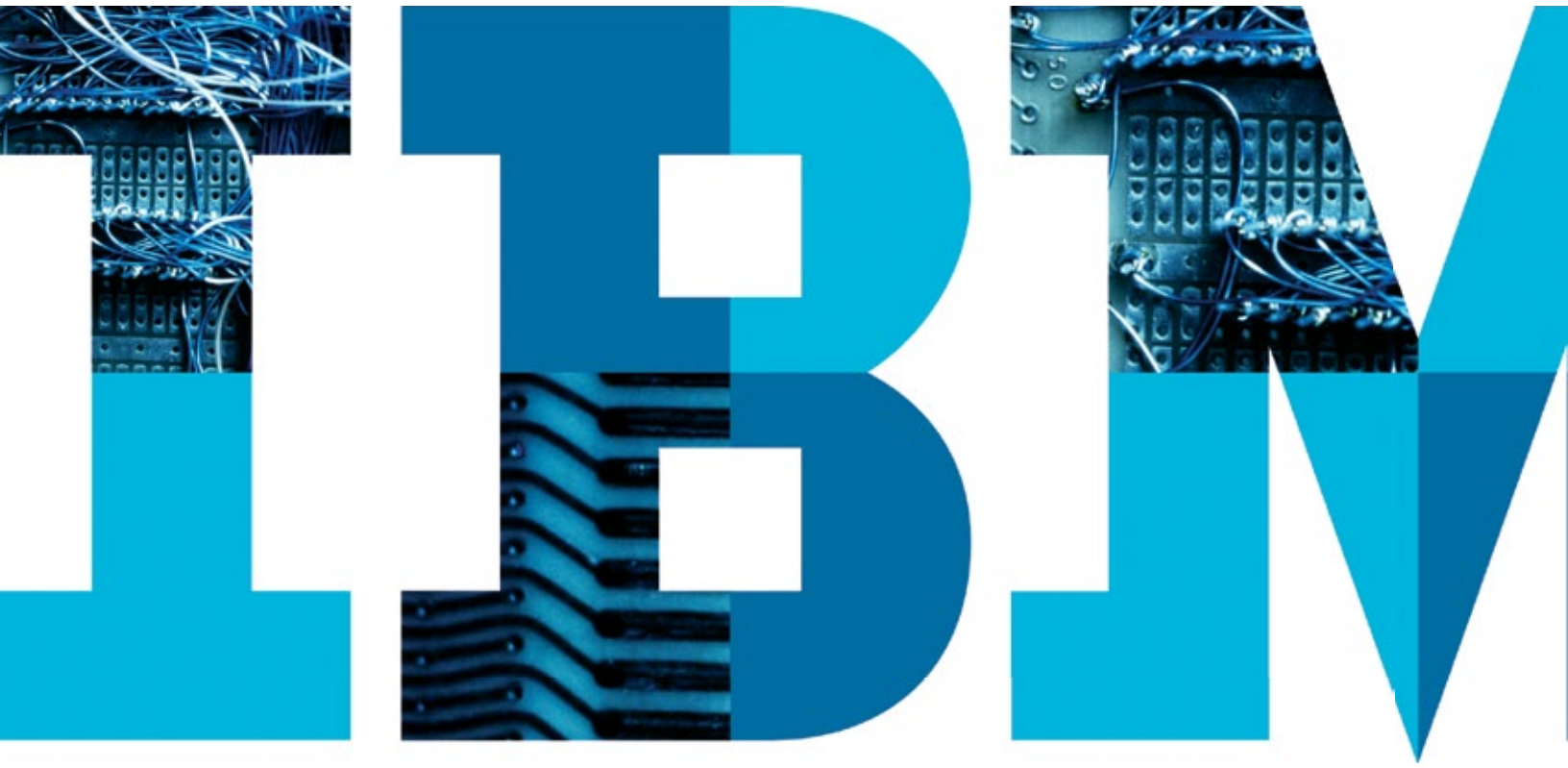
The next-generation data center

A software defined environment where service optimization provides the path



Contents

Introduction

Over the last two decades, IT organizations have faced an uphill battle against complexity. As business demands have multiplied, scores of technologies have been implemented in an attempt to respond. By and large, this steady organic growth has resulted in data centers that are costly and complex, with over-provisioned physical configurations and siloed management. Virtualization and cloud have helped to stem the tide, but even they are hampered by data center inefficiencies, specifically too little automation and orchestration between data center components.

There are inherent limits to what the data centers can do with a finite set of resources, however cloud-driven they may be. These days, with mobility, social business and big data analytics ratcheting up pressure for greater scalability, continuous availability and real-time responsiveness, the old data center mentality must change.

