Keeping the Data Center Competitive

Six Levers for Boosting Performance, Reducing Costs, and Preparing for an On-Demand World
Contact Information

Beirut
Ramez Shehadi
Partner
+961-1-336433
ramez.shehadi@booz.com

Berlin
Dr. Florian Gröne
Senior Associate
+49-30-88705-844
florian.groene@booz.com

Chicago
Mike Cooke
Partner
+1-312-578-4639
mike.cooke@booz.com

Frankfurt
Stefan Stroh
Partner
+49-69-97167-423
stefan.stroh@booz.com
Dr. Germar Schröder
Principal
+49-69-97167-426
gemmar.schroeder@booz.com

Hong Kong
Edward Tse
Senior Partner
+852-3650-6100
edward.tse@booz.com

London
Louise Fletcher
Partner
+44-20-7393-3530
louise.fletcher@booz.com

Milan
Enrico Strada
Partner
+39-02-72-50-93-00
enrico.strada@booz.com

Mumbai
Suvojoy Sengupta
Partner
+91-22-2287-2001
suvojoy.sengupta@booz.com

New York
Jeff Tucker
Partner
+1-212-551-6653
jeff.tucker@booz.com

Sydney
Chris Manning
Partner
+61-2-9321-1924
chris.manning@booz.com

Tokyo
Shigeo Kizaki
Partner
+81-3-3436-8647
shigeo.kizaki@booz.com

Christopher Schmitz and Christian Beekes also contributed to this Perspective.
EXECUTIVE SUMMARY

Cost and capacity pressures on the corporate data center are mounting. Increasing computing power demands, poor asset utilization, excess complexity, and growing concerns about energy usage and costs are forcing companies to reassess how they manage their data centers. Whether they operate their data centers for internal customers or as third-party providers of data center services to others, companies that don’t make the effort to rethink their data center strategy face a future of rising costs and declining performance relative to their competitors. Companies that do make the effort can expect to cut the cost of operating their data centers by as much as 40 percent.

Managers of data centers should look at six areas in which their operations can be improved. The greatest potential savings lie in improved utilization of data center assets, through server and storage virtualization and by making better use of the data center facility itself—including a careful analysis of the total worldwide footprint of data center facilities as well as how operations are organized. Understanding how and when data center resources are consumed can further improve asset utilization and save energy. Restructuring the data center’s operating model can increase efficiency, as can devising a global sourcing strategy for data center services. Finally, moving to a demand-driven model that rationalizes platforms and products will help set the stage for the creation of the data center of the future, one that can give corporate customers what they want: efficient and flexible computing capacity.
At a time when every large-scale organization is looking to cut expenses and streamline operations, the data center has come under increasing pressure to make its operations leaner. And the time is ripe: Traditional data centers are facing the upper limits of their data capacity even as they continue to underutilize computing assets, while their massive appetite for electrical power continues to raise concerns about their impact on the environment.

Any number of information-intensive companies—including the likes of Google, Microsoft, Facebook, Deutsche Bank, and DHL—are making major investments in state-of-the-art data centers in hopes of making their data center operations more efficient and less costly. Such efforts can lower total capital and operating expenses by as much as 40 percent. If corporations desire to maximize the value of their data center assets and reach the next level in performance, cost efficiency, and quality control, they must begin now to rethink the core structures of their data center production model.

Creating the data center of the future will require a reassessment of the current model in six specific areas: technology platforms, data center topology, consumption management, end-to-end process efficiency, global sourcing models, and commercial models. The risk of not doing so? Falling behind in the very competitive race for IT efficiency.

**RETHINKING THE DATA CENTER**

**Key Findings**

- Even as demand for data services is on the rise, the data center is under tremendous pressure to cut costs, reduce energy usage, and develop new delivery models.
- Those pressures, and the threat of rising costs, will force every data center operator to reassess how it does business if it wishes to remain competitive.

We believe there are six areas in which companies can work to improve their data center operations:

- Improve asset utilization through virtualization
- Consolidate the data center footprint
- Manage consumption to reduce usage and energy costs
- Restructure data center management and operating model
- Create a global data services sourcing strategy
- Modularize services offerings and rationalize payment schemes
Four factors are driving the need to rethink the data center. First, despite ever more powerful computing technologies, such as multi-core, 64-bit chip architectures, the installed base of servers has been growing 12 percent a year, from 14 million in 2000 to 35 million in 2008. Yet that growth isn’t keeping up with the demands placed on data centers for computing power and the amount of data they can handle. Almost 30 percent of respondents to a 2008 survey of data center managers said their centers will have reached their capacity limits in three years or sooner.

