White Paper

Intel Information Technology

Computer Manufacturing

Data Centers

Transforming a Global Data Center Environment

Intel IT has embarked on an enterprise-wide, eight-year strategy to transform our global data center environment, increasing efficiency and substantially reducing cost. Our strategy focuses on standardization, increasing utilization, and reducing the number of data centers. We plan to consolidate data centers and accelerate server refresh, as well as employ grid computing virtualization and energy-efficient computing. We are also optimizing storage, networking, support services, and IT governance. Ultimately, our goal is to run our data center environment much as we run our semiconductor factories, with similar high levels of efficiency and responsiveness to business needs. We expect to achieve overall program savings of USD 1 billion or more.

Data Center Efficiency Core Team, Intel Corporation

January 2008

IT@Intel

Executive Summary

Our goal is to run our data center environment much as we run our factories, with similar high levels of efficiency and responsiveness to business needs.

Intel IT has embarked on an enterprise-wide, eight-year strategy to transform our global data center environment, increasing efficiency and business responsiveness, while substantially reducing cost. We expect our strategy to achieve a nominal cost savings of about USD 1 billion (net present value of USD 550-650 million) through virtualization, standardization, and process improvements. Our goal is to run our data center environment much as we run our factories, with similar high levels of efficiency and responsiveness to business needs.

Our strategy focuses on standardization, increasing compute utilization, and reducing the number of data centers. Key elements include:

- Consolidating data centers, concentrating computing resources into fewer large, highly efficient hub data centers, and closing other local data centers
- Accelerating server refresh, while implementing grid computing and server virtualization to increase utilization, and optimizing storage and networking
- Standardizing processes and design specifications by using a strategic, long-range planning approach, including a reduction in the number of server and storage platform reference designs.
- Using green computing with power-efficient servers and energy-saving data center design, as well as developing our commitment to industry initiatives

We have already begun to achieve benefits in several of these areas, including grid computing, server consolidation, data center consolidation, and long-range planning. Our use of virtualization and grid computing reduces energy consumption and boosts utilization, enabling us to share compute servers by removing physical, geographical, and organizational boundaries and allowing Intel to avoid more than USD 30 million in 2007 capital purchases.

Since 2006, we have seen an 11 percent increase in server utilization, which will allow Intel to save approximately USD 77 million in 2008 capital purchases.

Ultimately, our goal is to run our data center environment much as we run our semiconductor factories, with similar high levels of efficiency and responsiveness to business needs.

Contents

Business Challenge

Intel's semiconductor factories are large, highly efficient, and fine-tuned for high-volume production. Our "Copy Exactly" strategy enables us to quickly implement nearly identical processes in each factory, helping us respond quickly to changing business needs by bringing new factories online or moving production between factories.

Intel IT aims to implement similar efficiencies, economies of scale, and business responsiveness throughout our data center environment.

Our data centers are at the heart of Intel's massive worldwide computing environment, which includes more than 240,000 network nodes and carries an average 137 million e-mail messages and 166 terabytes of WAN traffic each month.

Business needs are growing rapidly, increasing the demands on our data center resources. Server requirements for semiconductor design, which account for most of our existing server resources, are expected to grow at 15 to 23 percent a year. Storage is growing at an average of at least 40 percent a year. Server requirements for enterprise, office, and manufacturing applications are also increasing steadily.

Though our data centers have effectively supported Intel's requirements, we have lacked an overall long-term strategy to plan data center investments and capacity across the entire environment. Instead, our data centers have grown over time, largely due to acquisitions. As a result:

- The number of data centers has grown to 130, spread over 74 sites in 31 countries. Many of these data centers are small; however, our experience has shown that we can build large, high-density data centers that are more cost-effective.
- Many data centers face cooling, power, or space constraints. To overcome these, we have sometimes implemented tactical solutions to unforeseen problems, such as a need for more cooling or uninterruptible power supply (UPS) capacity. This fix is often more expensive than strategic long-term investments.
- Our four main computing application areas—design, office, manufacturing, and enterprise (DOME)—operate largely as silos, hindering our ability to flexibly share and allocate resources across multiple computing areas. For example, many data centers are dedicated to design computing.
- Storage utilization is only around 50 percent, largely because storage is managed independently for specific groups or applications. We also lack storage lifecycle management that could help reduce cost by transferring less-utilized or less-critical data to lower-cost media.
- It can be difficult to plan capacity because we sometimes lack clear, consistent demand signals from business groups.
- Different groups have met application needs by creating new reference designs (hardware, operating system, and application configurations) for computing platforms. As a result, we have accumulated nearly 400 different reference designs.

We realized that we needed a long-term, strategic plan for managing our data center investments and capacity. Our goals are to run our data centers more like we run our factories, concentrating computing resources into large, efficient data centers, implementing efficient strategies and processes across our environment, and responding more quickly and flexibly to business needs. We expect that this will significantly reduce cost.

Strategy

In 2006, we formed a worldwide team to create a long-term plan for our data centers. The goal of this initiative is to increase efficiency and achieve a significant transformation of our data center environment over eight years. We expect our strategy to achieve a nominal cost savings of about USD 1 billion (net present cost of USD 550-650 million) through virtualization, standardization, and process improvements.

We are employing three broad strategies:

Standardize. We are implementing standard processes to more effectively use our computing resources, improving throughput of design tasks and other business processes. By integrating our utilization, inventory, and forecast information, we will better enable just-in-time procurement and reduce the amount of time between installation and full utilization.

Increase Compute Utilization. We are using grid computing techniques to enable our global design teams to apply a common pool of computing resources to individual projects. We are also using virtualization software to consolidate enterprise and office servers. These techniques increase server utilization, reduce cost, and potentially accelerate business processes. To facilitate this change and further reduce cost, we are accelerating the rate at which we refresh servers.

Reduce Data Centers. We plan to drive down cost and increase flexibility by concentrating computing resources into a smaller number of large, strategically located hub data centers. We will focus investments on these high-density, highly efficient data centers, rather than undertaking more frequent upgrades across our current larger set of data centers, and we will eliminate local data centers where possible.

We are using these strategies to increase efficiency, while reducing the cost of key physical assets such as data centers, servers, and storage, which account for a significant proportion of our operating costs, as shown in Figure 1.

A separate Intel team is focusing on achieving efficiencies in our design applications. We are also working to improve processes and standards across our environment, including governance, data quality, and support services.

We are initially focusing on several areas that can deliver significant savings to Intel, including:

- accelerating server refresh
- aligning purchases with hub sites
- consolidating and virtualization of servers
- optimizing storage

We will also set a cadence for data center closures. Subsequently, we will expand our focus to include changes to our interactive workflow model, tiered services, cost transparency, business continuity (BC) and disaster recovery, and virtualization usage models.

Data Center Facilities

Traditionally, we have located one or more data centers at every Intel site. Most of these data centers are small: of those that support non-manufacturing applications, nearly 30 percent occupy less than 1000 square feet, as shown in Figure 2.

However, we have built several medium to large high-density, air-cooled data centers that are much more efficient and cost-effective than these

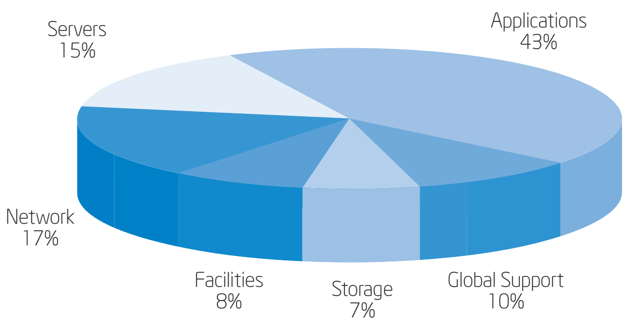

Figure 1. Cost breakdown of our current environment.