Data centers are no longer simply constrained by four walls. Increasingly, they are a dynamic collection of cloud and non-cloud resources that reside in multiple physical locations and have the automation and intelligence to go where they are needed to process applications efficiently. This is where the next-generation data center begins. It is an IT environment that is simplified, adaptive and responsive, one that allows IT to shift time and attention from maintaining systems to innovating business solutions. It is a composite of capabilities designed for a fast-moving world:

- Software defined environment in which IT resources are orchestrated dynamically and holistically, able to sense and respond to application demands in real time
- Hybrid environment where private and public clouds operate seamlessly with traditional systems
- Continuously available environment able to withstand component failures and maintain operations
- Cognitive computing environment where systems can learn and solve business problems using advanced analytics
- Global, managed ecosystem that integrates the elements of the IT and data center physical infrastructure and provides uniform management through a single console

This paper explores IBM's vision for the next-generation data center, its potential to be truly revolutionary and the prescribed pathway for getting there.

Re-thinking the data center

Today's data centers are no longer just a collection of physical assets. Though virtualization and cloud computing have expanded the boundaries of the data center and the capabilities of its resident hardware, most IT organizations still have a very hardware-centric view of the data center. The focus is still on optimizing individual infrastructure components—servers, storage, networks and facilities (such as generators and UPS systems)—to increase IT efficiency and business value. The problem is that optimizing these elements in separate towers invariably lowers the value to the applications they are meant to serve.

While the hardware components of a data center will always provide a critical foundation and never be reduced to completely interchangeable commodity parts, the advent of "software defined" technologies makes it reasonable to assume that change is on the way. In the next-generation data center, an increasing percentage of critical operational and management functions will be enabled in the software layer rather than by the underlying hardware. This will enable organizations to move away from current manually administered configurations to more dynamic, policy-driven configurations. The impact on IT will be dramatic, substantially lowering costs and risks while raising efficiency and quality of service.

Changing IT's hardware-centric perception of the data center is essential to making this transformation and extracting all the benefits made possible by software defined technology. However, this is not about replacing a hardware-centric mindset with a software-centric mindset. It's about seeing the data center as a business center and provider of the services that advance the business and facilitate innovation. It's about making critical business services—applications like email, customer relationship management and procurement—the focus of an organization's IT optimization efforts. This higher level, business-focused view of the data center is central to achieving greater value. It's what the next-generation data center is all about.

IBM's vision for the next-generation data center

The fast-changing business and IT landscape is driving the need for the next-generation data center. New opportunities cannot wait for resources to be manually procured and configured. Response needs to be immediate and spot-on to meet increasing expectations for availability, scalability and speed.

While virtualization and cloud-based delivery models have answered the need for greater agility, they have also increased management complexity and costs. What's more, the majority of IT provisioning and management tools are labor-intensive and increasingly unable to cope efficiently with the extreme performance demands of today's application workloads. The speed at which the DevOps model is delivering new capabilities and updates is making manual IT operational inefficiencies even more glaring. And despite organizations' best efforts to curb them, outages are on the rise, shining the spotlight on security, resiliency and compliance issues.

Clearly, the demand for more flexible, "always-on" architectures is mounting, advanced by the dynamic needs of mobility, big data and social business. These workloads are raising the stakes for today's data centers and increasing the urgency of the need for the next-generation data center.

So what makes a data center a next-generation data center? In short, the ability to remove many of the barriers that have inhibited IT to date. The next-generation data center provides a simpler, more adaptive infrastructure that is capable of responding to disruptive change, melting technology silos and integrating legacy and new architectures in a single, manageable ecosystem. Several attributes help define this visionary model and distinguish it from its predecessors:

- Innovation enabled by service optimization
- Software defined environment driven by patterns of expertise
- Open standards and support for heterogeneous infrastructures
- Delivery model integration and extensible APIs
- ITIL-based management and outcome-based metrics
- Global ecosystem supported by a converged infrastructure and data center infrastructure management (DCIM)
- Continuous availability
- Cognitive computing
- Integrated, active end-to-end security and automated compliance assurance
- Organizational and cultural change agent

Each of these attributes is described below.

The next-generation data center breaks down the functional silos and cultural obstacles that can get in the way of innovative development and service optimization.

Innovation enabled by service optimization

In a typical data center, optimization efforts are focused on improving IT metrics like utilization, response time and availability. In the next-generation data center, the focus is on optimizing the services that keep the business moving forward: sales and marketing, finance and accounting, procurement and so on. A retailer, for example, would implement cloud technology to optimize marketing services, not to increase scalability. Cloud would help the retailer deliver a more personalized customer experience and provide more targeted marketing. Even though scalability would be a likely outcome, it does not drive the decision to deploy cloud.

