Infrastructure & Operations: Top 10 Trends to Watch

David Cappuccio

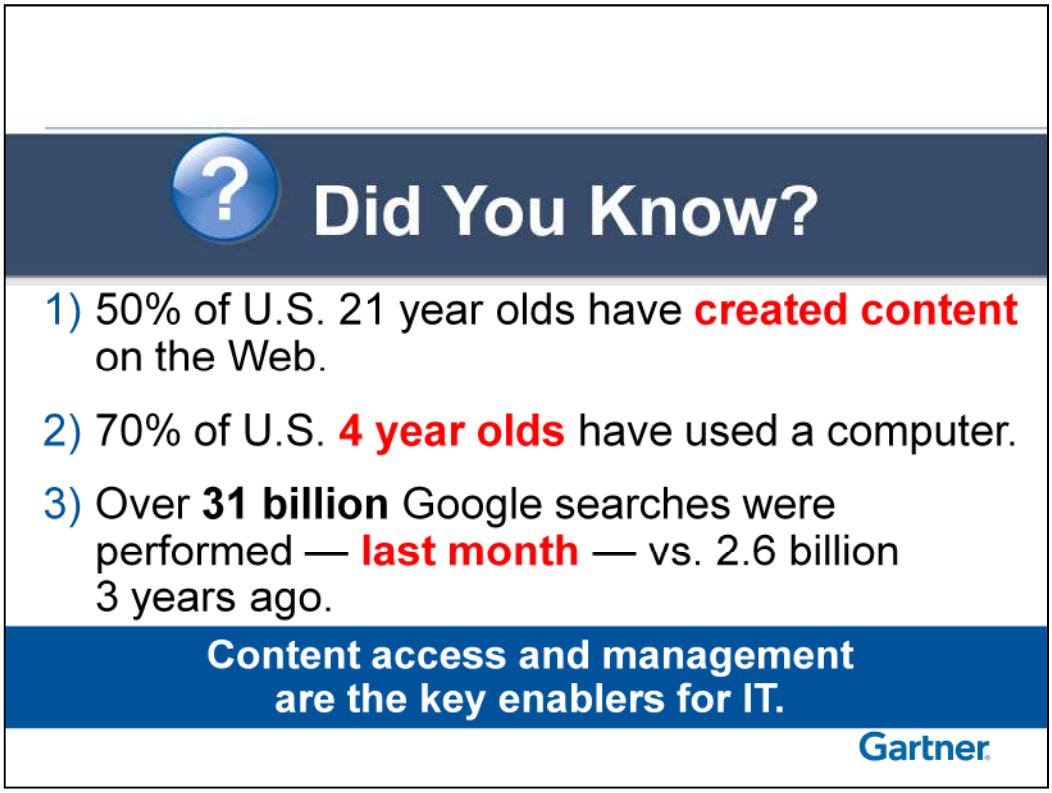
Gartner Symposium/ITxpo 2010

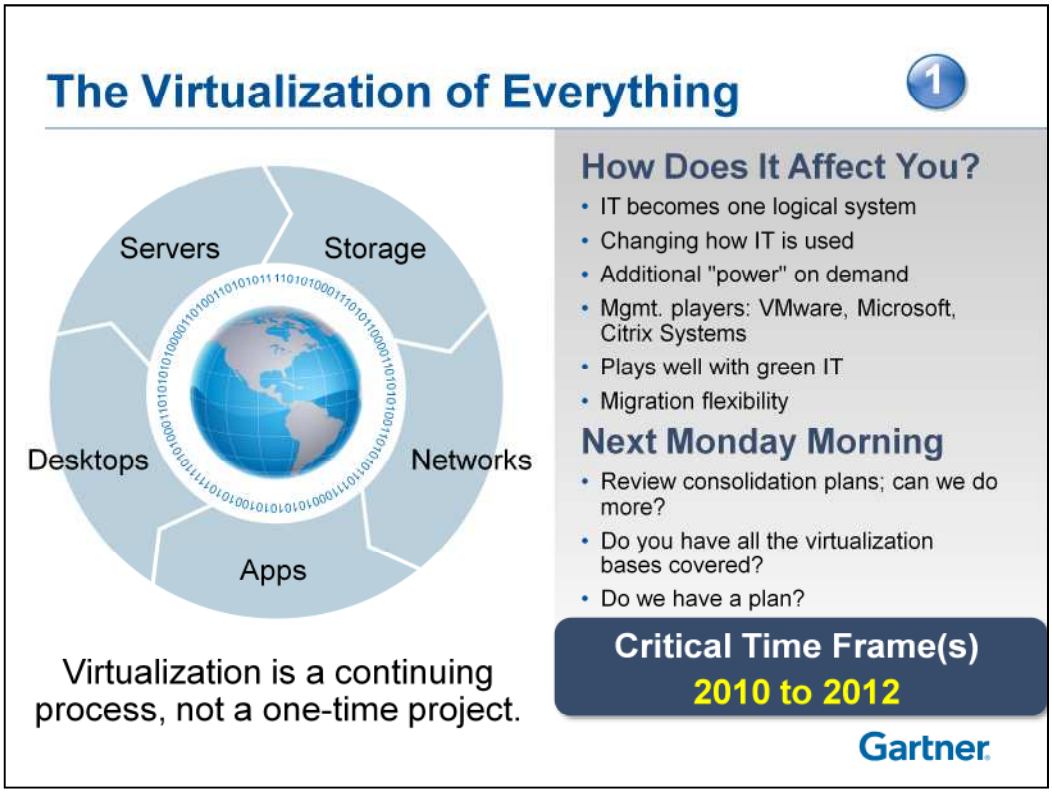
October 17-21, 2010
Walt Disney World Dolphin
Orlando, FL

Notes accompany this presentation. Please select Notes Page view.

Gartner

Ray Paquet





Virtualization is not simply a set of technologies buried in infrastructure. It also has important ramifications on business use of IT, and on business itself. Virtualization changes how IT is acquired, managed and used, and how software is priced and licensed. It will mean new forms of applications and it even becomes an enabler of cloud computing. Specifically, though, it can help lower the number of physical devices you require and help you make the most use of the ones you have. In addition to energy savings and lower capital expenses due to more efficient use of your hardware resources, you get high availability of resources, better management, increased security, and the potential of improved disaster recovery processes when you build a virtual infrastructure.

Virtualization can take several forms, from server to network to the broader domain of hardware and even desktop virtualization. Desktop virtualization creates a separate OS environment on the desktop. Employing the same approach as server virtualization, desktop virtualization lets you create separate virtual machines (VMs) on your Windows desktop, each of which virtualizes the hardware of a complete physical computer. The ability to deploy images virtually can increase the speed of deployment by a factor of approximately 30 times. Hence, the speed and flexibility of virtualization alone will likely make some form of chargeback mandatory, otherwise demand could skyrocket and more low-priority workloads could be deployed that don't justify their costs.

The result of virtualization will be to drive more companies to treat IT like a business. The danger during the next few years will be in following a specific vendor's vision, ignoring what other vendors are bringing to the table. It is unlikely that any one vendor's vision will prevail. Users should have their own visions of architecture control, and build toward it with a constantly updated strategic plan.

Ray Paquet

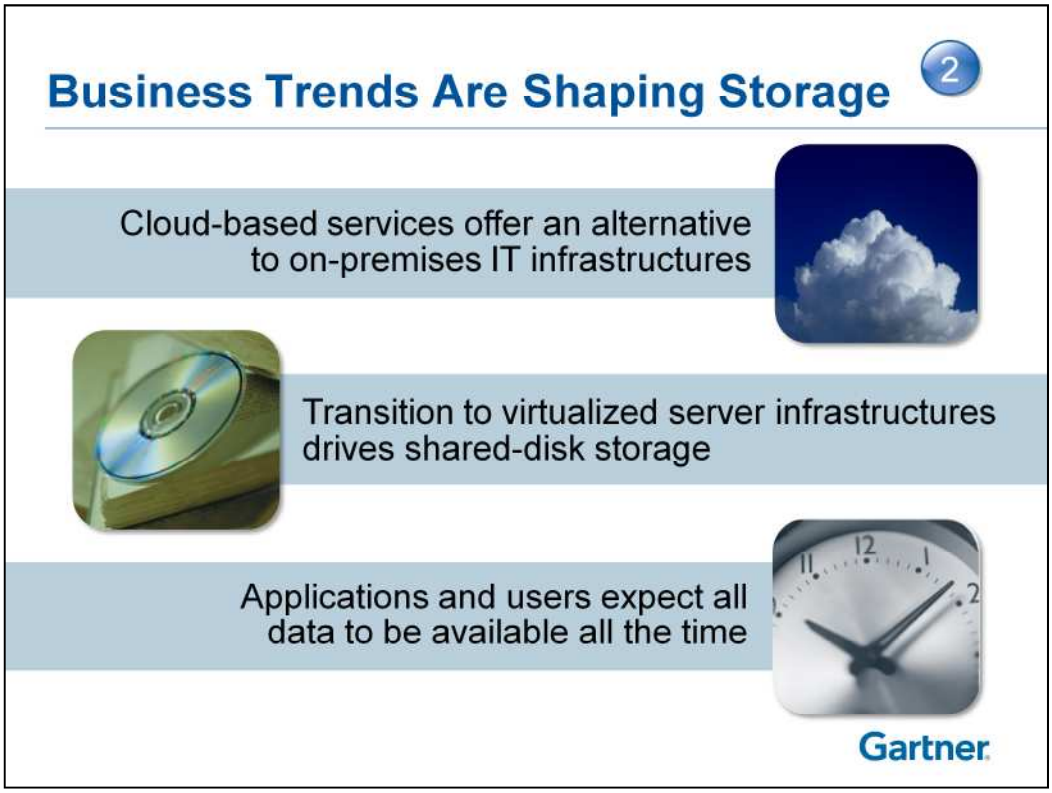
Strategic Planning Assumption: The number of virtualized PCs will grow from less than 5 million in 2007 to 660 million by 2011.