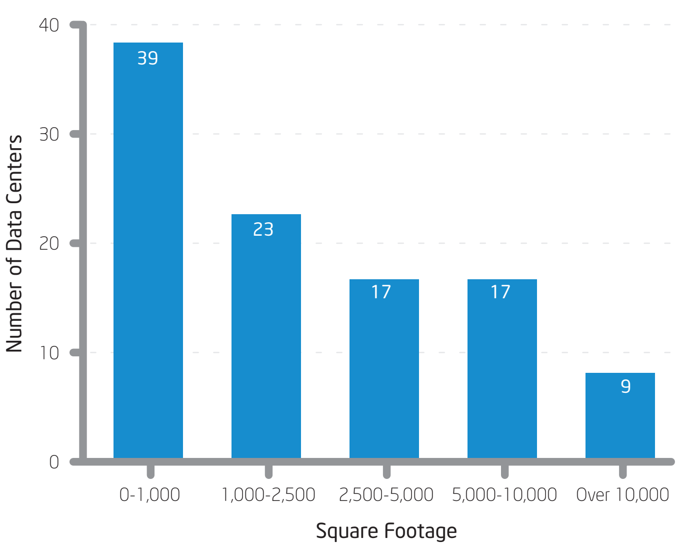

Figure 2. Current data centers by size. Excludes manufacturing data centers.

smaller facilities. By combining innovations and best known methods for air-cooled data center design, we have achieved breakthrough power and heat densities of 15 kilowatts per cabinet and more than 500 watts per square foot of server room area, with lower capital and operating costs than other data center designs. Our models indicate that it is possible to double that heat density in future.

Because of these potential efficiencies, we plan to concentrate computing resources into large, strategically located hub data centers. Our goal is to maximize the efficiencies by consolidating resources into the minimum number of data centers. As a result, we will reduce the total number of data centers by at least 50 percent, and we will no longer have a data center at every company site.

Closing local data centers will let Intel repurpose these rooms for labs, offices, or other environments. Challenging the need for local data centers will also provide us with more flexibility in site choice when leasing or constructing new facilities, potentially generating additional cost savings.

Some local data centers will remain for the foreseeable future. For example, manufacturing sites will need local data centers to run critical automation applications due to latency requirements. Some non-manufacturing sites may also need local data centers for similar reasons.

However, we believe that eight hub data centers can meet the needs of all other sites. To provide optimum response time, three will be located in each of the major geographical regions: the Americas; Europe and the Middle East; and Asia-Pacific, as shown in Figure 3. Each will be able to support all DOME computing areas, although we will tailor the mix of workloads at each region's data centers according to local demand.

By closing other data centers and designing these hubs for high density, we expect to reduce our total data center footprint from 458,000 square feet to 330,000 square feet by 2010, with additional reduction in footprint within the next three to four years.

We also expect to achieve further financial benefits as we concentrate IT equipment investment into these hubs. Hubs are often strategically located in areas with low utility costs or sales tax, which can both substantially reduce the total cost of ownership (TCO) for servers and other equipment.