At the same time, too many data centers are just not managed very well. Thanks to long histories of legacy systems and software, a lack of discipline regarding standards, and ineffective life-cycle management, the typical data center is saddled with a large and unwieldy inventory of operating systems, database software, and middleware. That in turn adds hugely to the complexity of systems administration and maintenance—not to mention adding to cost. And server utilization is often low, with too many CPUs idle or nearly so. Poor utilization is a central cause of inflated data center capacity requirements, unnecessarily high investments in hardware, and big energy bills.

Moreover, companies face growing concerns about the enormous amounts of energy their data centers use and the resulting high levels of CO₂ emissions they cause; indeed, a number of European countries are looking to regulate emissions even more strictly than they do now. European data centers currently consume more energy than the entire country of Denmark; by 2019 they are expected to use more than the Netherlands.
consume more energy than the entire country of Denmark; by 2019 they are expected to use more than the Netherlands. A European Code of Conduct is already in place; companies subscribing to the code must make a voluntary commitment to reduce power consumption by applying best practices such as energy audits, specific action plans for reducing emissions, and continuous monitoring of energy consumption. The auctioning of emission certificates to utilities, slated to begin in 2012, will further increase the cost of electricity.

There is only one future for companies whose data centers continue to be traditionally operated: a world of rapidly rising costs. Under current practices and assuming constant capacity, data center costs will rise 17 percent over the next four years (see Exhibit 1), even as both external and internal customers apply downward pressure on the price of data center services. Much of that increase is due to the rising cost of energy and higher labor costs, including wage inflation, an aging, more highly paid workforce, and a coming shortage in the skills needed to operate data centers, all of which will result in stagnating productivity. Although continuing improvements in hardware

---

**Exhibit 1**

*Over Time, Manpower and Energy Cost Inflation Will Eat Up Traditional Operators’ Margins*

---

**Source:** Booz & Company Econometric Data Center Planning Model
performance will mean that the ratio of price to performance will remain steady, rising energy costs will eat up those gains, as will higher costs of construction, operations, and facility services. The inevitable result: Driven by bottom-line cost pressures alone, providers of data center services will see their already thin margins erode further, turning average profits of 5 percent of revenues in 2008 into losses of 11 percent by 2012.

Only by reinventing how they run their centers—rethinking everything from technology platforms and data center topology to consumption management, global sourcing models, end-to-end process efficiency, and commercial models—can providers of data center services hope to thrive despite the pressures they face (see Exhibit 2). How should data center operators work to transform their businesses in each of these six areas?

Exhibit 2
CIOs Must Rethink the Core Structures of Their Production Model to Remain Competitive

<table>
<thead>
<tr>
<th>Structural Optimization Levers</th>
<th>Potential Gains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial Models</td>
<td>N/A</td>
</tr>
<tr>
<td>• Capacity-on-demand is the wave of the future, requiring service modularization and transparent payment schemes.</td>
<td></td>
</tr>
<tr>
<td>Global Sourcing Models</td>
<td>10%–15%</td>
</tr>
<tr>
<td>• Optimizing the global delivery strategy involves balancing commodity applications with more complex tasks.</td>
<td></td>
</tr>
<tr>
<td>E2E Process Efficiency</td>
<td>5%–15%</td>
</tr>
<tr>
<td>• Restructuring both the data center management organization and the operating model can lead to better capacity planning and lower administrative costs.</td>
<td></td>
</tr>
<tr>
<td>Technology Platforms</td>
<td>15%–20%</td>
</tr>
<tr>
<td>• By increasing utilization, server and storage virtualization has the greatest potential for lowering data center costs.</td>
<td></td>
</tr>
<tr>
<td>Data Center Topology</td>
<td>15%–30%</td>
</tr>
<tr>
<td>• Footprint consolidation and a proper tier structure can save money, but a balance must be struck between scale and complexity.</td>
<td></td>
</tr>
<tr>
<td>Consumption Management</td>
<td>5%–25%</td>
</tr>
<tr>
<td>• Better utilization of data center assets and reduced energy consumption offer significant benefits in cost and reduced complexity.</td>
<td></td>
</tr>
</tbody>
</table>