Key Issue: How will virtualization change technology?

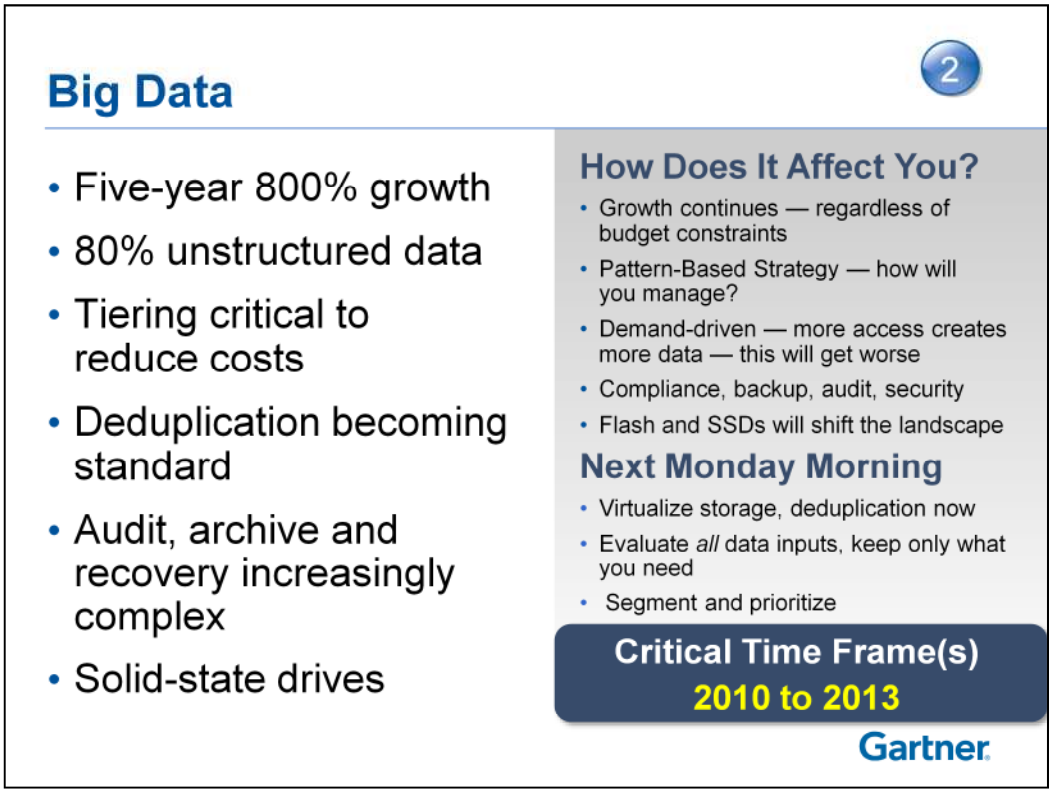
The key technology for defining new rules for device "footprints" is virtualization — a decoupling technology that breaks the close ties between hardware and software. A standard PC installation consists of a stack of multiple layers, the most important being hardware, the OS and applications. Because of the way these layers interact, the configuration of each is tightly coupled with the configuration of the layer below. This is the cause of much of the management complexity of today's PCs, because hardware changes regularly, which has a geometric impact on everything above. Virtualization breaks these dependencies, so the installation of each layer is independent of the configuration of the layer below. On the PC, it occurs at two levels: between hardware and the OS (machine virtualization), and between the OS and applications (application virtualization). On the PC, the impact of virtualization is to decouple the main functional layers. Application virtualization is gaining considerable interest, because key market changes are taking place. This type of virtualization is highly valuable for dealing with current PC management challenges, but it cannot help in the personal vs. computing argument. And although more immediately accessible to you, its long-term impact will be far less significant than that of machine virtualization, which is the focus of this presentation. This is the technology that will really make personal computing more manageable, flexible and secure, by enabling users to define multiple isolated footprints on the same device.

Ray Paquet

Massive content everywhere we look — and it all needs to be managed and stored (and archived). The default today is to keep everything, for fear of deleting the wrong thing.



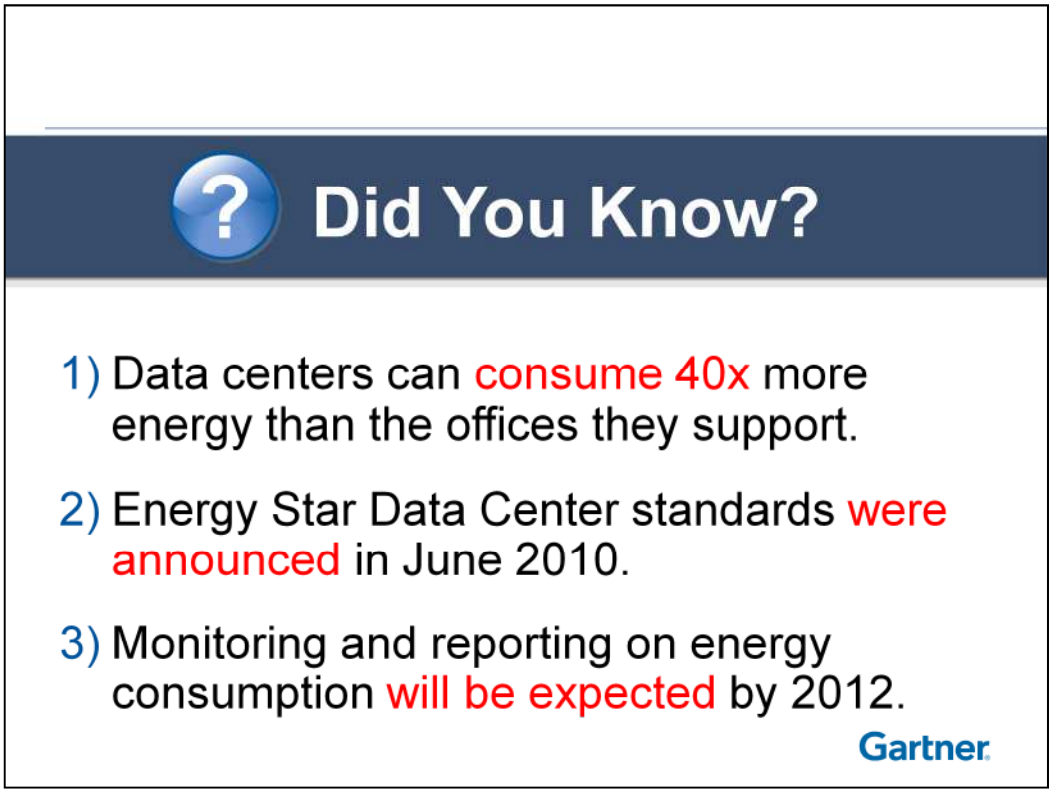
Storage continues to grow at an incredible rate, averaging a CAGR of 50% to 60% in most enterprises, but the fundamental changes happening now are more about storage access than storage growth. The advent of cloud-based alternatives has forced companies to look again at where storage is located, and how storage should be viewed from the business perspective. Is data a critical asset that needs complete 24-hours-per-day, 7-days-per-week availability, with robust security and auditability? If so, traditional methods of maintaining storage infrastructures remain in place. If, however, some data is deemed not so critical, or not required for immediate access, then storing it on a less costly cloud platform might be a good solution, assuming service levels can be agreed. Virtualization is also driving changes in storage with the realization that a virtualized server has very little benefit if its workload is tied to a direct-attached storage device. But with network-attached or SAN storage, the number of possibilities for workload management, business continuity and disaster recovery planning become much greater. Also, the rapid growth of storage is causing a new look at energy consumption, or the amount of kilowatts per access (or gigabyte) consumed. This is opening up new conversations about the types of storage used, and how a multitiered approach might give greater capacities overall, while reducing energy consumption and floor space. And, lastly, with the proliferation of smartphones and the subsequent extension of work hours to "whenever needed," employees, partners and consumers have adopted an expectation of "always available" for any information (data) they need, wherever they are. These three factors are changing how we design, install, manage, configure and extend our storage infrastructures.



