

DATA CENTER INFRASTRUCTURE MANAGEMENT

Collecting detailed information to improve planning,
increase agility and reduce costs

Executive Summary

Today's agile data centers need updated management systems, tools and best practices that will let their managers plan strategically, operate these environments at a low cost and analyze workflow to identify improvement options. After all, there is no such thing as an information recession. Quite the opposite; most organizations are struggling to handle more data processing and storage demands, not fewer.

Given budgetary constraints and other concerns, technology staffs need to stretch available resources while reducing costs. And that includes what their organizations spend on physical space and energy consumption for technology services. That's where data center infrastructure management (DCIM) comes into play.

This white paper will explain DCIM as well as provide technology advice on the tools and practices for its implementation and use, and how to take advantage of DCIM to improve energy efficiency and effectiveness in the data center.

Table of Contents

-
- 2 **DCIM: The Big Picture**

 - 3 **Situational Awareness: Analytics and Monitoring**

 - 4 **A Habitat for Technology: Facilities and Resource Planning**

 - 4 **Making It Happen Day to Day**

 - 6 **Efficient, Effective Power and Cooling**

The value proposition for DCIM includes the following:

- Insight into IT resources (facilities, hardware, software, tools and people)
- Timely access to information about IT resources and inventory for decision-making
- Metrics for service planning, budgeting compliance and audits
- Enhanced efficiency and productivity, which allows available budgets to be stretched further
- Agility and flexibility, to adapt to changing conditions
- Improved reliability, availability, serviceability and quality of service (QoS)
- Elimination of islands of underutilized hardware and software
- Reduced complexity and waste, which yields cost savings
- Ability to meet service-level agreements (SLAs) and service-level objectives (SLOs), which can drive up return on investment (ROI)

DCIM: The Big Picture

DCIM, as the term implies, focuses on managing IT resources. In the data center, that covers a wide swath of items, including physical floor and cabinet space, power and cooling, networks and cabling, physical servers and storage, and other hardware and software management tools.

For some enterprises, DCIM is more facilities-oriented, focusing mainly on physical floor space, and power and cooling. Other organizations will have a converged view that encompasses hardware, software and facilities and how they are used to deliver information services cost-effectively.

Common to all DCIM practices are the use of metrics and measurements along with other related information about available resources to gain situational awareness. Situational awareness enables visibility into what resources exist; how they are configured and used (by what applications); and their performance, availability, capacity and economic effectiveness (PACE) to deliver a given level of service.

In other words, DCIM, when properly enabled, will provide an organization detailed insight into its data center operations and help it make timely and effective decisions.

An overarching DCIM approach comprises the following areas:

- Facilities, power (primary and standby, and distribution), cooling and floor space
- Resource planning, management and tracking
- Hardware (servers, storage and networking)
- Software (virtualization, operating systems, applications and tools)
- People, processes, policies and best practices for operations management
- Metrics for analytics and insight (situational awareness)

The evolving DCIM paradigm is around elasticity, multitenancy, scalability and flexibility, and is metered and service-oriented. So it involves a combination of being able to rapidly provide new services while keeping end-user experience and satisfaction in mind. Also, as part of the focus on the end user, DCIM takes into account the ability of an enterprise to compete with outside service offerings as it focuses on being more productive and economically efficient.

Although specific technology domain areas or groups may concentrate their attention mainly on a particular area or mission, identifying and understanding the interdependencies across IT resource areas is essential to achieving an efficient virtual data center. For example, provisioning a virtual server relies on configuration and security of the virtual environment, physical servers, storage and networks, along with associated software and facility-related resources.

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 of the reliance on abstracting physical resources for applications and IT services.

Because of the complexity of the environment, common DCIM activities therefore include the following:

- Measurement, accounting, reporting, analysis, audit, billing and chargeback
- Configuration discovery, remediation and change validation management
- Monitoring of resource usage, and tracking the effectiveness of everything from facility components to application software
- Establishment of service templates, blueprints and guides for configuration
- Performance, availability and capacity planning
- QoS, SLAs, SLOs and service delivery management

DCIM takes an end-to-end (E2E) management approach, relying on situational awareness across a data center's different technology domains and organizational boundaries to understand how IT resources are used to deliver service at any given level. This means having insight into facilities, along with power and cooling, combined with server, storage, networking, hardware, and applications software effectiveness and usage.