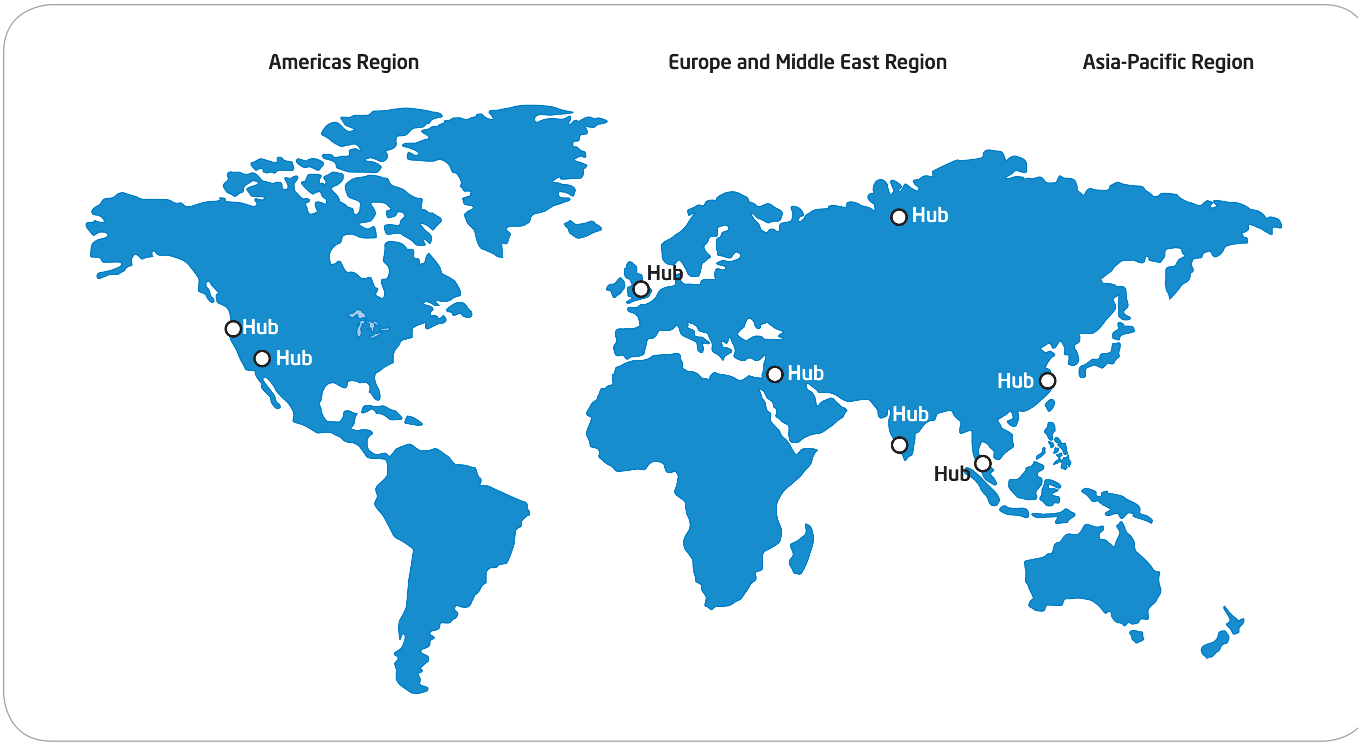

Figure 3. Strategic hub data centers.

Our strategy allows us to take advantage of this lower TCO. When we add or refresh servers, we will install the new systems at these high-density hubs rather than upgrading other sites.

As we consolidate resources into our strategic hubs, we will drive standardization in our other key focus areas, such as servers and storage.

Servers

We plan major changes in our server strategy to increase efficiency, while reducing cost and energy consumption. These changes include accelerating refresh rates, implementing consolidation and virtualization, and standardizing reference designs.

Accelerated Server Refresh

In the past, we focused on maximizing the useful life of servers. Like many organizations, we depreciate the value of a server over four years. However, we typically continued to use servers even after they were fully depreciated, rather than replacing them. While well-intended, this practice actually has increased TCO, because older machines run less efficiently and require more maintenance.

We are changing this policy and plan to refresh 100 percent of servers once they are fully depreciated. Server refresh will enable us to accelerate the shift to more power-efficient and cost-effective servers based on multi-core Intel® Xeon® processors.

These servers will help us realize other efficiencies through consolidation and virtualization.

In some cases, we may refresh servers within less than four years because we can achieve other financial benefits by doing so. For example, at a major data center facing power and space constraints, we replaced existing servers with new systems based on multi-core Intel Xeon processors. Because these new servers were more powerful than the ones they replaced, yet consumed less energy, we were able to add compute capacity without investing in a new facility, which would have been much more expensive.