Service optimization can dramatically reduce the operational cost and risk of services that are integral to the business. It allows IT to reduce the resources spent running the operation—often greater than 65 percent of the IT budget[1]—and allocate more resources to business innovation. This is important. Increasingly, innovation is viewed as the best defense against competitive threats. IBM's 2013 Global C-Suite Study found the pursuit of extensive innovation to be especially prevalent in top-performing organizations.[2]

As hardware becomes more virtualized and abstracted, these innovations will take place largely above the hardware. Tomorrow's services will be designed and developed by tapping into a near-infinite pool of compute, storage and network resources without any knowledge of the underlying devices. They will also benefit from DevOps, which will bridge the gap between IT operations and development teams, making them better able to work together toward a common service delivery goal. New services can be deployed faster and more cost-effectively.

The next-generation data center breaks down the functional silos and cultural obstacles that can get in the way of innovative development and service optimization.

Software defined environment driven by patterns of expertise

In the next-generation data center, the IT infrastructure won't be controlled manually by administrators making hardware decisions. It will be controlled by software that is programmed to make those decisions automatically. This software defined environment (SDE) optimizes the compute, storage, network and facilities infrastructure, enabling the whole to adapt dynamically to the type of work required. It transforms a static IT infrastructure into a resource-smart, workload-aware infrastructure.

Software defined environments change the rules governing how resources are deployed, literally programming in an organization's business objectives and capturing what IBM calls "patterns of expertise" to define how those objectives will be met. These patterns are essentially best practices for workload deployment, configuration, integration and other complex IT tasks that have been captured by subject matter experts and then codified in templates to be reused again and again. Patterns encapsulate all of the elements required to automate processing for workloads, including the policies that govern it (like application requirements and service levels).

When a workloads runs, the associated template is invoked. The SDE automatically orchestrates the infrastructure resources to meet workload demands in near real time, scaling to meet shifting demand and using predictive analytics to achieve desired performance outcomes. It enables the infrastructure to be exceptionally responsive to unpredictable market changes.

What differentiates the next-generation data center's SDE from other pattern-driven, software defined solutions is its ability to encapsulate infrastructure patterns holistically, codifying the infrastructure in a way that has not been done before. Instead of managing software defined compute, storage, network and facilities resources in silos, it includes the respective middleware and application stacks to manage the complete workload. This enables IT to automate almost every aspect of the data center, from network provisioning to storage setup and provisioning, even power and cooling capacity.

Open standards and support for heterogeneous infrastructures

The next-generation data center is built on open standards. With support for platforms like OpenStack, Linux/KVM and OpenDaylight, it enables organizations to achieve true interoperability between cloud and traditional models. It also facilitates integration of today's heterogeneous infrastructures, enabling organizations to bring their legacy systems into the software defined world. This allows them to preserve legacy applications, hardware, workflows, roles, procedures and data center practices so they can continue satisfying business needs while responding to growing cost pressures.

By providing an open platform, the next-generation data center facilitates the information and service sharing that is crucial for collaboration and holistic management. The IT infrastructure can easily be managed as a collective set of business resources, rather than as discrete compute, storage and networking elements. Its open design enables organizations to exploit new technologies more easily, and it prevents vendor lock-in, increasing the long-term viability of data center investments.

Delivery model integration and extensible APIs

In the next-generation data center, developers will continue leveraging application programming interfaces (APIs) to assimilate functionality from disparate service providers and extend services to a growing array of mobile and other devices. They will also enable hybrid cloud interaction between public and private cloud resources. While APIs are designed to facilitate the integration of cloud models, proprietary APIs limited to specific platforms can complicate integration and slow down the pace of new development. IBM® Watson™ provides a good example of how the next-generation data center will get around them.

Instead of making IBM's cloud-based cognitive supercomputing technology available via platform as a service (PaaS), IBM Watson is being offered under the software as a service (SaaS) model as a proprietary software stack with open APIs. The open APIs allow developers to interact with IBM Watson to build rich applications or run complex analytics in real time. This means they can tap into leading edge—and still proprietary—technologies more quickly and speed the time to value.

ITIL-based management and outcome-based metrics

With the organization served by multiple delivery models and a heterogeneous array of systems, service management is essential to achieving business benefits from IT at a controlled cost. IT initiatives are more likely to stay within budget, and service costs are less likely to spiral out of control. IT Infrastructure Library (ITIL) has long been associated with a service-focused approach to IT management, so it only stands to reason that ITIL-based service management, with metering and analysis for accurate trending, capacity management and chargeback, should be an integral element of the next-generation data center.