A fact with which data centers have lived for many years remains true today: data growth continues unabated. From an IT perspective, one of the main issues is not awareness of the issue, but prioritization of the issues. We have spent so many years dealing with this, and surviving, that storage management projects are usually initiated from the ground up, rather than top-down, relegating many of these to "skunk works" status with little long-term funding.

Leading-edge firms have realized the problem and are beginning to focus on storage utilization and management as a means to reduce floor space usage and energy usage, improve compliance and BC/DR programs, and improve controls on growth within the data center. Now is the time to do this, because most of the growth during the next five years will be in unstructured data — the most difficult to manage from a process or tool point of view.

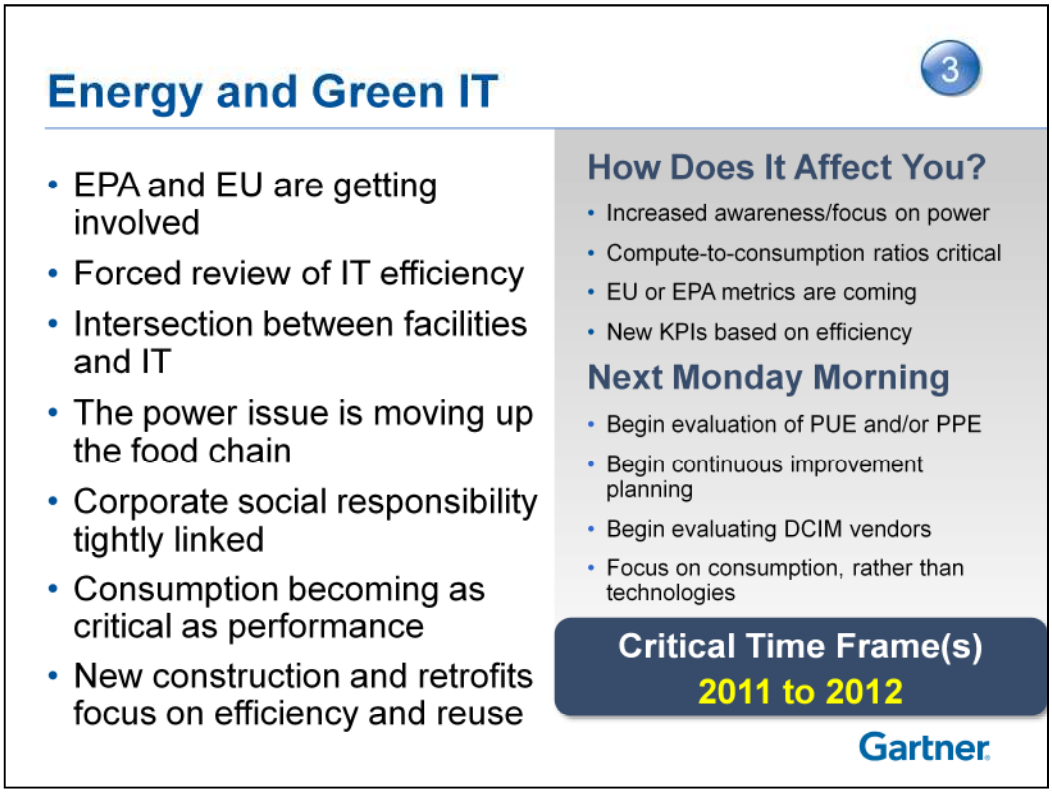
Technologies that will become critical in the next few years are in-line deduplication, automated tiering of data to get the most efficient usage patterns per kilowatt, and Flash or SSD drives for higher-end performance optimization, but with significantly reduced energy costs. NAND pricing continues to improve at a rapid pace, moving from $7,870 per gigabyte in 1997 down to $1.25 per gigabyte today — and this trend will continue.



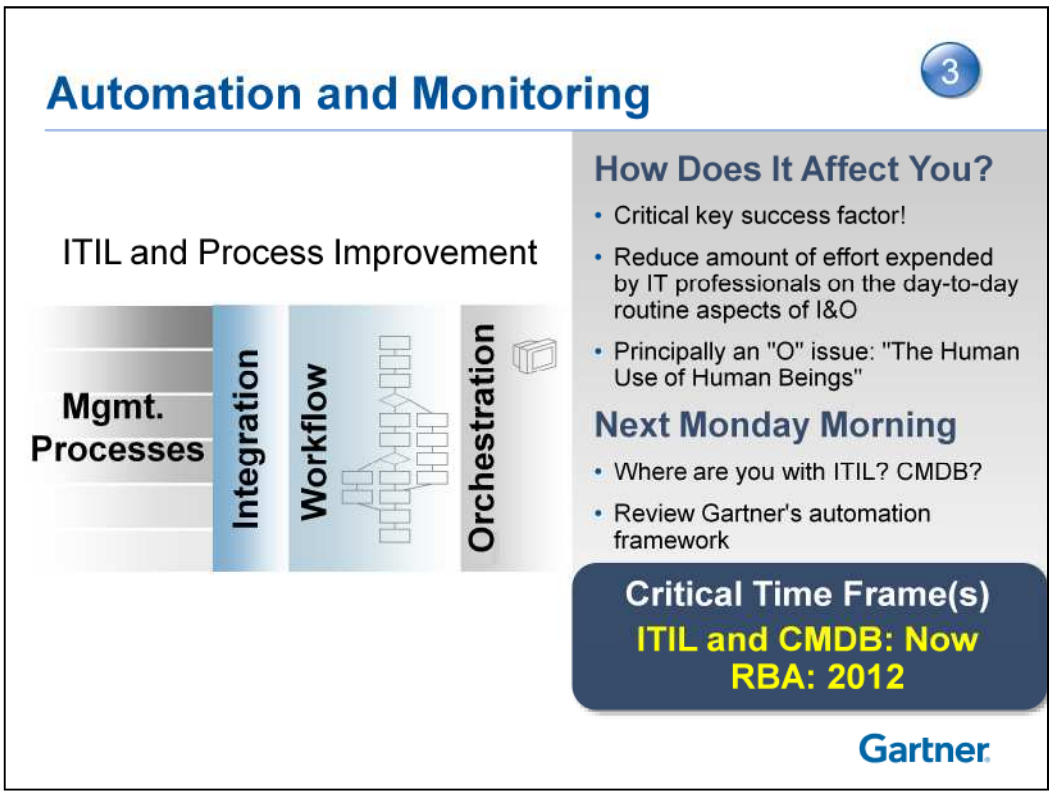
A little FUD from the EPA. Energy Star was announced in June 2010 for data centers. NetApp was the first vendor (public company) to be certified, with a 99 out of 100 rating. Most public companies with CSR programs will likely begin to follow — with continued escalation as peer pressure mounts.

As the trend toward consolidation and higher densities continues, companies are beginning to take a closer look at energy consumption relative to compute capacities. Historically, this didn't happen because the energy costs were usually buried in the facilities budget, combined with all other building systems, whether IT-related or not. However, with the increased attention given to power consumption, it has become apparent that many systems are highly underutilized. At low utilization levels, they use a high percentage of their total energy draw.

An average x86 server that is turned on, but idle, will draw upward of 65% of its name plate wattage, so if an IT organization doesn't have a clear inventory of what compute resources are doing what workloads, there is the potential for significant waste of energy. If that idle server were to have a 50% workload applied to it (via virtualization as an example), the incremental energy draw to move from 1% to 50% would be 10% or less, netting a significant improvement in compute power per watt than before. While this may seem inconsequential at the single device level, if applied across hundreds or thousands of servers, the savings can be dramatic.