In addition to usage and utilization, E2E management and awareness means being able to tell whether users are getting the response time, availability and level of service they are expecting while also knowing how underlying resources are performing or behaving.

E2E management can be accomplished using a single tool that cuts through different technology layers collecting data via

application programming interfaces (APIs), Simple Network Management Protocol management information bases (SNMP MIBs) and other mechanisms.

Another option is to leverage tools already familiar to the technology and applications management groups within the organization. These sometimes have crossover capabilities.

For example, a facilities management tool might see how power and cooling is being used, as well as which servers, network or storage hardware are consuming that power. Server and storage tools might monitor how power and cooling are being consumed, as well as the amount of CPU, memory and input/output operations per second (IOPS) used by particular applications.

Application tools might track how business functions are being processed, as well as provide details about underlying resource usage. The net result of integrating the information gathered by these tools is situational awareness. As a whole, they can let an enterprise track key performance indicators that are critical to making timely informed decisions instead of siloed or compartmental choices.

On Your DCIM Short List

Here is a short list of steps to take to get a data center infrastructure management plan going:

- Establish key performance indicator metrics to measure resource effectiveness and usage for a given level of service delivery.
- Gain end-to-end (E2E) situational awareness of resource usage and effectiveness.
- Implement a converged development and operations paradigm as an extension of an effective DCIM strategy.
- Define – and then manage – service-level objectives and agreements (SLOs and SLAs) capable of meeting user and customer expectations.
- Streamline workflows and align services to stretch resources while reducing waste.

Situational Awareness: Analytics and Monitoring

Performance and capacity planning can be combined as complementary activities with systems or storage resource management (SRM) and utilization, or handled as separate tasks.

Performance tuning and optimization can initially be viewed as reactionary tasks that respond to specific situations. A performance plan and ongoing performance-tuning initiative, however, will let an organization shift from reactionary to tactical and adopt a long-term strategic

management approach. Moving to a performance-planning approach that relies on performance and usage analysis and the ensuing optimization will support overall growth that maximizes IT spending.

DCIM reporting and monitoring tools should allow an IT administrator to see across different technology domains and from virtual server to physical storage for a complete data center picture. In addition, capacity and resource usage tools will add performance and activity reporting to traditional space and capacity utilization, which will provide a more holistic view of resource usage for servers, storage and networks.

Systems or storage resource analysis (SRA) tools step beyond basic resource reporting tools by providing event correlation and other analysis capabilities across multiple technology domains. Some SRM products have (or will morph into) SRA tools over time. And some SRA products can also function as SRM tools, in addition to providing a configuration management database (CMDB) repository, policy management, workflow, reporting, visualization, automation and some orchestration tasks.

A CMDB houses information about IT resources: servers, storage and networking; hardware; software; heating, ventilation and air-conditioning (HVAC) and computer room air-conditioning (CRAC); uninterruptible power supply (UPS) systems; and power distribution units (PDUs). Information in the CMDB can vary from basic asset tracking or inventory to current health and status, usage or activity, location (aisle, rack or cabinet), power usage and cooling requirements, configuration and related data.

The more robust the CMDB, the more data it can correlate to support planning and decision-making. Here's some additional information that an organization should consider capturing about physical resources: whether a resource is purchased or leased, lease and purchase terms, software versions loaded on hardware, all applications assigned to data center support, and maintenance histories.

In addition to information about physical resources, a CMDB can also contain information about performance, availability, capacity, and related historical and real-time data. This can be used for planning and analysis, troubleshooting and monitoring for abnormal behavior. It's also possible to maintain that activity and capacity data in a separate performance management database (PMDB) that has links to the CMDB.

DCIM tools can also leverage the CMDB for storing information about policies and procedures and to establish a knowledge base that will capture workers' experiences, which will help facilitate cross-technology and cross-functional team support.