IT metrics are another integral element. They continue to measure component-level performance (availability, utilization, recovery time) against service level agreements (SLAs), with financial penalties for service outages. This is especially important with outage costs on the rise. According to a Ponemon Institute study sponsored by Emerson Network Power, a single unplanned outage in 2013 cost businesses more than USD7,900 per minute, up 41 percent from the cost in 2010, with the average outage resulting in over USD900,000 in damages.[3]

What does change in the next-generation data center is the emphasis on outcome-based metrics, like customer satisfaction, productivity and the quality of the user experience. Consider the metrics currently being collected by cities using intelligent transportation systems. These systems provide real-time traffic information and alerts to help drivers avoid congestion and help cities improve roadways to meet citizens' needs. Instead of measuring CPU, memory and disk utilization to assess the systems' success, cities are measuring reduction in traffic, fuel consumption and carbon emissions.

Global, managed ecosystem with a converged infrastructure and DCIM

At this point, it should be clear that the next-generation data center is not just a physical structure, but a global, managed ecosystem with the capacity to share resources within and beyond physical boundaries. It operates on much the same principle as a converged infrastructure, eliminating the need to tie server, storage, network and facilities infrastructure together manually and instead delivering a pre-integrated, optimized architecture that is much simpler to manage and scale using orchestration, automation and policy-driven templates. It delivers cohesive, centralized management in real time, providing increased visibility into all the elements of the physical and virtual infrastructure via a single management entity.

The next-generation data center also addresses the increasing trend toward data center infrastructure management (DCIM), automating control of critical facilities systems like power and cooling. Despite advances in other areas of the data center, management of these systems often relies largely on manual input and has been complicated by the erratic demands of virtualized workloads. The pattern-driven automation inherent in the next-generation data center increases the need for DCIM solutions while also enhancing their capabilities.

In the next-generation data center, data can be extracted from facilities systems and analyzed to drive dynamic adjustments of power and cooling capacity and to automatically identify and correct conditions that could lead to outages. This means that an increasing workload in one area of the data center does not impact the power and cooling systems in that space. It also means that as loads are dynamically shifted from server to server or data center to data center to increase utilization, information about the physical state surrounding those servers can be factored in to drive even greater efficiency. Modular power and cooling systems enable loads to be shifted easily to reduce cooling demand in an area of the data center. In turn, cooling capacity and servers in that area can ramp down automatically, providing considerable energy savings.

Continuous availability

High availability clusters and multisite disaster recovery capabilities used to be the gold standard in data center design. But social business, mobility and the continuing consumerization of IT are mandating even higher levels of availability. "Continuous availability"—99.999 percent uptime (the equivalent of only 27 seconds of downtime per month)—is the standard by which data centers will increasingly be judged. It puts the focus on uptime instead of recovery, since the latter is no longer needed. Outages still occur and maintenance is still performed, but without disrupting service to users.

Continuous availability is the objective of the next-generation data center. To achieve it, the next-generation data center will stretch virtualization of an organization's network, storage and computing platforms across multiple sites. Clustering technologies will be deployed inside the data center, and instances of the data center will be replicated at other locations. Redundancy and resilience become software controlled functions. This allows for dynamic, concurrent updating of data within and between data centers, and enables maintenance to be performed where needed while the overall system stays up and running. In addition, business services will have the automation and intelligence to run from multiple data center locations with minimal intervention. This lessens the need to invest in highly redundant power and cooling systems, and instead focus on optimizing the uptime of those critical business services.

Continuous availability—99.999 percent uptime—is the standard by which data centers will increasingly be judged, and it is the objective of the next-generation data center.

Cognitive computing

Today's data center systems deliver tremendous productivity benefits through automation. Cognitive systems represent the next productivity wave, with the ability to simulate the human thought process at extraordinary speed. They have the capacity to sense, reason and interact with people in new ways. They can process massive volumes of fast-moving data, recognizing patterns, detecting anomalies and making complex decisions in seconds. Their ability to adapt and learn over time and process natural language is what distinguishes cognitive systems from traditional analytics.

In essence, cognitive systems magnify human capability, delivering powerful insights faster than is humanly possible. They can help doctors evaluate and prescribe targeted treatments by accessing globally available medical literature instantaneously. They can help financial services companies make critically timed investment decisions by analyzing vast amounts of data about trading patterns, credit risks and market conditions. They can help cities assess historical weather patterns against current landscapes and resources to determine the probable impact of a weather event and develop a plan for responding. IBM Watson is already demonstrating the enormous potential of these systems for science and industry (see sidebar *IBM Watson—speeding the way to cancer diagnosis and treatment with cognitive computing*).