Reference Designs

Different Intel groups have designed reference server configurations for specific requirements and applications. Over time, this proliferation resulted in nearly 300 reference platforms. We plan to standardize these platforms, sharply reducing their number to approximately 25. This reduction will enable us to implement other efficiencies, such as hosting a wider variety of applications on each platform. Through consolidation and virtualization, we will also be able to stack more applications onto each server. We expect this will improve utilization, reduce cost, and shorten the time to provision a new instance of an application.

Consolidation and Virtualization

We are using virtualization software to consolidate enterprise and office computing servers. We plan to consolidate the workloads of as many as 15 to 20 older servers into virtual machines onto each new virtualization host server based on multi-core Intel Xeon processors. This consolidation will improve utilization, further reducing cost and energy consumption. We expect that we will also be able to provision new instances of an application more quickly so that we can respond to business needs faster.

Grid Computing for Design

Traditionally, Intel semiconductor design groups have relied largely on servers in local design computing data centers. To obtain the growing computing power needed for their projects, design groups typically added enough local capacity to meet peak demand. The ebb and flow of use meant that in between the demand peaks, many of these local servers were underutilized. Worldwide, this added up to significant underused capacity.

We are addressing this issue with a standards-based grid computing approach that lets design teams use idle servers at remote data centers. We can then apply Intel's global computing resources to individual projects. Using these resources will increase overall server utilization and smooth fluctuations in demand, reducing the need to add capacity and cutting costs. Potentially, design task times will shorten as well.

Our goal is to make using remote resources highly automated and transparent, using an approach based on open service-oriented architecture (SOA) standards. Design teams should be able to use computing resources anywhere, without being concerned about their physical location.

While not all design activities can benefit from this approach, those that can account for about 50 percent of all compute cycles and 70 percent of the batch demand.

This grid computing strategy is an essential element of our broader data center efficiency program because eliminating site dependencies enables us to consolidate data centers. Our goal is to increase utilization from 55 percent to 80 percent or better.

Storage

To date, we have, to a large extent, managed storage independently for specific groups or applications.

However, this approach has resulted in a relatively low average utilization of around 50 percent.

Improving storage utilization has become a priority. Storage requirements are increasing rapidly at 40 percent a year or more in some areas, and storage already accounts for a significant part of the overall cost of running our environment, as shown in Figure 1.

Our goals are to increase utilization from about 50 percent currently to as much as 90 percent for storage area networks, and 70 percent for network-attached storage. We are investigating several approaches to achieve these goals.

Tiered Storage and Information Lifecycle Management (ILM)

Tiered storage and information lifecycle management (ILM) will help us drive down the average unit cost of storage by transferring less-critical data to lower-cost media. We also plan to use ILM and related approaches to archive or purge stale data more quickly, further reducing the amount of storage that is kept online. These strategies will free up disk space more quickly for reuse, reducing the need to purchase more storage.

Self-Service Provisioning

We have reduced the overhead and cycle time required to assign disk partitions to users with an automated storage provisioning tool. This tool allows a user to specify a desired quantity of storage and duration of use. The user is then assigned the requested disk space if the request meets the criteria specified by an IT policy. At the end of the requested duration, the user receives notice that his or her storage is about to be purged. A short time later, the tool reclaims the disk space and makes it available to other users. We expect that this capability, like ILM, will reduce our overall storage demand by winnowing out data that is not referenced or needed.

Thin Provisioning

Thin provisioning allocates physical disk space only when data is written, rather than when logical volumes are created. The amount of unused disk space is reduced, letting us use existing storage more efficiently and potentially reducing the need to purchase more storage. This approach also allows us to forecast and provision storage based on usage data and trends rather than trusting application owners' estimates. Ultimately, we expect to manage to use a safety buffer based on the time needed to purchase and configure storage into our environment.

Scale-out Storage

In the past, the higher performance we require for some applications has only been available with large, relatively expensive storage solutions. However, solutions are now emerging that offer high performance without the need to purchase a large system. These solutions may enable us to reduce costs of storage for design computing.

Network

As we consolidate our data centers, we will drive more standardization in network configurations. Standardization includes physical interconnects, cabling standards, and physical and logical network topologies.