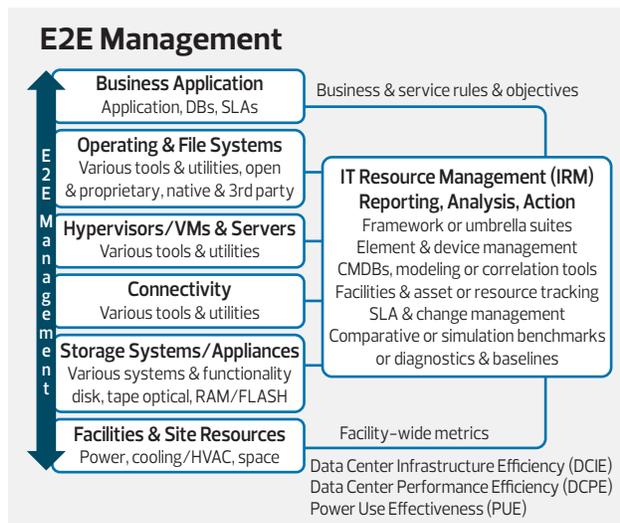
Having timely situational awareness gives you insight into how resources can be more effectively and efficiently deployed. E2E situational awareness removes blind spots from IT

services delivery. Combining situational awareness with service delivery and DCIM activities will also improve cycle times.

This means the tech team can get resources into production faster and minimize the time needed to decommission resources when they need to be replaced. It will help with proper load balancing and elasticity to meet changing workload demands or seasonal surges. All of these have business and organizational benefits, including reduced per-unit resource costs.

The business benefit of virtualization or other forms of abstracting is to provide transparency and agility. But an additional layer of complexity is introduced that requires E2E cross-technology management. Storage and networks are needed for IT departments to effectively manage resources and deliver applications services to users, and SRA tools are needed to support collection and correlation of data from servers.

The *E2E Management* sidebar illustrates E2E management and awareness across different technologies and resource domains by implementing DCIM activities and leveraging a CMDB.



Another critical component in optimizing the data center and integral to DCIM is availability. Initially, it might seem that there's no link between availability and performance and capacity planning. However, 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.

So what are the components of resource usage and capacity planning? They include the following:

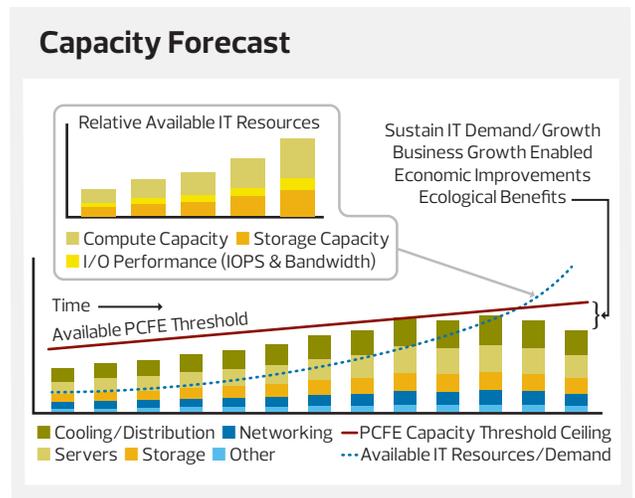
- Status and resource usage monitoring, accounting, event notification and reporting
- Determination of which resources can be consolidated and which need to be scaled

- Performance, availability, capacity and energy usage reporting, planning and analysis
- Diagnostics, troubleshooting, event analysis and proactive management of resources
- Interdependency analysis between business functions and various IT resources
- Cause and effect, high utilization versus poor availability or response time
- Physical and virtual asset and facilities management

A variety of businesses and organizations use capacity planning practices. In a manufacturing company, for example, they apply to inventory and raw goods. Airlines use capacity planning and management to determine when to buy more aircraft. Electric companies use these practices to decide when to build power plants and transmission networks.

By the same token, data centers should use capacity planning and capacity management to derive maximum value and use from servers, storage, networks and facilities (power, cooling and floor space), while meeting service-level objectives and other requirements.

An example of a simple capacity forecast, combined with usage tracking information, can be seen in the *Capacity Forecast* sidebar.