Integrated, active end-to-end security and automated compliance assurance

Interconnectivity increases the risk of exposure, especially for businesses that are already compromised by rigid security architectures, manual controls and a multitude of dedicated security appliances. The next-generation data center lessens the risk by extending the software controlled environment to security and compliance.

Software manages and orchestrates everything from identity management to intrusion detection and policy enforcement. Security resources are abstracted from individual physical devices and pooled across system boundaries. An organization's assets, processes and information are protected using automatically executed security policies. Analytics are used for security monitoring and compliance checking to proactively identify and prevent security events.

Without dependence on physical hardware, security controls are easier and less costly to deploy however and wherever they are needed. Software controls are also more scalable, so they are able to better accommodate changing business needs and new applications. Security can be logically wrapped around applications in a way that is not possible with hardware-based security.

IBM Watson—speeding the way to cancer diagnosis and treatment with cognitive computing

With valuable data from patient care and clinical trials locked in the heads of clinicians and researchers and in medical databases around the world, physicians are rarely able to access all the information they need to treat patients. IBM Watson is helping MD Anderson Cancer Center change that by improving the knowledge available to physicians and helping them uncover crucial insights about patient treatment.

IBM Watson runs patient data against new information sources and vast libraries of medical literature, helping physicians deliver evidence-based treatment options that are personalized to patients. Physicians can see mutations that are responding well to certain treatments, and they are using that information to identify the best patients for clinical trials. As new therapies and investigational protocols are rolled out and compared, the results are being used to advance future care.

The next-generation data center leverages an integrated security strategy to better understand threats and vulnerabilities in terms of business impact, better respond to security events with optimal business results, and better quantify and prioritize security investments. It is an end-to-end, business-driven approach to security, compliance and risk management, operating within a governance framework to ensure that IT's recommendations and practices maintain alignment with business goals.

Organizational and cultural change agent

Any description of the next-generation data center would be incomplete without discussing the organizational and cultural impact. Converged infrastructures and software defined environments blur the lines between technology silos and, in so doing, force a significant cultural change on business and IT. In truth, the success of the next-generation data center depends as much on the integration of stakeholders as it does on the integration of systems and tools.

Data center comparison: evolution from the traditional to the next generation

	Traditional data center	Software defined data center	Next-generation data center
Definition	Siloed, hardware-centric infrastructure, often oversized and complex, requiring skilled IT specialists to operate and manage	Infrastructure delivered as a service and controlled by software that orchestrates IT decisions automatically across cloud and non-cloud environments	Converged, software defined infrastructure leveraging codified patterns of expertise to automate and execute infrastructure functions holistically
Operational paradigm	• Workloads assigned manually to server, storage and networking resources • Manual IT optimization • Reactive approach to opportunities and competitive threats	• Workloads assigned automatically to best-fit resources by policy-driven software • Software-driven IT optimization • Proactive approach to opportunities and competitive threats	• Workloads assigned automatically to best-fit resources based on infrastructure patterns • Dynamic IT optimization with cognitive learning • Proactive approach to opportunities and competitive threats
Management	• Siloed management of servers, storage and network resources • Multiple tools needed to accommdate heterogeneous, multivendor systems	• Holistic, centralized management of IT resources • Orchestrated infrastructure management via shared software tools	• ITIL-based service management • Increased visibility and simplified controls via a single console • DCIM
Metrics	• Hardware-based metrics, capturing component-level performance • Manual analysis required	• More integrated view of performance • Analytics used for assessing performance and addressing infrastructure issues	• Emphasis on service-based metrics focused on business outcomes • Real-time analytics insights used for dynamic infrastructure optimization
Security and compliance	• Rigid, complex security architecture with appliances dedicated to individual systems • Compliance checking, reconfigurations and fixes handled manually	• Automated protection and compliance provided uniformly via software-driven policies and rules	• Automated protection and compliance based on policies and rules • Heavy use of analytics to predict events and maintain compliance

Formerly fragmented business units and IT organizations must work together as a team on budgetary issues, performance metrics and everything in between to make decisions in the best overall interest of the business. The next-generation data center also challenges traditional support roles, forcing many system, storage and network practitioners to move to a more generalist role that demands greater knowledge about the infrastructure as a whole while maintaining their domain expertise.