With the increased awareness of the environmental impact data centers can have, there has been a flurry of activity around the need for a data center efficiency metric. Most that have been proposed, including power usage effectiveness (PUE) and data center infrastructure efficiency (DCiE), attempt to map a direct relationship between total facility power delivered and IT equipment power available. Although these metrics will provide a high-level benchmark for comparison purposes between data centers, what they do not provide are any criteria to show incremental improvements in efficiency over time. They do not allow for monitoring the effective use of the power supplied — just the differences between power supplied and power consumed. For example, a data center might be rated with a PUE of 2.0, an average rating, but if that data center manager decided to begin using virtualization to increase his or her average server utilization from 10% to 60%, while the data center itself would become more efficient using existing resources, then the overall PUE would not change at all. A more effective way to look at energy consumption is to analyze the effective use of power by existing IT equipment, relative to the performance of that equipment. While this may sound intuitively obvious (who wouldn't want more efficient IT?), a typical x86 server will consume between 60% and 70% of its total power load when running at very low utilization levels. Raising utilization levels has only a nominal impact on power consumed, and yet a significant impact on effective performance per kilowatt. Pushing IT resources toward higher effective performance per kilowatt can have a twofold effect of improving energy consumption (putting energy to work) and extending the life of existing assets through increased throughput. The PPE metric is designed to capture this effect.



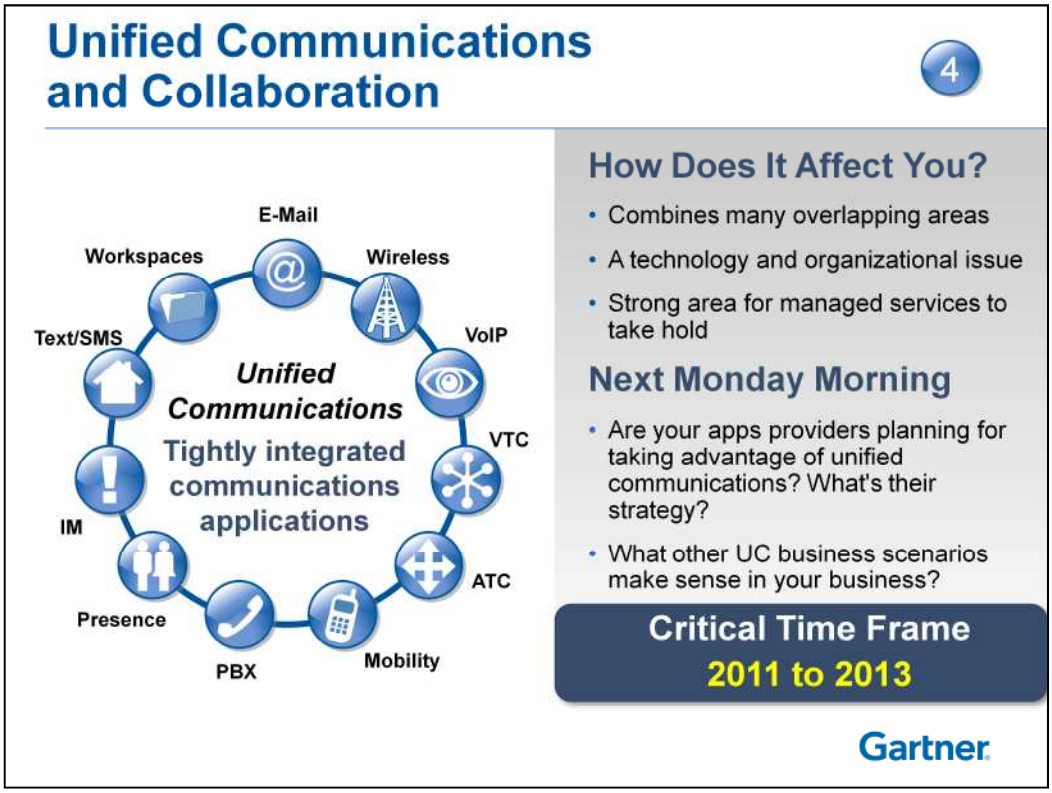
Just as a car company drives efficiency through parts standards, reusability and repeatable processes, so too must the IT organization. IT services are defined in conjunction with business requirements, but are predicated on architecture and standards. IT organizations that lack architectural standards (in infrastructure, software/applications, process and operations) will have greater costs and less agility than those with standards. Standards enable more rapid deployment because services are based on repeatable processes. Architectural standards help IT operations sustain a high rate of change with low error rates, because the behavior of the underlying systems is well understood, tested and predictable.

It is expensive, time-consuming and ultimately impossible to automate an environment with no architectural standards. If every system is different, you cannot achieve any economies of scale. Automation that is developed for one system cannot be leveraged on the others. Moreover, SLAs cannot be agreed to without understanding the architecture of the services provided. Finally, optimizing an environment that is not based on standards requires individual attention, rather than leverage across IT services.

Action Item: Define standard infrastructure and architecture building blocks, publish the standards and start requiring compliance with new projects first. Migrate established applications to new standard infrastructure over time on an opportunistic approach when updates are planned or funding is available.

Ray Paquet

IT is reacting to the massive growth and technology continues to keep pace. But will organizations and process do the same? The stat above is based on throughput of a Cisco Nexus 7000 switch.

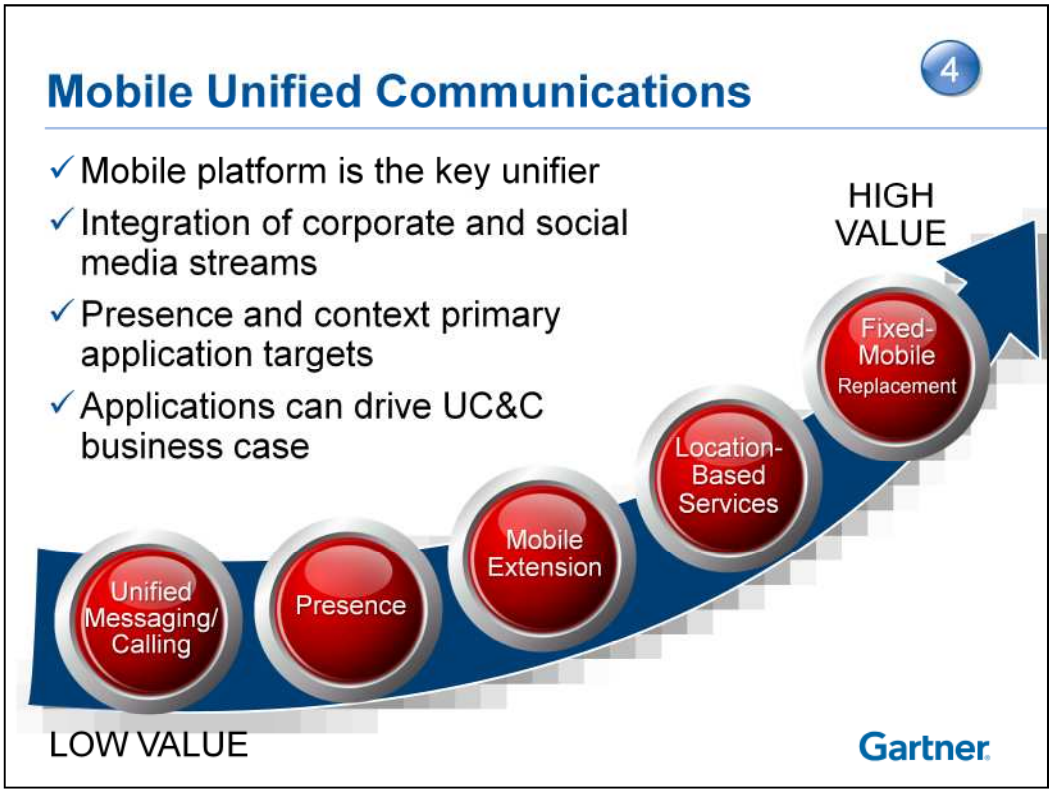


A different generation of workers is entering the workforce. For this generation, many different modes of communication options have always been present. For them, communications is something that meets their needs, not something to which they should adapt. Companies must develop a road map for migrating to unified communications (UC). First, planners should identify and inventory current products that are deployed in each of the key UC areas. Most enterprises have a mixture of vendors and products. These were originally acquired by different departments and groups, based on entirely independent sets of requirements. This phase will take time, but once an initial pass has been accomplished, the more important vendors can usually be identified. Then they should execute their plans with selected strategic partners in phased plans. The number of products for each key UC area should be reduced. Not all users will need the same level or type of UC support, so targeting functions is a critical part of the process.

Action Item: Use the strong competitive market and industry research to ensure the best solution and price.