This composite chart depicts servers, storage and networking along with associated power, cooling, floor space and environmental (PCFE) requirements; thresholds for PCFE; and anticipated growth demands. Shown on the right are improvements made to PCFE demands and server, storage and networking optimization. The result is some breathing room to support demand that requires more processing, network and storage resources while fitting into an existing environment.

A data center accomplishes PCFE optimization by fine-tuning servers, storage, networking and virtualization to achieve a data footprint reduction and drive up the energy efficiency of

devices. Newer HVAC and CRAC devices and PDUs also can help optimize PCFE capabilities in a holistic DCIM approach.

Capacity planning can be applied as a one-time or recurring exercise to determine how much and what types of resources are needed to support an application or an array of services. A nontactical approach to resource needs assessment and sizing would be to simply acquire resources (hardware, software, networks and people) as requirements change.

A strategic approach would be to evolve from short-term tactical capacity planning to long-term focused planning so that the data center makes the most use of its resources now and in the future. For example, plotting out resource needs ahead of time would let an enterprise take advantage of special vendor incentives to acquire equipment. Similarly, planning might provide the option to delay a purchase if current terms aren't acceptable.

Ultimately, poor metrics and insight will lead to poor decisions and management. It's essential to look at servers from more than a percent utilization perspective; it's equally important to consider response time and availability.

Think about storage from the perspective of IOPS and bandwidth performance. Also factor in response time or latency, and available capacity. And for networking, take into account latency, cost per given bandwidths and percent utilization.

A Habitat for Technology: Facilities and Resource Planning

One of the most useful aspects of DCIM software and monitoring tools is that they can provide visual representation of resources on a local, metro or wide area geographically, and from a macro or high level with the ability to drill down into specific details.

A high-level view would show general health and status of a data center or equipment room with indicators that display state changes as well as errors, activity levels and other event highlights. A further drill-down should enable identification of a failed or poorly performing component so that the IT team can analyze and troubleshoot in response to the information.

In addition to displays, feeds and event notification, it pays to integrate DCIM tools into the data center's performance and configuration management databases. By doing so, an organization's policy manager and the data center's automation tools can take action based on applicable rules depending on the type of event or activity.

Beyond monitoring, tracking and taking action based on events or activity, DCIM tools also support trend analysis and what-if planning. What-if planning should become a central part of strategic capacity planning and management. It's useful because it allows an IT team to simulate and validate reconfiguration or technology updates that factor in power

(primary and standby), cooling and air flow with other considerations.

Some DCIM analysis software can also validate availability design considerations such as loss of an electrical circuit or cooling device along with impact to servers, storage and networking. Likewise, these programs typically can model the impact of application servers or storage systems failing over for unplanned downtime, planned maintenance or load balancing (along with the affect on the underlying infrastructure).

Other DCIM planning tools address floor and cabinet loading. Smart and intelligent cooling technologies are available to leverage computational fluid dynamics and determine turbulent airflows, cooling bottlenecks and other areas of inefficiency in data center design that are ripe for tinkering. Plus, some DCIM intelligent cooling tools can be integrated with power distribution, cooling and other technologies to further optimize their effectiveness.

Making It Happen Day to Day

Energy efficiency today sometimes simply equates to energy avoidance. However, in the future, the emphasis will shift to doing and enabling more work and storing increasing amounts of information for longer periods.

This will have to be accomplished while consuming less energy and using less floor space. The only way to do more work in a more productive manner with less energy will be to use improved technologies, techniques and best practices to gain efficiencies.

Does the enterprise currently have key performance indicator metrics and dashboards? Are those focused on a specific technology area, such as facilities, power, cooling or availability? Or do they also look at application delivery, response time, server and storage utilization, network errors and reliability?

The Green Grid's power usage effectiveness (PUE) metric shows the ratio of how much total power or energy goes into a facility versus the amount of energy used by IT equipment. As a macro metric, it has a broader focus on how well the physical facility, including the power distribution infrastructure, operates as opposed to how effective a given data center is at delivering services.

For example, if 12 kilowatts (in a small room or site) or 12 megawatts (in a large site) are required to power a data center or computer room, and of that energy load 6kW or 6MW goes to IT equipment, then the PUE for both scenarios would be 2.