Becoming a next-generation data center

Most data centers, with their siloed, heterogeneous mix of systems, processes and delivery models, are inherently complicated. Executing the next-generation data center vision requires transforming the IT infrastructure through the abstraction, automation and orchestration of server, storage, network and facilities assets, and that can take time. Most organizations will want or need to take a phased approach. IBM has identified four stages along this transformation continuum:

- Consolidation and integration
- Optimization and standardization
- Simplification and automation
- Dynamic optimization

Transformation continuum

The transformation continuum outlines an evolutionary path to the next-generation data center. Each stage moves further away from the complex, labor-intensive, reactive methods that have dominated the traditional data center toward a more simplified, automated and proactive data center operation. Certainly, most companies have already started along this path, consolidating and virtualizing servers, virtualizing storage and network resources, investigating or implementing DCIM, and automating some functions. A few are starting to venture into the software defined realm, recognizing the potential of software to quickly configure development resources and drastically simplify management.

Where an organization enters the continuum is based on its data center maturity, including the scope and depth of virtualization, standardization and automation of its systems and processes. It is not unusual for an organization's servers, storage and networks to be at different levels of maturity and enter the continuum at different stages. Infrastructure components that support legacy applications might enter the continuum at the beginning and only go through one or two stages before maxing out their value. For those components, the automation and shared services enabled in the later stages could prove too costly to deploy. By comparison, components that support born-on-the-cloud applications are likely to enter the continuum at a later stage. These components are not only able to benefit from the final stage; they actually need to progress to that stage to deliver their highest value.

At any given time, an organization's server, storage, network and facilities components are in varying stages of transformation. Each stage of the continuum is designed to take an organization closer to the next-generation data center and deliver important efficiency gains.

Consolidation and integration

The first step is to rein in the uncontrolled infrastructure sprawl that has left organizations without the flexibility to respond quickly to new technologies, business requirements or rapidly shifting workloads. Consolidation addresses inefficiencies, relieving the cost of maintaining more systems and applications than necessary and saving energy, space and labor while improving asset utilization.

Once the organization rationalizes and consolidates services and their corresponding workloads, it is possible to consolidate and integrate the systems that support them. Integration begins with virtualization, abstracting physical systems from the services that are trying to use them. This ability to separate the logical from the physical provides hardware independence, and that enables IT resources to be shared and dynamically

served up as needed. Virtualization allows the organization to do more with fewer physical assets, reducing the energy demands on the infrastructure while expanding IT capacity.

Analytics are taking the guesswork out of identifying systems for consolidation and virtualization and radically speeding the process. By providing a more complete and accurate picture of the IT infrastructure, they are enabling organizations to make better-informed decisions about optimizing it.

At this stage, the lack of standardization and automation limits what virtualization technologies can do. Getting the most out of the virtualized infrastructure requires progressing through the next stages of the continuum.

Optimization and standardization

Service optimization is at the heart of the next-generation data center. When IT services are optimized, it is possible to redirect IT spending from day-to-day operations to innovations that drive business growth. But service optimization requires taking a step back and establishing standards for the services operating in the data center.

Service standardization helps manage IT cost, quality and risk. It allows IT administrators and users across the organization to accomplish common IT tasks consistently and reliably each time they perform them—one of the key principles of IT service management (ITSM).

To standardize, organizations must define and prioritize the services operating in the data center—a process that helps determine which services are business-critical and which ones are not. This puts organizations in a better position to make decisions about rationalizing and right-sizing the data center, including decisions about the number and size of data centers and the availability levels to be provided. They can standardize services and workloads that are most valuable to the business and retire those that are least valuable and redundant. Fit-for-purpose analytics and business criticality assessments can help simplify these decisions (see sidebar *Service optimization—simplified through automation and analytics*).

Service optimization—simplified through automation and analytics

For years, IBM's fit-for-purpose analytics have been helping organizations map workloads to their best-fit platforms. Potential platforms are assessed analytically against more than 30 workload and processing variables to identify the most cost-effective platform.

The next-generation data center leverages a fit-for-purpose architecture to continuously assess workload requirements against available resources and speed the mapping of workloads to their optimal platforms. Workloads are shifted between available platforms to satisfy fast-changing demands, while maximizing utilization and performance at the lowest possible cost.

IBM also uses business criticality assessments to lower cost and risk, assigning workloads to different tiers of service based on the their value to the business. The most business-critical workloads are typically assigned to the top tiers, which offer the highest levels of availability, support and redundancy at the highest cost. The top tier might provide greater than 99.7 percent availability, 24x7 support for all severity level 1 problems, and multisite redundancy. The bottom tier might provide less than 95 percent availability, limited support and no site redundancy.

Segmenting workloads into service tiers allows organizations to pay for the highest levels of availability and redundancy only when it is needed. Considerable savings have come from shifting less critical workloads to lower level tiers. In the next-generation data center, these shifts are handled dynamically with policy-driven automation and enforcement.