Note: All figures should be read as “total cash-out reduction potential”; they include both capital and operational expense reductions.
Source: Booz & Company analysis
Of all the steps that can be taken to reduce costs, using data center assets more efficiently has perhaps the greatest potential for generating significant savings. Underutilization of both computing and facility assets remains a large problem in data centers: Servers typically run at less than 10 percent of capacity, and it is not uncommon for more than 50 percent of data center floor space to sit underutilized as well. The result: significant amounts of operational and capital expenditures that could be better used elsewhere. After a detailed analysis of its data center costs, one large European corporation found that a large-scale virtualization program had the potential to lower its overall data center costs by 29 percent.

The economics of virtualization are powerful (see Exhibit 3). The technology has the potential to lower
total costs of ownership by as much as 40 percent. Consider hardware. Dedicated servers frequently have utilization rates of less than 10 percent, whereas servers run virtually can often more than double those rates. And even though virtual servers typically cost more per CPU, the overall benefit can be hardware cost savings of between 15 and 35 percent.

Despite the very real benefits of virtualization, it rarely makes sense to try to virtualize everything. Indeed, the “sweet spot” for generating the maximum return on investments in virtualization—primarily high-volume platforms such as Windows and standard Unix or Linux—lies somewhere between 20 and 60 percent of assets, depending on the production model and computing footprint. Trying to virtualize more than that will usually involve the virtualization of more exotic platforms with lower server counts, such as legacy Unix systems, and the effort simply won’t bring the returns expected.

Achieving the maximum return requires a careful review of the preconditions in the data center for successful virtualization. That includes, on the hardware front, an inventory of assets—the number of harmonized vendor clusters and machines with fewer than four CPUs—and an analysis of utilization levels. As to software, the inventory should include harmonized operating systems, database software and middleware clusters, and standard, multi-platform certified applications such as Web and e-mail and standard ERP and CRM. Finally, it’s important to ascertain whether any applications have technical restrictions, such as maintenance liabilities, that might restrict the use of virtualization.

Dedicated servers frequently have utilization rates of less than 10 percent, whereas servers run virtually can often more than double those rates.
Large multinational corporations use different strategies for siting their data centers. Some may find themselves running dozens of data centers around the world. Hewlett-Packard, for instance, maintains about 60 centers worldwide. Others, such as ING, maintain just one primary hub. Data center topology, however, creates a dilemma: Scale or resilience? A topology that includes a small number of large-scale centers offers the benefit of scale, but the lack of diversification can pose a security risk, and individual centers risk being simply too complex to operate efficiently. A plan that includes many smaller centers runs the opposite risk: Security concerns and complexity are eased, but the individual centers may not be large enough to reap the maximum benefits of scale. Where is the happy medium?

To operate at peak efficiency, data centers should be about 10,000 square meters. At that size, each center’s annual operating expenses are minimized, but the added costs created by excess complexity have not begun to make themselves felt (see Exhibit 4).

Another way to put the problem is in terms of utilization. Obviously, data centers are expensive. Looked at in terms of cost as a function of utilization rate, however, unit costs come down rapidly. But the benefits go only so far. After about 90 percent utilization, data centers run a real risk of losing operational flexibility. Generally speaking, the utilization goal should be about 80 percent, which leaves adequate headroom for peak demand. Service providers will want to leave somewhat more room depending on their mid-term deal pipeline, which may add sudden large demands on their data centers. By the same token, in coping with capacity bottlenecks, before expanding capacity in a lower-tier center and reducing its utilization rate, consider the possibility of using higher-tier capacity with better utilization and overall lower unit costs.

Exhibit 4
When Consolidating Data Centers, It Is Vital to Find the Right Balance between Scale and Complexity

![Data Center Cost by Size](chart)

Source: Booz & Company analysis
On the other side of the utilization coin is the issue of consumption management, involving the consumption of both computing assets and energy. Significant savings can be found in the effort to reduce the use of assets and to optimize the kinds and number of assets being used. The key here is to implement an efficient capacity planning process. On the consumption side, begin by identifying the resources required for each software application. Then retire or move any resources that do not get accessed frequently. Together with the application development team, work to limit increases in consumption that may occur when new application releases are rolled out. Set up a program to closely monitor changes in resource utilization in order to understand why they occur. These steps can reduce consumption of assets by up to 15 percent.