A PUE of 2 is an indicator that 50 percent of the energy powering a facility or computer room goes toward IT equipment (servers, storage, networks, telecom and related equipment), with the balance used to run the facility or environment, which typically has had the highest percentage going to heating and cooling.

If an organization focuses only on optimizing the physical facility, PUE is a macro metric that can help gauge that environment in comparison to others. But if the focus is on broader DCIM, including overall effectiveness and productivity, then additional considerations and metrics will be needed. The focus should not be on having the best PUE value, but rather on having an effective PUE that represents a highly productive and efficient data center.

Efficient, Effective Power and Cooling

The increasing density of IT equipment needed to sustain organizational growth and support new applications, capabilities and services has resulted in an array of tough power and cooling challenges for data center managers.

Power consumption per footprint varies widely by operational focus and type of application, the ratio of servers to storage, the amount of online active storage compared with near-line or static and inactive data, and networking bandwidth.

For example, a compute-intensive video rendering, modeling or simulation-based environment might have a high ratio of servers to storage; an online hosting environment might have an even balance of servers and storage; and a fixed or static content provider or managed backup service might have a high ratio of storage to servers.

In addition to occupying less physical space within a cabinet, rack or blade center, each successive generation of technology also provides an increase in net processing or compute power along with a boost in memory and IOPS capabilities.

With heating and cooling accounting for up to half of all electrical power consumed in some data centers, any improvement in airflow and cooling will have a positive impact. For example, if 500kW are being used to cool a data center, and raising the room temperature by 5 degrees Fahrenheit can reduce energy consumption by 5 percent, that's a savings of 25kW.

That might be enough to support near-term growth or transition to more energy-efficient servers and storage. In general, one BTU of heat is generated from 0.293W of energy; on average, 3.516kW of energy (or 12,000 BTU) requires 1 ton of cooling (cooling capacity is measured in tons).

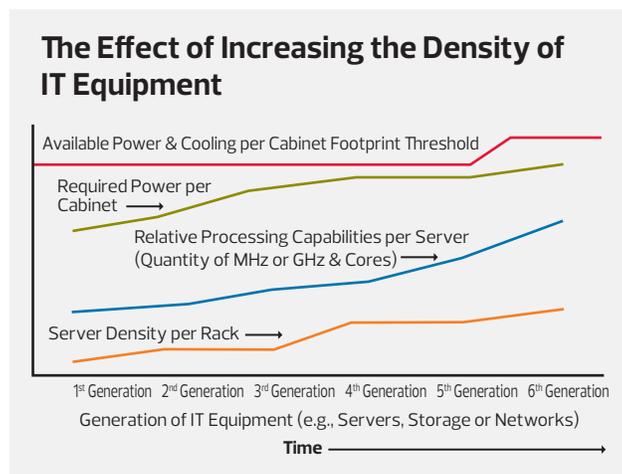
Data centers increasingly are looking to virtualization technology and techniques to help mediate these issues. Through virtualization, they can consolidate servers and storage resources in many environments, which will boost resource utilization and contain costs. With a growing focus on PCFE requirements and the greening of IT in general, consolidating servers, storage, networking and facilities is a popular and justifiable use of virtualization.

But server consolidation using virtualization is only part of an approach to addressing PCFE issues. Consolidation by

itself deals only with those servers and applications that lend themselves to being consolidated. It does not optimize applications that need to scale beyond the limits of a single or even multiple servers. So in addition to server consolidation for candidate servers and applications, deployment of new-generation, faster and more energy-efficient equipment can address PCFE issues as well.

Planning so that dense IT equipment can support growth without aggravating or further compounding PCFE issues requires a delicate balance. Over time and successive generations of IT equipment, more gear can occupy less physical space within a cabinet, rack or blade center. But these various generations of servers also have to exist with the available primary and secondary (standby) power as well as the available cooling footprint.

Each successive generation of equipment also typically provides an increase in net processing or compute power along with a boost in memory and I/O capability. The sidebar *The Effect of Increasing the Density of IT Equipment* details primary and secondary power constraints until a future upgrade to the facility boosts the per-cabinet power footprint. In this example, the relative power required per cabinet increases as the quantity of servers, their speed and the number of processing cores grows.