In today's heterogeneous, rapidly evolving data center environment, decisions about which services to keep and optimize are more complicated. Workloads are not one-size-fits-all. Each workload's resource, integration, service level and data requirements have to be factored in, as does its impact on the network. IT needs to determine the accessibility and availability of workload data and its portability between applications.

Service optimization and standardization lay the groundwork for automation. By establishing standard policies for repetitive IT procedures, it is possible to automate those procedures and assure greater quality and predictability each time they are performed.

Simplification and automation

This stage represents a major step forward as it turns IT from a cost center to a service provider. Instead of just delivering infrastructure, automation simplifies the provisioning and management of IT services by enabling their delivery without the need for human intervention.

Automation lets IT step out from behind the management console. It allows users to direct the deployment of new application environments and other IT tasks by accessing a self-service portal and selecting from a standardized catalog of services. Each time users execute a service request, the request is fulfilled via an automated workflow in the cloud. Virtual machine images are created, host servers are selected and storage is allocated automatically in the cloud according to established template guidelines. The fact is, workload automation is an indispensable part of cloud computing. Without it, many of the features and benefits of cloud cannot be achieved at an effective price point.

Automation simplifies cloud processing. By mechanizing the provisioning and scaling of resources, no workload is left to struggle or use excess capacity. The self-service interface provides an intelligent front end, enabling users to track the status of their requests online and eliminating delays and errors. Automation also facilitates handoffs, such as between application development and testing, providing consistent processing and more predictable outcomes.

Dynamic optimization

In the final stage of transformation, the focus is on becoming a software defined environment, freeing applications completely from the physical infrastructure so that they can operate more easily and pervasively in the cloud. This stage extends what cloud and automation can do.

Logic is programmed into the software to orchestrate the delivery of virtualized infrastructure resources, along with corresponding middleware and applications. This allows the infrastructure to dynamically adapt to changing workload requirements. Using analytics and cognitive computing, the software is also able to effectively optimize its response to provisioning and other service requests as they occur. It learns from experience and improves with each interaction and outcome. So instead of operating by a static set of rules, organizations are continuously invoking and updating those rules in real time based on situational intelligence and cognitive learning.

This ability to dynamically optimize helps IT accommodate the growing number of service requests from line-of-business users while maintaining security and compliance. It enables these users to execute big data analytics workloads whenever needed, without delay and without committing compute resources any longer than it takes to run the workload. And if the workload needs more bandwidth, the associated policy can respond automatically and make the adjustment, then capture and apply that knowledge to future workload requests.

Simplifying the journey

The software defined movement is justifiably gathering momentum, and there is no shortage of providers lining up to participate. Most are offering component-level solutions for servers, storage or networks while recognizing the benefits of a more holistic view. As a result, some providers are joining forces or acquiring companies to be able to offer the full complement of capabilities required for a software defined data center. Buyers should be aware that simply bundling existing solutions under one brand is not sufficient.

When disparate technologies are merged and new layers are added to the existing management stack, there can be serious integration challenges. These can counteract the benefits of a next-generation data center, inhibiting scalability and increasing management complexity. True integration is essential, and it is harder to accomplish after deployment. With it, orchestration between applications, services and infrastructure resources can be highly automated and transparent. Without it, automation around access, security and performance can be difficult, and delays are likely to occur.

Software defining the infrastructure as a whole, rather than as siloed components, allows for the integration required. So it only makes sense that the journey to the next-generation data center should begin with this more holistic view.

Toward that end, expert systems may provide the quickest route to the integrated, software defined environment required for the next-generation data center (see sidebar *The advantages of expert integrated systems*). Integrated by design, these systems automate best practices for server, storage and network resources right out of the box while enabling companies to tune for their own workloads, service levels, security and compliance requirements.

The advantages of expert integrated systems

Operational simplicity is a key benefit of expert integrated systems. Because system resources are deeply integrated and software controlled, day-to-day provisioning, deployment and management of physical and virtual resources takes very little time, enabling IT expertise to be channeled to more innovative pursuits.

IBM PureSystems® offerings combine compute, storage, network and management software into preconfigured platforms optimized for specific workloads and customer environments. IDC interviewed 10 companies to assess the impact of IBM PureSystems offerings on their IT efficiency.

On average, companies experienced:

- **47 percent less time spent managing the infrastructure for improved staff productivity**
- **55 percent faster system deployment**
- **67 percent faster application deployment using predefined patterns to provision development environments**
- **57 percent reduction in operating expenses resulting from platform automation and centralized management**
- **32 percent reduction in server hardware requirements, reducing the data center footprint by 29 percent and lowering capital costs.[4]**

The best expert systems can sense and anticipate the resources required for optimal workload performance. They have the ability to provide unified management for the entire IT environment via a single console, offering a truly integrated single point of control rather than a coordinated set of management tools. Furthermore, they are built on an open architecture with a standard set of interfaces that developers can use to innovate and deliver substantial business value.