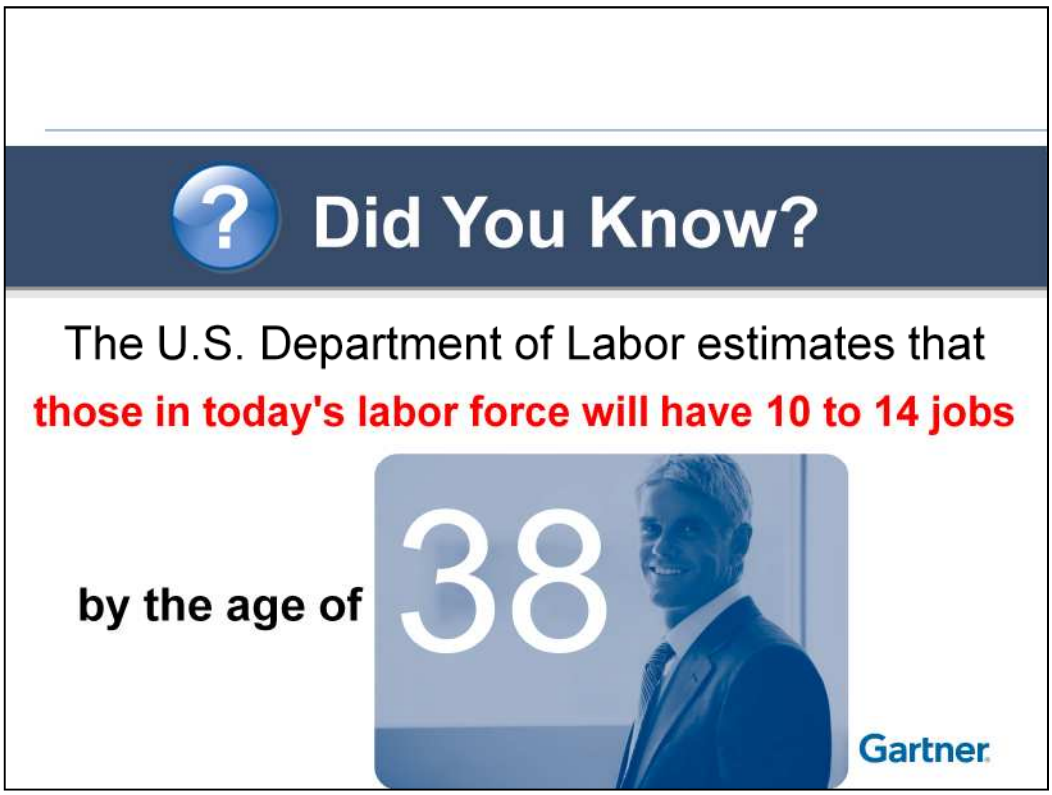
Ninety-five percent of communications for 18 to 25 year olds is via text. What are they saying? And what will IT do about it?

Strategic Imperative: Create a business plan that promotes the adoption and use of MUC technology aligned with the benefits of IT and support of corporate initiatives.

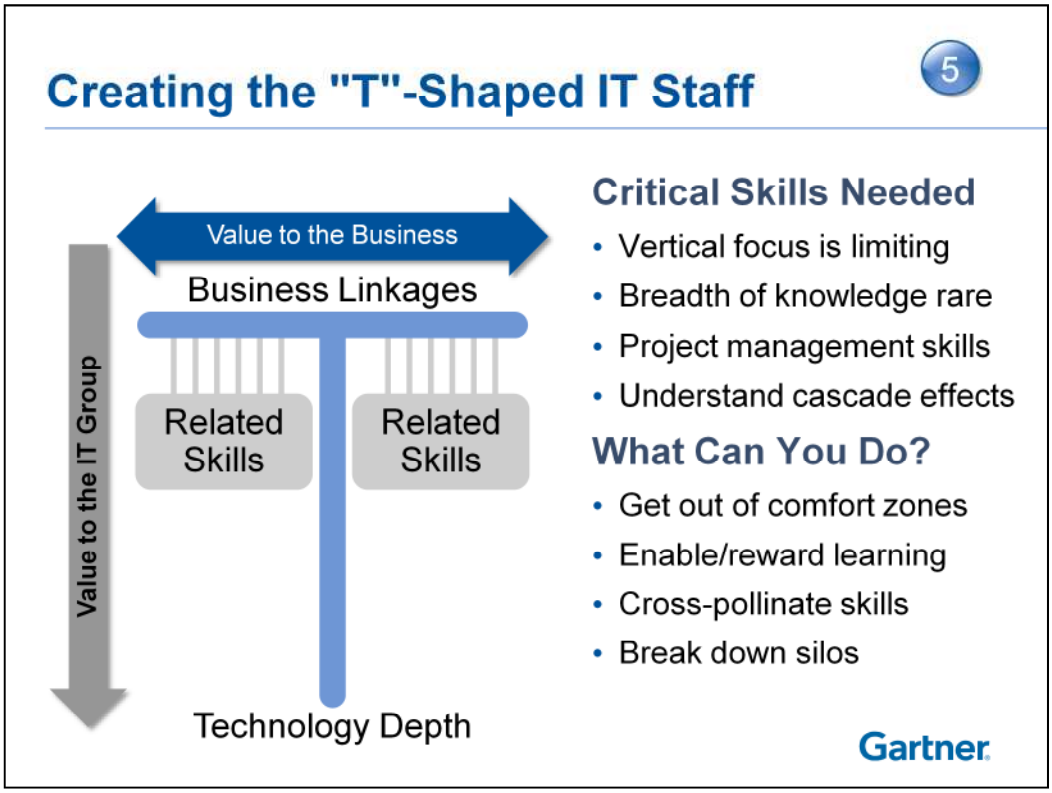


Mobility has always been a tough business case to make. It has a lot in common with UC. Most companies believe both are "a nice to have," and realize there is value to adopting them, but have a hard time quantifying the benefits or making a business case. By combining mobility and UC under mobile UC, there is a greater opportunity to define a business case and measure demand. Mobility can drive the business case for UC in the enterprise by providing clear benefits, namely incorporating mobile devices into the enterprise. Financial benefits will vary depending on the type of implementation, the type of mobility and the calling used. Adding capabilities such as presence can start to add value. Extending desktop phone capabilities to mobile phones increases that value, and advanced features such as location-based and context-aware services continue to move it up the scale. These capabilities can help save money, as well as increase productivity. In the end, the greatest impact will be by full fixed-mobile replacement, which will help consolidate devices and services, as well as offer least-cost routing and connection to private networks to help reduce cellular calling costs.

Action Item: Identify which mobile requirements can help drive your business case. Identify key mobility constituents and develop a strategy for MUC support based on critical mobile communications needs.



How do we motivate new employees? How to create loyalty to the company (or just to the job).

