to the operating system.

Thin provisioning

Thin provisioning is a technique for making the most efficient use of SANs by allotting storage as needed on a dynamic basis rather than in bulk up front based on anticipated future requirements.

Tiering

Tiering refers to the storage of data in the most appropriate medium based on its intended use. Data needed on demand would be top-tier and stored on solid-state or fast disks. Data rarely needed would be archived on the lowest tier, usually optical disks or tape (sometimes offline).

Topology

The term topology refers to network design. Planning data center operations and enhancements require both physical and logical topologies.

Virtualization

Virtualization involves the encapsulation of an application, operating system and memory as a self-contained software unit, known as a virtual machine (VM), that can reside with other VMs on a single server. A VM is not tied to a particular physical machine and can move easily from machine to machine based on load balancing, backup or recovery needs.

WAN

A WAN is a backbone network that serves geographically disparate users — consisting of a combination of dedicated lines, virtual networks over the Internet and wireless technologies.

Disclaimer

Index

10 Gigabit Ethernet (10 Gig-E) 8, 11, 14

Blade servers 4–8, 13–14, 17

Bring your own device (BYOD) 12–13

Byte caching 12

Cable planning 14

Capacity planning 30–32

Chargeback 32

Cloud computing 4, 6, 8, 11

Consolidation 4–6, 13–14, 16, 28

Data center bridging (DCB) 13

Data center in a box 4, 7–9

Data center infrastructure management (DCIM) tool 28

Data deduplication 7, 16

Data tiering 16

Hot migration 8

In-band (and out-of-band) virtualization 15

Infrastructure as a service (IaaS) 7–9

IT Infrastructure Library (ITIL) 29–30

Media Access Control (MAC) 8, 13–14

Network interface card (NIC) 10

Network-attached storage (NAS) 7, 15

Orchestration 5, 27–28

Power and cooling 7, 17

Protocol optimization 12

Redundancy 10

Reference architectures 9

Resiliency 10–11

Server selection 13–14

Service level agreement (SLA) 6, 14, 17, 29, 30

Service-oriented architecture (SOA) 11

SPECvirt_sc2010 13

Storage area network (SAN) 5, 14–15

Storage management 15–17

Storage optimization 15

Storage virtualization 7–8, 15–16

Vendor management 32

Virtual machine (VM) 4, 5, 8–12, 14–17, 27, 29, 32

Virtualization 4–9, 11, 13–15, 27–28

ABOUT THE CONTRIBUTORS

PAUL SCHAAPMAN is a solution architect for CDW. With more than three decades of experience in IT infrastructure, he has a strong background in virtualization (server and client), server and storage engineering, IT architecture, and IT consulting. He was awarded VMware's Virtual Vanguard Award in 2007 for his work on a large virtual infrastructure for the Virginia Farm Bureau.

HOWARD WEISS is one of CDW's thought leaders and presenters on data center architecture, private cloud and next-generation networking. He has a 12-year tenure with CDW and currently runs the Network Architecture team for the western half of the United States.

LOOK INSIDE FOR MORE INFORMATION ON:

- Determining what role cloud computing has in the data center
- Updating and fine-tuning storage architecture
- Making sense of orchestration and automation tools
- Bringing order to data center management processes

120125 108713