DevOps can also facilitate the journey to the next-generation data center by implementing the processes and tools required for software-driven automation. DevOps can help virtualize across all data center resources, make those resources programmable and deliver the automation capabilities that enable the software defined environment.

The urgency of transformation

One way or another, most organizations are moving towards the next-generation data center. They are taking steps to virtualize infrastructure elements and break down management silos. They are venturing into the cloud to increase scalability and disconnect applications from physical resources that limit their availability. Still, it has been difficult to keep up with the staggering rate of technological change brought about by the consumerization of IT and the combined impact of social, mobile, analytics and cloud technologies.

Part of the power of the next-generation data center is the opportunity it creates to assimilate emerging technologies seamlessly. Perhaps even more important are the business opportunities that could be missed without it, including the chance to:

- Leverage vast social media resources to understand, anticipate and capitalize on the desires of digitally empowered customers and constituents
- Spur runaway collaborative innovation by openly democratizing data across customers, partners and employees
- Create personalized customer experiences 24x7 using sentiment mining, cognitive learning and other analytic tools

Responding to opportunities like these cannot wait for new hardware to be procured, installed and configured. It has to happen immediately.

The truth is that as technologies continue to evolve and shape the way people work, play and live their lives, falling behind can have detrimental business consequences. Organizations have a critical choice to make. They can wait for competitors to lead the way to the next-generation data center, or they can move past the technological and organizational hurdles and actively begin transforming now. Getting the right people and processes in place to help drive this kind of change is paramount for any company looking to stay ahead of curve.

Part of the power of the next-generation data center is the opportunity it creates to assimilate emerging technologies seamlessly. Perhaps even more important are the business opportunities that could be missed without it.

IBM knows data centers

IBM has long been a leader in the technologies that make the next-generation data center possible, inventing and steadily improving virtualization, centralized infrastructure management and the hardware that enable greater control and automation of IT. While other vendors focus on providing point solutions for software defined server, storage or network components, IBM is focused on delivering an end-to-end software defined environment. Ours is a solution—and a vision—that crosses all of these data center domains and addresses the IT infrastructure holistically.

Our deep understanding of heterogeneous environments, workload-aware infrastructures, pattern-driven automation, cross-domain integration and orchestration, and analytics-enabled optimization underlies this vision. It accelerates the deployment of the simplified, responsive and adaptive infrastructure required for the next-generation data center.

IBM continues to recognize the importance of open standards to break down technology barriers, fuse silos of expertise and deliver on the promise of emerging technologies. This is especially important as line-of-business users play a more dominant role selecting and deploying IT services and as DevOps teams work to facilitate integration and interoperability across the enterprise.

Conclusion

The next-generation data center represents the next evolution of the converging IT infrastructure, where server, storage, network and virtualization resources are abstracted from the underlying hardware and workloads run on the most appropriate combination of resources, whatever they may be. In this environment, software provides the intelligence for managing the infrastructure dynamically and holistically, based on real-time workload needs. The next-generation data center transforms a static IT infrastructure into a dynamic, workload-aware infrastructure that can anticipate demands and respond with incredible speed.

Becoming a next-generation data center requires change, not just in the physical and operational aspects of the data center, but also organizationally and culturally. The evolution from hardware-centric thinking to service optimization is critical. Organizations that want to speed the transformation will choose software defined solutions that come fully integrated and ready to orchestrate the entire IT environment. Those that succeed will be opened up to new ways of working and innovating—above the hardware—making it easy to operate in a hybrid environment, speed the development of portable applications and capitalize on rapidly evolving cloud, mobile, social and analytic technologies.

For more information

To learn how IBM is helping organizations deliver on the next-generation data center, please contact your IBM representative or IBM Business Partner, or visit
ibm.com/services/resiliency

IBM Global Services
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
September 2016

[1] IBM, *"Data center operational efficiency best practices: Findings from the IBM Global Data Center Study,"* April 2012.

[2] IBM, *"Reinventing the rules of engagement: CEO insights from the Global C-Suite Study,"* November 2013.

[3] Ponemon Institute, *"2013 Cost of Data Center Outages study sponsored by Emerson Network Power,"* December 2013.

[4] IDC, *"IBM PureSystems: Delivering IT efficiency,"* IDC #242685, August 2013.

RLW03027-USEN-03