By studying this example, it becomes clear that the required power per server does not scale linearly, as some improvements in energy efficiency and power management are being factored in to offset the growth. So an important takeaway is that capacity planning for servers, storage and networks needs to incorporate a facility's PCFE capabilities to address how technology improvements will help sustain growth in existing footprints.

IT data centers require electricity to operate servers, storage and networks as well as associated cooling and ventilation equipment. Energy usage may not be well understood by users, or even by most IT professionals or vendors. For a data center IT team to make effective energy allocation and configuration decisions, it will need more

in-depth information about energy use than what is provided on equipment nameplates.

Data on equipment nameplates typically provides for growth without clearly stating what energy is immediately required and what is included for growth. The result is overconfiguration. If a data center overestimates electrical power, it will also likely overconfigure subsequent cooling capabilities, resulting in excessive cooling and energy consumption.

Depending on the size and importance of applications that are hosted in a data center, multiple feeds from different power sources or access points to the broader electric power grid may exist. The feed from the electric utility is considered a primary feed, with secondary or standby power provided by onsite generators.

To maintain availability between a power outage and the short time before generators start up and are ready to provide stable power, UPS systems are used. Power management switches enable automatic or manual transfer from primary utility or self-generated power to standby power for maintenance, diagnostics and troubleshooting, or for maintenance bypass with load balancing.

PDUs transform high-voltage power supplied from energy management systems to IT equipment where it is needed. Electrical power cables can be found either under raised floors or in overhead conveyance systems. PDUs can be located throughout a data center in different zones, providing power to servers, storage and networking devices along with heating and cooling units.

UPS systems bridging the gap between the time utility-supplied power is disrupted and the time a generator is started produce clean power. Depending on the size and type of generator, as well as the power management or control system, the generator may start automatically when the control unit detects loss of power.

The generator may start up within a few seconds, plus another few seconds for the power to stabilize for use. During the gap between loss of utility power and stable power from the generator, either a battery-powered UPS or a motor generator (whose flywheel continues to rotate from kinetic energy) provides bridge power to keep IT equipment running.

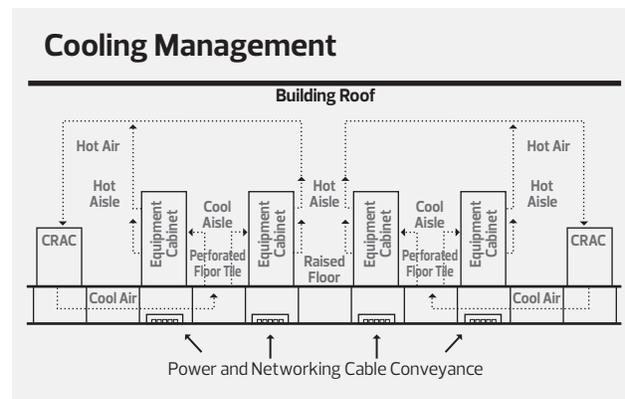
A significant piece of the DCIM puzzle is managing the cooling with the data center to make the most efficient use of the power grid in the facility. In the *Cooling Management* sidebar, the data center has an exposed ceiling and no overhead return air ducting for heat removal.

Hot air is pulled back in from cooling units that produce cool air forced under the raised floor. The cool air under pressure exhausts through perforated floor tiles and is pulled into equipment cabinets arranged in hot and cold aisles.

The cold aisle is the cooling row, where equipment intakes face one another in a row, pulling cool air through the perforated floor

tiles and exhausting hot air into the hot aisle. In the hot aisles, the exhaust sides of the equipment face one another, with warm air being pulled to the ceiling and back to the cooling units.

Also under the raised floor, cable management and cable conveyance systems for networking and electrical power reduce airflow obstructions, reducing the power required to push air throughout the facility. Working with technology suppliers and consultants, air flow and temperature analysis can be performed to assess areas of improvement for optimizing airflow and cooling.



For example, the IT team will want to identify hot spots and cold spots, as well as areas of airflow contention so it can make adjustments. Possible changes might include raising or lowering the temperature in certain sectors, rearranging equipment locations or redesigning exhaust patterns.