A further 15 to 20 percent reduction in the use of resources can be achieved by identifying, and balancing, differences in utilization by region, season, time of day, and line of business. Work with application owners and application development teams to identify the factors driving utilization and to develop measures for reducing consumption, including the renegotiation of service-level agreements, the restructuring of job networks, and the redesign of applications to run at peak efficiency. Again, devise a program...
Data center operators can take a variety of steps to save money on the energy side (see Exhibit 5). Of the various data center components, the CPUs consume the most energy. Ideas for reducing the amount of power consumed by CPUs include virtualization and the use of more efficient multi-core processors and processors with dynamic scaling. Gains can also be made in the area of power distribution by reducing the number of AC-DC/DC-AC conversion cycles and by converting to high-efficiency power distribution systems. Cooling costs can be lowered by the use of district cooling, heat pumps to reduce fan loading, desiccant cooling driven from waste heat, variable speed fans, and direct liquid cooling.

Exhibit 5
Data Center Operators Can Deploy a Number of Effective Measures to Optimize Energy Consumption

<table>
<thead>
<tr>
<th>Component</th>
<th>% Power Consumed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>41%</td>
</tr>
<tr>
<td>CPUs</td>
<td>41%</td>
</tr>
<tr>
<td>Power Supply Units</td>
<td>37%</td>
</tr>
<tr>
<td>Chilled Power Supply</td>
<td>20%</td>
</tr>
<tr>
<td>Uninterruptible Power Supply</td>
<td>12%</td>
</tr>
<tr>
<td>Voltage Regulators</td>
<td>8%</td>
</tr>
<tr>
<td>Server Fans</td>
<td>7%</td>
</tr>
<tr>
<td>Computer Room Fans</td>
<td>6%</td>
</tr>
<tr>
<td>Power Systems</td>
<td>4%</td>
</tr>
<tr>
<td>Water Pumps</td>
<td>2%</td>
</tr>
<tr>
<td>Total Improvement Measures</td>
<td>1%</td>
</tr>
</tbody>
</table>

Source: Data Center Energy Briefing, U.S. Department of Energy; Intel Corporation; Booz & Company analysis
THE EFFICIENT DATA CENTER

The typical data center faces a further challenge: the lack of truly efficient organizational structures and processes. The causes of inefficiency are many: Too many data centers find themselves focusing on day-to-day troubleshooting rather than on strong system architecture and design. Furthermore, data centers often move into production mode prematurely, before they have completed proper testing and deployment procedures. The result is a low degree of standardization in commodity operations activities and no clearly modularized service and product portfolios, which makes both sales and product management needlessly complex. Incoherent process routines and lack of fully transparent end-to-end service management often lead to the delivery of service levels over and above what has been agreed to (and is being paid for) by the customer—24/7 support, for instance, becomes the default setting.

Many of the causes of inefficiency can be attributed to organizational problems such as understaffed demand and capacity planning functions and the lack of an integrated operating model. After a careful analysis of its employees’ activities, one company running a midrange hosting operation discovered that employees were spending far too little time on planning and building out their systems, and far too much time on daily operations and ad-hoc troubleshooting. The result: Day-to-day operations struggled with a poorly integrated operating model—and efforts to standardize infrastructure and increase utilization were doomed from the start.

The consequences of a poorly organized operating model can be dire (see Exhibit 6). In the model on the left, the various functions of the data center, from hosting to storage to connectivity, are effectively siloed, with

Exhibit 6
A Cross-Platform Planning and Management Capability Can Improve Efficiency

• Current data center management is fragmented into many layers, with too many handoffs of core processes, such as problem resolution and operations and change management, between functions.

• By integrating service management and thinking in terms of “services,” not “servers,” data centers can achieve better capacity planning and management and lower administration costs.