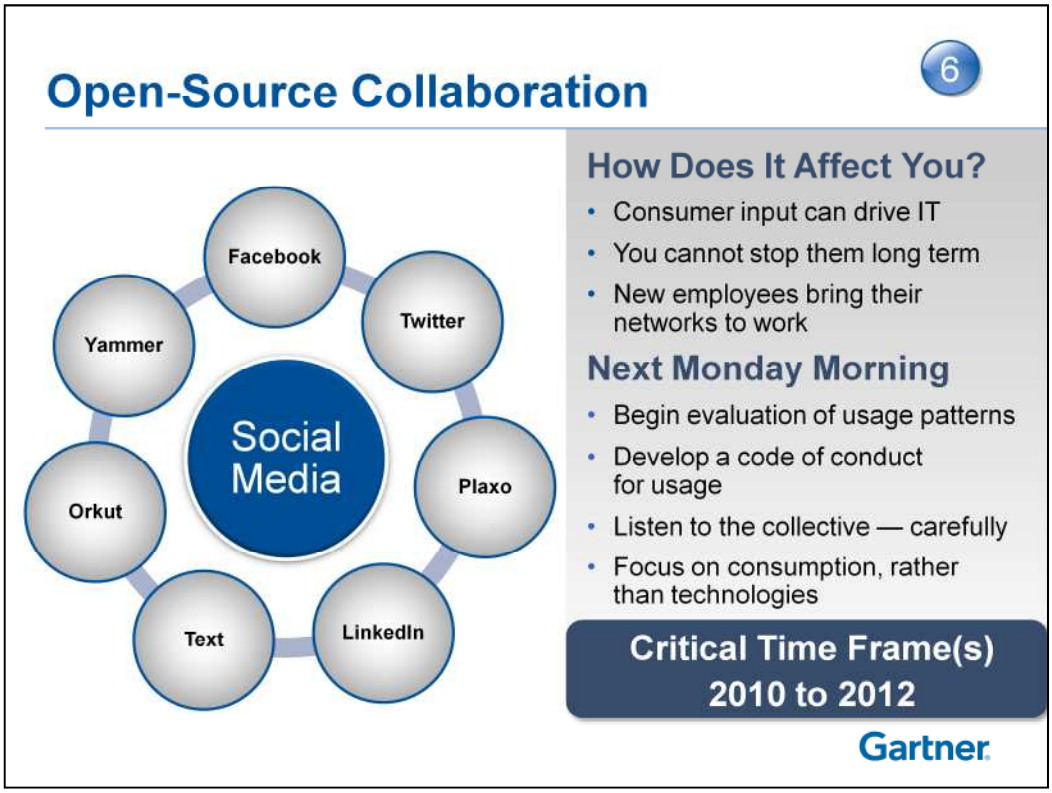


If you look at any employee within the IT group, try to imagine them as represented by a capital letter "T," where the center post of the T represents their depth of knowledge in their primary skill set. Some people can drill down that center post to a frightening level of detail, but, in a sense, the depth of the T represents their strength as a technologist, and, in most organizations, it's the key to their credibility in the organization, and to their success within IT. It's their core. The crossbar of the T is not about technology per se, but about the relationship between converging technologies (and the business) and how much that employee understands about these relationships. As you increase the number of relationships (or technology linkages), the broader the crossbar on the T becomes. Now sit back and think about the most valuable people in the IT organization — those people who always get the projects handed to them, because we know they'll get things done, regardless. If you look at the strengths they bring to any project, it's rarely depth of knowledge (the vertical T), but breadth of understanding (all the linkages). Drill-down specialists are available in all disciplines, but the linkage masters are hard to find.

This then becomes a motivational tool for organizations. Let's begin to recognize what we really value in IT — not depth (except in junior people), but breadth across multiple disciplines *coupled* with depth in a primary discipline. It's difficult to train people with this knowledge, except through real-world experiences with a business context — your business, and your projects. In some cases, it may take forced change in disciplines, but force the change in a related discipline with the idea of expanding knowledge and creating those linkages. The most effective IT people are always looking for new things to learn, and, in many cases, the most interesting areas are in the unknown, not the known areas. Enabling this learning — even creating incentives for it — is a critical success factor as we move toward fully virtualized environments. And when employees realize that their value is not only how much they know in a discipline, but how much they understand the linkages between disciplines, IT as a whole will become a much stronger organization, and more able to adapt to these changing environments.

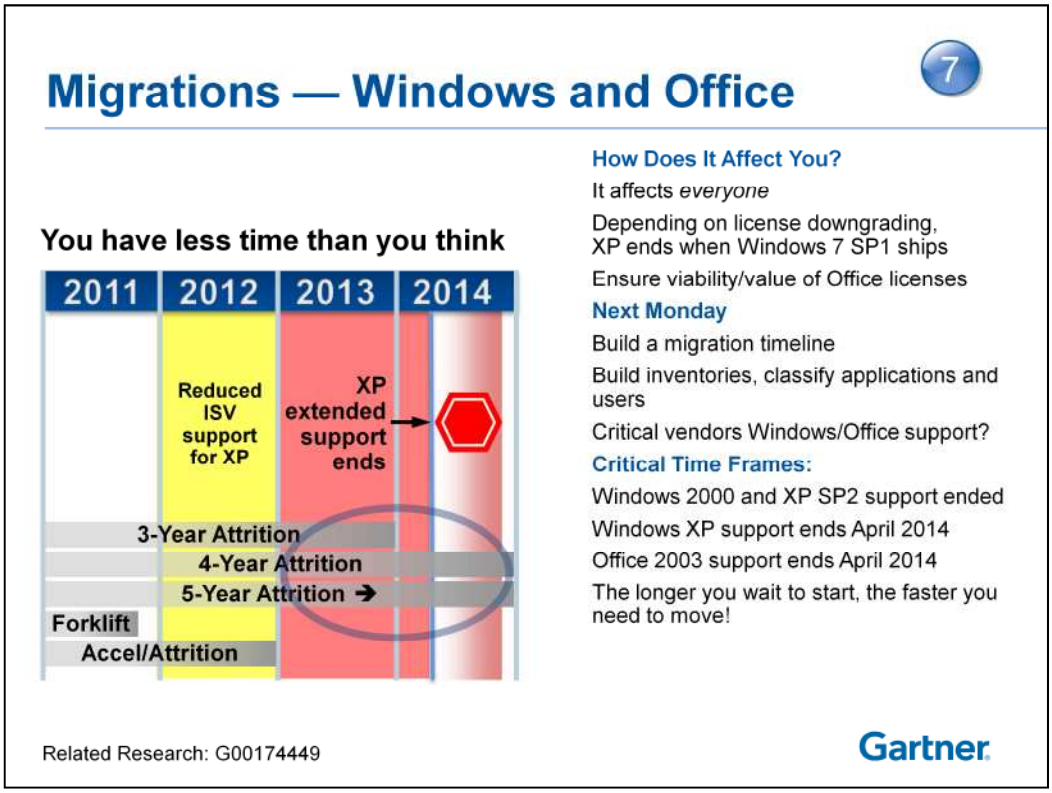
Ray Paquet

Social networking is a force that continues to grow. Its impact on society will have a direct impact on business. Facebook and Twitter now get more daily searches than Google.



Not only do you unify communication internally, but outside social networks are also impacting internal communications, especially with the next generation of employees. Get ready for it. Social software is a broad range of technologies that can be divided into four major categories. **Social networking** covers social profile management products, such as MySpace, Facebook, LinkedIn and Friendster. It also covers social-networking analysis technologies that use algorithms to understand and use human relationships for the discovery of people and expertise. **Social collaboration** covers technologies such as wikis, blogs, IM, collaborative office, crowdsourcing and virtual worlds. **Social media** includes technologies that assist in sharing content and building expansive content repositories. YouTube and Flickr fall into this category. **Social validation** gets the community involved in assessing (products, content, people). Examples include social rating, social commentary, social tagging and social bookmarking. Digg and Delicious fall into this category. Consider adding a social dimension to a conventional website or application. Analyze how users interact with the site, define engagement metrics, and instrument the system accordingly. Survey business processes, and categorize them as "structured" or "unstructured." View unstructured processes as candidates for deriving business value from social sites and social platforms. Analyze social platforms for the potential to support them. When building a social site or Web application, you operate in the shadow of giants (Facebook and Google-led OpenSocial consortium). Thus, your long-term strategy must include coexistence or collaboration with these platforms. Top-down is about business value, organizational goals and user needs. Bottom-up is about core entities (e.g., people and their roles and relationships). Resting on this social substrate is the evolving vocabulary of social interactions and gestures. This dynamic space is just emerging in the enterprise.

Strategic Planning Assumption: Plan to be off Windows XP by YE12 to avoid ISV support issues.



Key Issue: How long can you run your current software and what are your options, including ones from vendors other than Microsoft?

Microsoft will support Windows XP with security fixes (the support deadline that worries most organizations) into April of 2014. However, past experience has shown that ISVs will stop testing new versions of their software on Windows XP as early as 2010, and this will be a common problem by 2012. That means that new releases of critical business software may require Windows Vista or Windows 7 long before Microsoft support for Windows XP ends.