Cable management and cable conveyance solutions keep cables grouped together to simplify maintenance as well as maintain good airflow. By keeping cables grouped together, underfloor blockage of airflow is minimized, enabling cooling and heating systems to operate more efficiently.

The same holds true for cabling within cabinets and equipment racks. Even in this location, organized cabling will result in more efficient cooling. Many new data centers are being built or remodeled with several feet of clearance below the floor, sometimes as much as 5 or 6 feet. These higher raised floors facilitate maintenance, the addition of new cabling and better airflow.

A common problem in many environments is a lack of available floor space for new equipment, even though existing equipment cabinets may not be fully utilized because of a lack of available power, cooling or standby power. To help boost available power in a given footprint, energy-efficient cooling (including liquid cooling technologies) can be used, thereby freeing up electricity to power equipment.

To address internal data center policies limiting the amount of power or cooling per physical floor space footprint, new packaging and cabinet setups are available. Cabinets with built-in liquid- or air-cooling can enable denser equipment deployments per physical footprint (assuming adequate power is available).

Modular side, top or in-line cooling shared between racks or dedicated to a rack can also boost available cooling per footprint. For example, current racks and cooling limited to 10 kilowatt-hours could be boosted to 30 or even 60kWh in the same or smaller footprint to support dense server, storage and networking gear.

Virtualization also provides a way to replace old technology and deliver newer, faster, higher capacity, denser, energy-efficient gear to maintain or recapture power, cooling, floor space and environmental footprint capabilities.

For large-scale or extreme mega-scale bulk storage or compute-intensive server farms or grids, large blocks of servers or storage can be replaced on an annual basis with modular virtualization blocks that combine hardware, software and networking.

For example, assuming that an organization refreshes its technology in five-year cycles, it could replace one-fifth of the older technology in its data center environment each year with

one of these new modular blocks to achieve a quick boost in performance and energy efficiency.

Obviously, a chief goal of all these optimization efforts is the ability to extract reliable use and cost data. Executives are increasingly focused on understanding infrastructure and operations in data centers so they can distribute these costs fairly to their users.

The integrated approach of DCIM lets managers gain a more complete and accurate picture of their organization's electric, power and cooling consumption – and, therefore, budget outlays – as they work to drive down use and cost.

That also means the IT group can then establish more realistic customized chargeback reports (supported by actual data) for purposes of billing users and customers, if that is how the organization allots spending for data center services. That information, in turn, can plug back into the capacity planning cycle as the data center looks to the future.



Sustain critical operations that simply can't go down. Emerson Network Power delivers a full range of Liebert uninterruptible power supply (UPS) systems and power distribution units (PDU), from individual data center solutions to integrated systems that keep network closets, computer rooms and data center infrastructure up and running.

CDW.com



dcTrack software is an effective tool to manage data center capacity, assets and change, and helps improve staff efficiency, database accuracy and data center utilization. dcTrack's asset tags, intelligent capacity search and placement, and powerful connectivity management tools help accurately track assets down to the 1U for better overall space, power and cooling utilization.

dcTrack, unlike spreadsheets and Visio diagrams, has a data center-centric point-and-click operation, intelligence to correlate resources, advanced visualization capability and complete power chain and network connection management.

CDW.com/raritan



Organizations struggling with the complexities of availability, server consolidation, virtualization and energy management are increasingly looking for more intelligent physical infrastructure management systems. APC by Schneider Electric brings customers proactive data center management through integrated software applications enabling design, real-time monitoring, inventory management and planning through predictive simulation. StruxureWare for Data Centers' end-to-end management approach allows data centers to run more smoothly on a daily basis, and its built-in business tools for short- and long-term planning and budgeting help keep capital and operational costs and energy use in line.

CDW.com/apc



The information is provided for informational purposes. It is believed to be accurate but could contain errors. CDW does not intend to make any warranties, express or implied, about the products, services, or information that is discussed. CDW®, CDW-G® and The Right Technology. Right Away® are registered trademarks of CDW LLC. PEOPLE WHO GET IT™ is a trademark of CDW LLC.

All other trademarks and registered trademarks are the sole property of their respective owners.

Together we strive for perfection. ISO 9001:2000 certified

108299 – 120425 – ©2012 CDW LLC