Source: Booz & Company client example
each function running its own administration and service management and delivery. The resulting fragmentation creates the need for an excessive number of handoffs when problems occur, and any effort of the various functions to work together to change operating procedures becomes very difficult. Instead, planning, administration, and service management should be integrated across all the functions, as in the model on the right, allowing for better capacity planning and lower administration costs.

Can the automation of data centers help improve efficiency? That, of course, is the hope of every operator of data centers, especially as both process and management complexity increases. The desire to raise the efficiency of IT processes themselves is strong, with the goal of automating routine, labor-intensive tasks such as trouble ticketing, fault management, and performance management. And the number of automation tools is growing fast, as are the different configurations of these automation systems, and the move to virtualization will only add to that complexity. As long as each platform possesses its own management silo, moreover, each silo will look to automate its own platform operations.

Vendors are offering a variety of automation suites that can aid in the process of automation. Such tools already include configuration management functions such as automatic asset discovery and resource transparency and are beginning to offer dependency mapping and advanced configuration item reporting. New audit and control features can help with compliance and risk management tasks, and some suites offer the ability to manage workflow aligned with standard ITIL processes. The maturity of such suites remains a concern, however. Many of them still lack real depth and interoperability, and they do not typically cover critical areas such as storage area networks and other network functions. The market includes a number of niche players in such areas as server provisioning, migration to virtual machines, patch management, and storage allocation. The result: Automation efforts still require a patchwork of tools—BMC Patrol combined with MS System Center, for instance, or VMware vCenter Server and EMC ControlCenter. Buyer, beware.

One company running a midrange hosting operation discovered that employees were spending far too little time on planning and building out their systems, and far too much time on daily operations and ad-hoc troubleshooting.
GLOBAL DELIVERY

As corporations look to rationalize and save money on their overall data center footprints, the opportunity to offshore and nearshore a variety of data center services—and to farm out some services to third-party providers—continues to grow quickly. A well-planned and well-executed global delivery model for data center services can generate considerable savings, depending on which services are sent offshore, where they are sent, and to whom. Offshoring or nearshoring suitable “commodity” activities such as application management, database management, monitoring, and engineering will bring the greatest cost decreases—thanks primarily to lower salary and benefits costs, significant process improvements, and lower overall management costs.

Moving application management efforts to a combined onshore/nearshore model probably offers the greatest benefit—cost reductions of up to 22 percent—primarily because of the large reduction in labor costs. Manpower typically makes up fully 69 percent of the cost of onshore application management; moving to a combined model, however, can reduce labor costs to just 48 percent of the total. Cost savings can also be found in other remote management models, including mainframes, infrastructure, and storage (see Exhibit 7).

Exhibit 7
The Market for Offshore and Nearshore Services Continues to Gain Momentum, Offering Significant Cost Reduction Opportunities

RIM = Remote Infrastructure Management
Source: Booz & Company analysis
Companies have taken a variety of routes in their efforts to reap the cost and flexibility benefits of global sourcing. India has long been a prime region for such activities. One large data center operator, for example, turned to an Indian outsourcer for services that included 350 servers, 46 separate databases, 3,200 network elements, and 10 firewalls. In addition, the provider offered virus management, as well as backup and storage management. The benefits obtained included service levels of 99.95 percent, 24 hours a day, and higher productivity, as well as significant wage differentials.

Such benefits can be found closer to home as well, as regions such as eastern Europe gain expertise in running large-scale data centers. Looking to create a shared onshore/offshore infrastructure delivery model, one German corporation recently set up a nearshore subsidiary to provide its more commoditized second- and third-line services on a 24/7 basis, in addition to its onshore data center, where the tasks requiring a more highly skilled workforce took place. Again, the benefits came primarily in the form of lower labor costs. Wages were about 40 percent lower than those in Germany, and the average age of the nearshore employees was likewise lower. The client also found a strong talent pool surrounding its new facility, with adequate English and IT skills, thanks to the presence of nearby universities.

A third company, a large European industrial manufacturer, outsourced all of its data center operations, including remote management of more than 5,000 servers, to a provider in India. In addition to data services, the scope of the project included incident management, monitoring, and change execution, all provided on a 24/7 basis. The project began with just 50 full-time employees, but within three years that number was approaching 150. Again, the advantage in labor costs provided the greatest savings: Wages in India averaged less than half of those in the European headquarters. And the project also allowed the client to enforce a high level of process standardization and automation.