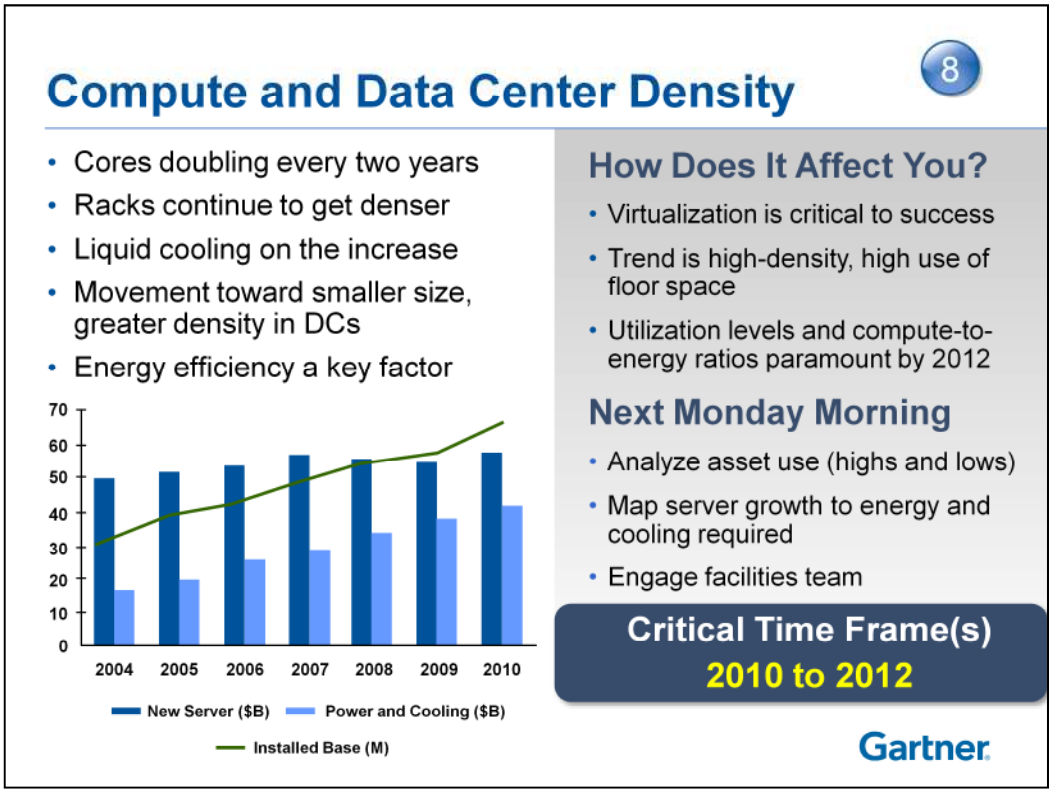
Therefore, to be conservative, we would suggest that organizations plan to have all their users off Windows XP by YE12. This provides two benefits: (1) it ensures that new versions of critical business applications will run on most users' PCs by 2012 and 2013; and (2) if you run late with your project and don't make your YE12 goal, you have a buffer of 15 months of continued security patches for Windows XP. Organizations that require extended support from Microsoft to get security patches beyond April 2014 will likely have to pay a minimum of $200,000 for the first year of custom support, money which would be better spent moving off Windows XP and onto Windows 7.

Because most organizations will not be able to begin a migration to Windows 7 before early 2011, that provides two calendar years (2011 and 2012) to do the migration.

Action Item: If you have not done work on Windows Vista by now, you're skipping it. Start your Windows 7 work to get XP out by YE12.

Ray Paquet

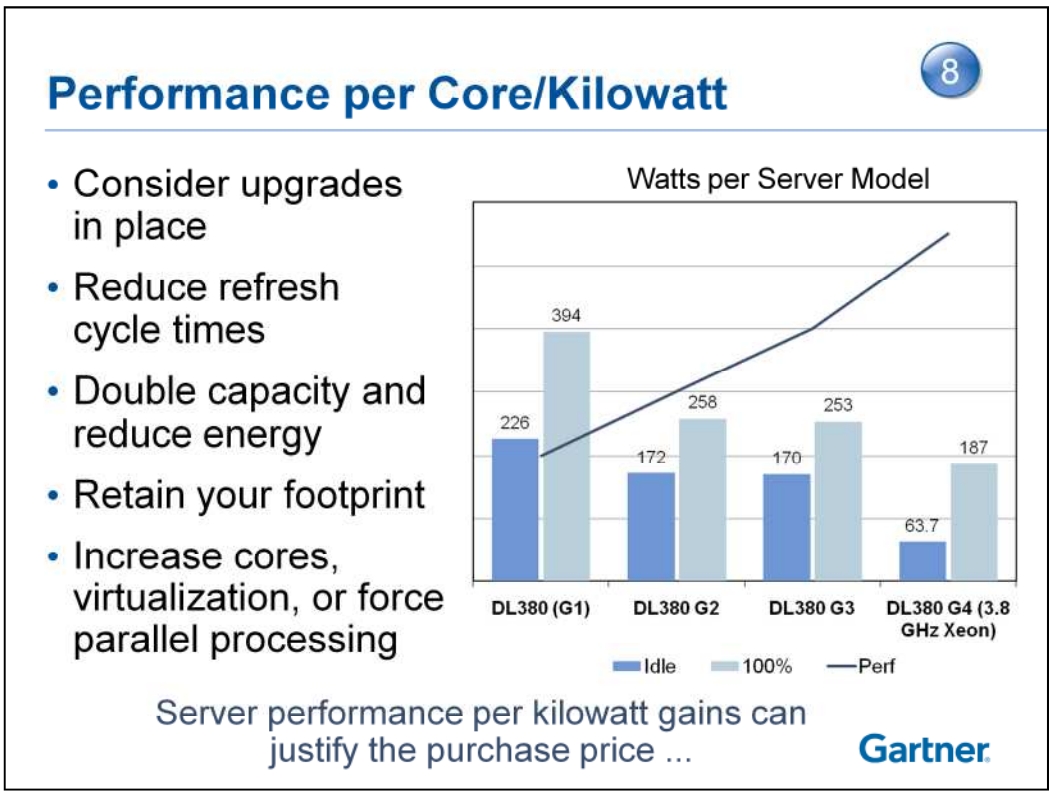
SSD and NAND technology will continue to be the driver of performance, capacity and energy efficiency gains across the IT spectrum.



The system density issue will not go away. In today's economy, the number of large capital budgets allocated to building new data centers has shrunk, forcing many companies to consider scaling their data centers vertically (via density), rather than building new ones. This movement will put even greater emphasis on getting more out of servers, while severely taxing the power and cooling infrastructures of existing sites.

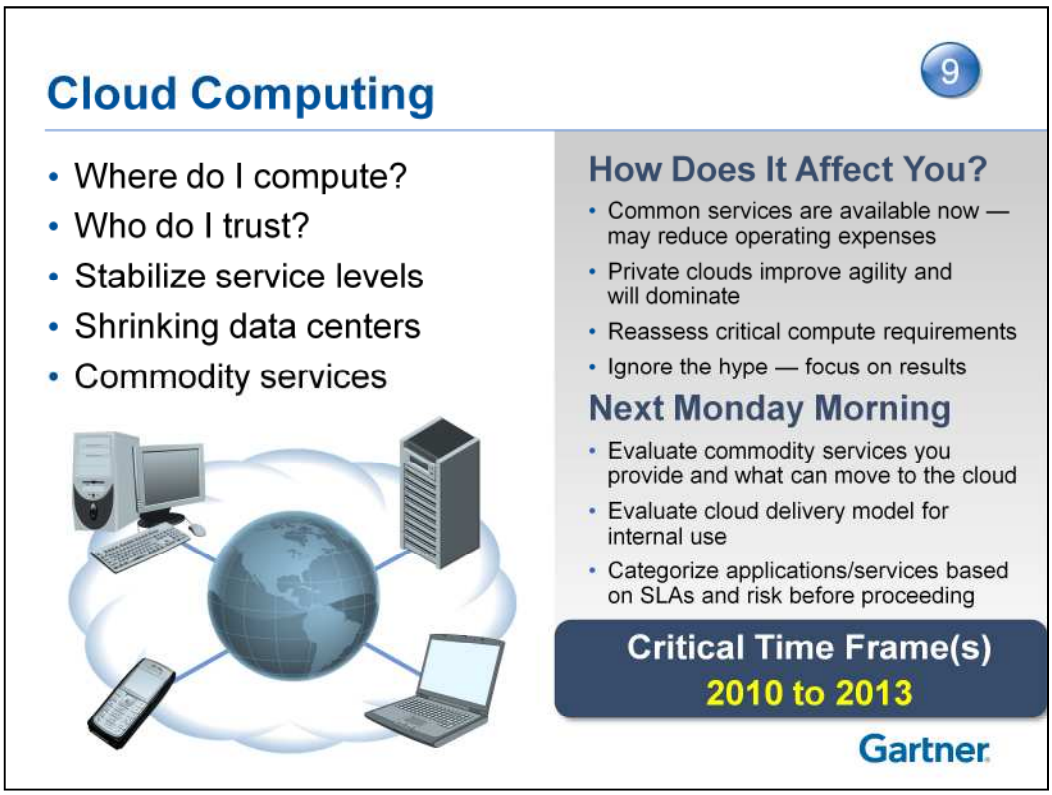
Virtualization is one of the most critical components being used to increase densities and vertically scale data centers. If used wisely, average server performance can move from today's paltry 7% to 12% average to 40% to 50%, yielding huge benefits in floor space savings, energy savings and agility (for example, provisioning speeds). Two factors that need to be considered are the number of cores per server (four- and eight-core systems are becoming common, and 16 cores will be common within two years), and overall energy consumption trends in the data center. The core issue will have to be addressed within IT operations from a performance/licensing point of view, and from applications during initial design phases of new projects, where old methods of linear coding will not be effective. Parallel coding techniques need to be addressed soon, and starting the training process early is a wise move. Also, public companies will soon have to track energy consumption and efficiency metrics at the data center level, so IT managers should begin tracking performance and consumption levels in more detail, in preparation for this change.