We expect significant savings in LAN costs resulting from improved computing efficiency. For example, we expect that server consolidation will reduce the number of network ports, reducing cost because we require fewer network components.

As we shift computing resources from local data centers to our strategic hubs, we expect we will need to increase WAN investment in order to provide enough bandwidth to maintain service level agreements (SLAs) for many applications.

However, we plan to use WAN optimization technologies to minimize the investment required. In lab and production tests, we found that these tools can be extremely effective in reducing the load on our WAN, substantially accelerating data transfers and reducing transmitted file sizes by as much as 99 percent.[1]

Automation and Manageability

We see great opportunities to reduce complexity and cost by developing and implementing consistent manageability and automated services across our environment.

We are working to integrate a robust monitoring and management framework across all aspects of our computing environment, from data center facilities all the way up through the application layer. We will require similar data feeds from each area for more effective decision making. Similarly, we expect to gain efficiency through improved event correlation and mitigation capabilities that span servers, storage, networks, and even facilities.

We expect to realize increasing benefits as our various business groups and application owners adopt more consistent manageability capabilities. We are also working to reduce the number of management tools across the enterprise, from the several dozen in use today, to a smaller, more manageable set.

1 Verall, D., D. Wescott, and Lilin Xie. "Optimizing WAN Performance for the Global Enterprise." http://www.intel.com/it/pdf/optimizing-wan-performance.pdf (PDF).

Energy Efficiency

Energy efficiency is a key focus area for Intel IT. We are using several approaches to reduce our data center energy consumption and carbon footprint.

Power-Efficient Servers

Increased server energy efficiency is a key benefit of our accelerated server refresh approach and an important consideration in our overall server strategy. New servers based on multi-core Intel Xeon processors are much more power efficient than the systems they replace and play a major role in minimizing energy consumption and cost. These new servers also deliver substantial benefits at power-constrained data centers by allowing us to add compute capacity without increasing overall power consumption.

Data Center Cooling Techniques

We have used a combination of innovations and best known methods to develop highly power-efficient cooling for our high-density data centers.

For example, we designed a data center heat recovery system that will play an integral role in Intel's first planned "green" building. The system employs heat recovery chillers that capture heat produced by servers in the building's data center and reuse the heat to warm offices in winter and provide year-round hot water. This recovery eliminates the need to add boilers for heating the building. We expect that this approach will save energy, accumulate points towards green certification, and deliver annual cost savings estimated at hundreds of thousands of USD.

We have also used many other energy-saving techniques, including wet-side economizers that provide "free cooling" at several data centers.

Industry Initiatives

Through our involvement in industry research and technology development groups, we are evaluating new technologies with the potential to reduce our energy consumption even further.

We are a member of the Green Grid Consortium, with the goal of developing guidelines, standards, and best practices for data center efficiency.

Our commitment to the Climate Savers Computing Initiative, which aims to reduce computer-related greenhouse gas emissions by 50 percent, includes adopting energy-efficient practices and technologies, including servers with 80 percent efficient power supplies.

Processes and Governance

Our strategy requires clearly defined governance and BC plans. We are concentrating our computing resources into a smaller number of strategic hubs and reducing unused server and storage capacity. This consolidation and reduction makes it even more important to plan for events that could further reduce available capacity, such as an outage at a hub.

Fortunately, our strategy will also make effective BC planning easier because we will have a smaller number of sites, well-defined mapping of services, and fast and efficient provisioning.

We have already seen BC benefits in our adoption of grid computing for design workloads: we can quickly adapt in the event of a site outage by redirecting demand to another site. In the past year, we experienced unplanned power outages in both small and large data centers, but we minimized the impact by rerouting and reprioritizing jobs in our hub locations.

We intend to run our data center environment in much the same way that we run our factories. Our goal is to manage supply with clear forecasts, allowing just-in-time technology purchases triggered by demand signals.