Manpower typically makes up fully 69 percent of the cost of onshore application management; moving to a combined onshore/nearshore model, however, can reduce labor costs to just 48 percent of the total.
Data center operators looking to offer their services to others on a commercial basis frequently face the same problem: Too often they don’t have the ability to conduct effective end-to-end capacity management and cost steering, or the means to provide transparency for costs and services. Typically, their catalog of services is too diversified, with every customer enjoying its own set of specific, personally configured services. The result is an overly complex set of offerings that is difficult to benchmark or rationally and transparently charge for. Making the transition to a demand-driven model will require significant changes in how data centers operate. They must move to a standardized set of limited platform products that can be strictly managed and maintained, easily compared with the offerings of competitors, and straightforward to cost out.

In the area where service and production issues merge, the problem is similar. Many data centers struggle to maintain a clear, logical connection between how services are produced and the revenues obtained from them. Without that connection, they cannot clearly link capacity planning and revenue projections, and they cannot determine clear targets for their costs of production, given what the market expects. Solving this problem requires simple, transparent connections between services and modular production building blocks and the capacity to understand the market and link it to target cost planning.

A further consequence of the typical collection of diversified services maintained by most commercial data centers is a lack of transparency regarding the total cost of the data center’s operations and poor governance in managing those costs. The mechanisms by which costs are allocated become overly complex, thanks in part to poor organizational alignment. Here again, the solution lies in introducing a cost-reporting mechanism that is simple, transparent, and based on total cost of ownership, and in realigning the organization into “production towers” that can help organize the process of cost reporting.
The capacity-on-demand model is clearly the future of data center operations (see Exhibit 8). Customers no longer want to pay for capacity they aren’t always using, and they don’t much care anymore about the specifics of the hardware and software being employed. In order to stay competitive, providers must move quickly to offer data center services that are scalable depending on the customer’s needs, and based more on computing capacity and performance than on specific hardware and software configurations. That means developing efficient new ways to deploy capacity and shut it down when not needed, to better balance total capacity usage, and to standardize both hardware and software platforms so they can scale up capacity quickly.

**Exhibit 8**
The Continued Move toward Capacity-on-Demand Models Will Force Data Centers to Rationalize Platforms and Services Offerings

These six areas in which data center operations can be enhanced offer data centers the potential for major improvements in their performance, and significant benefits in the form of reduced operating costs. Both corporate data centers and centers providing computing services to others should consider some or all of the improvements suggested if they wish to maintain their competitive position. The data center is rapidly moving toward a new model in which what matters is delivering as much computing capacity as customers need, when they need it. Will you be ready to give it to them?

**IN SUMMARY**

These six areas in which data center operations can be enhanced offer data centers the potential for major improvements in their performance, and significant benefits in the form of reduced operating costs. Both corporate data centers and centers providing computing services to others should consider some or all of the improvements suggested if they wish to maintain their competitive position. The data center is rapidly moving toward a new model in which what matters is delivering as much computing capacity as customers need, when they need it. Will you be ready to give it to them?
About the Authors

Stefan Stroh is a partner with Booz & Company in Frankfurt. He leads the global transportation technology practice and works for leading players in the international railway, logistics, aviation, travel, high tech, and consumer products sectors.

Dr. Germar Schröder is a principal with Booz & Company in Frankfurt. He focuses on IT strategy, large-scale transformation programs, and finance IT, primarily for the telecommunications industry. He also supports IT service providers in business model development, strategy, and operational efficiency.

Dr. Florian Gröne is a senior associate with Booz & Company in Berlin. He supports telecommunications companies and ICT Service Providers in developing their market positioning strategies and improving IT operations efficiency. He also works on CRM strategy and architecture across industries.
Booz & Company is a leading global management consulting firm, helping the world’s top businesses, governments, and organizations.

Our founder, Edwin Booz, defined the profession when he established the first management consulting firm in 1914.

Today, with more than 3,300 people in 58 offices around the world, we bring foresight and knowledge, deep functional expertise, and a practical approach to building capabilities and delivering real impact. We work closely with our clients to create and deliver essential advantage.

For our management magazine strategy+business, visit www.strategy-business.com.

Visit www.booz.com to learn more about Booz & Company.