Ray Paquet



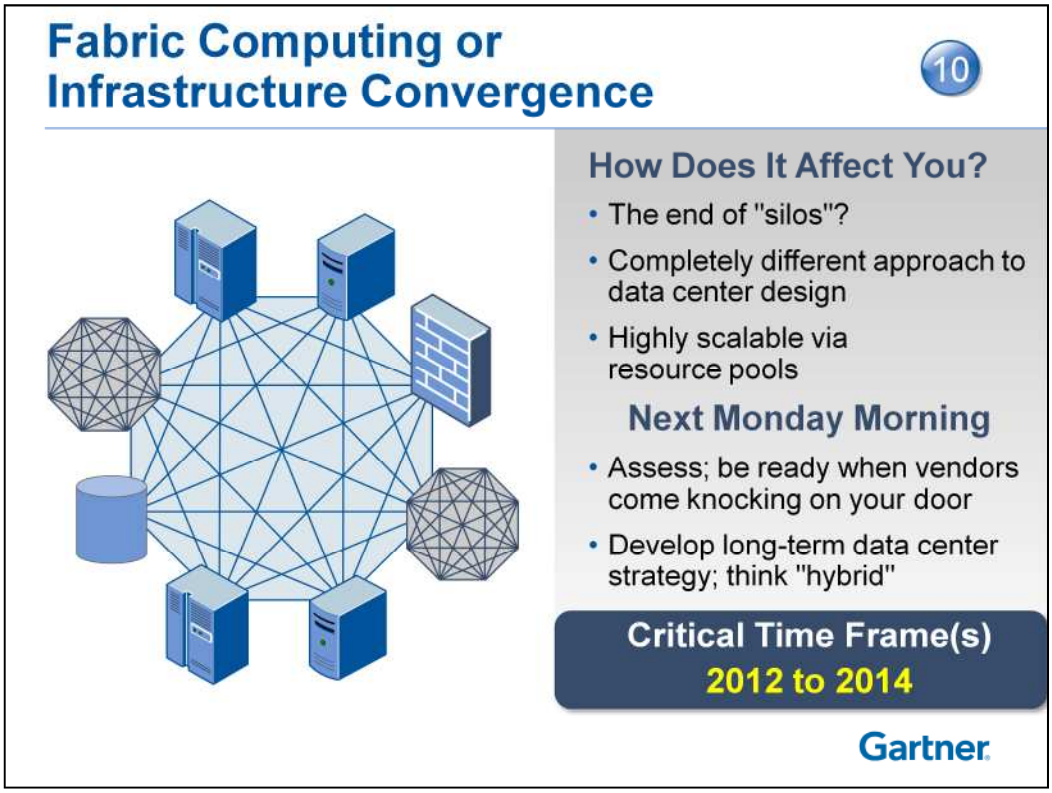
Server power consumption is dropping, and power efficiencies are increasing. Today's Nehalem processors, at full load, consume less energy than a five-year-old server at idle — with 10x the performance.

Traditional asset life cycles for servers have averaged between four and five years, with many companies keeping older servers around even longer for use in less mission-critical (or compute-intensive) environments. While this was often done to reduce capital costs in replacements, the cure was often worse than the disease, as the energy consumption of these older servers was significantly more than newer ones. One alternative view to this type of asset life cycle management is to look at Moore's Law and apply it to upgrades in place. As with all technology changes though, there are often hidden cascade effects. In the case of increasing processor and core density, the question that needs answering is, "How best to use them?" While some enterprises will continue the virtualization push, others are beginning to realize that business-critical applications designed for x86 may need to be rewritten to take advantage of four-core (and greater) systems. This will reintroduce the concepts of parallel processing and parallel development methodologies into the AD Teams, while creating its own series of cascade effects across the IT group. The requirements of parallel processing in applications will drive significant change with IT development teams, both from a skills perspective and core methodologies. Senior staff may not be able to adapt as quickly as needed, which will force a new look at acquiring outside talent to augment the staff.



The hype surrounding cloud computing grew exponentially during 2009. Vendors increasingly use cloud computing as a marketing label for many old technologies and offerings, devaluing the term and trend. Although cloud computing is a natural evolution of various enterprise and Web-based technologies and trends, it is a mistake to simply relabel these older technologies as "cloud computing." Cloud computing emerges from the synergistic intersection of select elements of these trends and technologies. This new computing model drives revolutionary changes in the way solutions are designed, built, delivered, sourced and managed. Cloud computing isn't defined by one product or technology. It's a style of computing that characterizes a model whereby providers deliver varied IT-enabled capabilities to consumers. The key characteristics of cloud computing are: (1) delivery of capabilities "as a service;" (2) delivery of services in a highly scalable and elastic fashion; (3) using Internet (by extension, Web) technologies/techniques to develop and deliver services; and (4) designing for delivery to external customers.

The final item warrants examination. Cloud-computing services are designed based on an assumption that the service provider will deliver capabilities to third parties that act as consumers. Cloud computing is heavily influenced by the Internet and vendors that have sprung from it. Companies such as Google deliver various services built on a massively parallel architecture that is highly automated, with reliability provided via software techniques, rather than highly reliable hardware. Although cost is a potential benefit for small companies, the biggest benefits of cloud computing are built-in elasticity and scalability, which reduce barriers and enable these firms to grow quickly. Web-centric startup firms such as SmugMug have leveraged Amazon's EC2 and S3 as a foundation for their computing environments. As certain IT functions industrialize and become less customized (e-mail, for example), there are more possibilities for larger organizations to benefit from cloud computing.



As a working definition for IT organizational planning, we define infrastructure convergence as: the vertical integration of server, storage, and network systems and components with element-level management software that lays the foundation to optimize shared data center resources efficiently and dynamically. Converged infrastructures enable such building blocks to be assembled, integrated, connected and repurposed with interoperability in physical, logical or virtual domains. It represents a critical connectivity layer that has the potential to enable speed and agility in provisioning, configuration and repurposing. One type of implementation of such convergence is known as fabric computing (see "The Impact of Fabric-Based Computing on the Server Market" [G00166921]).

Gartner proposes that IT organizations use a model that calibrates a degree of convergence as expressed in the definition. Comprising a five-point scoring system, the model presented in this research will help users assess the level of efficacy of convergence initiatives. From such criteria, we anticipate developing a maturity model to guide users toward a staged and incremental approach to optimize and modernize an IT infrastructure. We do not believe that convergence must necessarily be viewed as the be-all and end-all of IT infrastructure planning. IT organizations that take a wait-and-see approach to convergence may be losing valuable time. It is possible to be cautious and proactive. An ad hoc sandbox could be used to test against current pain points as a proof of concept. The vendor PR effect will be in high gear in 2010, and users need a systematized approach to unravel the reality from the hype. The model provided is one important step to aid the decision process and create a collaborative approach with the vendors.

The message is clear — what we've seen, which has been pretty incredible, is nothing like we will see in the next generation.

The message is that although IT continues to change at an ever-increasing rate — and I&O will have to deal with it — the pace will continue to escalate and we all need to get ready for it, from an infrastructure, process and organizational perspective.

Source: Ray Kurzweil, "The Singularity Is Near."

Recommendations

- ✓ Take virtualization to the next level — focused on all facets of IT.
- ✓ Scale vertically first, horizontally second. The new KPI will be about compute per kilowatt.
- ✓ Staff skill sets will reach a crisis point if not addressed soon. Look again at incentives and motivational techniques.
- ✓ Social networks will be an impact on all enterprises — get ready for it.
- ✓ Fabric-based systems are the future — eventually.

Gartner

Your Action Plan

CIOs and enterprise architects should …

- **Monday Morning**
  - *Assess* all projects and their relationships — look horizontally.
  - *Prioritize* based on risk, reward and long-term impact.
- **Your Next 90 Days**
  - Look again at capacity planning — without presumptions.
  - Assess *staff skills* based on breadth of knowledge and value to the business, not just value to IT.
- **Your Next 12 Months**
  - *Review* opportunities to converge networks, infrastructures and skill sets.
  - *Establish* impact on other business plans.

Gartner.

Related Gartner Research

- **"Microsoft's Desktop Virtualization Strategy Is Becoming Clearer"**
  by Mark A. Margevicius (G00175894)
- **"Understanding 'Client in the Cloud'"**
  by Stephen Kleynhans (G00174116)
- **"IT Infrastructure and Operations Key Initiatives, 2010 and 2011"**
  by Jay E. Pultz (G00174909)
- **"Grow Disk Storage 800% or More, Without Increasing Power or Cooling Costs, in the Same Space"**
  by Dave Cappuccio (G00175118)

Gartner