These demand signals will come from a mix of sources, including tracking data and corporate long-range plan information. We will depend largely on utilization data to trigger purchases. To budget appropriately, we will analyze project requirements and correlate them with project type and complexity.

We will automate as many of these processes as possible, basing them upon a foundation of management tools to perform discovery, provisioning, monitoring, and remediation in a consistent, policy-driven way.

Tiered Service Levels

Historically we have aimed to provide high availability for enterprise applications. Over time, this has meant that the vast majority of our enterprise applications are supported with the same high level of service. This system has yielded a higher cost of ownership, due to hardware expense and support required to deliver this service level.

We are working with application owners and business groups to develop formal descriptions of a range of differing service levels. We will then translate these descriptions into specific requirements and capabilities that we need to deliver at each service level.

Providing differentiated service levels for applications will increase efficiency, including reduced hardware and support costs. We also plan to be able to easily migrate applications to different service levels as business needs change.

Results

Though our strategy is still evolving, we have already made substantial progress in several areas. These include grid computing, along with enterprise server consolidation and virtualization.

Grid Computing for Design

Our global grid is delivering results. Since 2006, we have seen an 11 percent increase in server utilization, allowing Intel to avoid more than USD 30 million in 2007 capital purchases and approximately USD 77 million in 2008 capital purchases.

Approximately 40 percent of all batch jobs now run remotely. We have developed tools to make this possible, including a scheduler that assigns jobs to server pools spread across multiple data centers, and software that automatically provisions operating systems to as many as 100 servers in 30 minutes.

While increasing utilization, we also improved throughput time for many projects. This improved time enabled design groups to meet or improve on key milestone dates for their projects.

Enterprise Server Consolidation and Virtualization

We have begun consolidating enterprise computing workloads onto servers based on multi-core Intel Xeon processors. This consolidation has reduced the number of servers by about 1800, while increasing the overall number of cores.

We have also validated key operational goals from virtualization, hosting as many as 11 virtual machines on one physical machine and migrating workloads from one physical machine to another to perform preventative maintenance.

Next Steps

Over the next 12 months, we plan to execute changes in several key areas.

Data Center Consolidation

As we finalize plans for consolidating our data centers, we will build an asset and application inventory to help us optimize and manage our environment overall. Because consolidation means some users are farther away from computing capacity, we will also characterize our applications' WAN latency sensitivities and mitigate where necessary.

We will prioritize our data center closures based on a number of factors, including business value and the simplicity of accomplishing the closure. We are also looking for cases in which we can repurpose data centers as lab space to avoid construction or leasing costs.

Grid Computing

We are aiming to drive utilization still higher, from the current 66 percent to 80 percent, generating potential additional savings of tens of millions USD. We also plan to consolidate design batch computing into three data centers.

As we do this, we are streamlining our ability to reprovision servers based on demand by integrating some reprovisioning capability into our scheduler. We expect that this will enable us to respond more quickly to changes in demand by tailoring the configurations of hub server pools to the needs of incoming workloads.

We have also begun adjusting our SLAs for batch capacity to define an agreed range between minimum and expected service levels, rather than a single number representing a specific level of service. This adjustment will further smooth out the peaks and valleys of our utilization trends, helping us increase average utilization.

Servers

We will continue to consolidate workloads onto new multi-core Intel Xeon processors, while characterizing applications and server performance to lay the ground for even more extensive consolidation. We will validate new workloads on our standardized reference platforms. We will improve our support capabilities for virtualized servers by incorporating tools and processes for tracking the deployment and migration of virtual machines on physical servers. For example, we plan to maintain a real-time mapping of virtual instance to physical location to help manage complexity and speed recovery in the event of a physical server failure.

Storage Optimization

We plan to deploy horizontal shared storage services across all compute areas, while working to establish an ILM plan. By deploying shared services, we expect to be able to improve our ratio of support staff to storage capacity, effectively increasing our support capabilities.

Cost Transparency

As we move toward shared, standard services and away from customized solutions, we will be able to provide our business partners with much more detailed information on performance and cost of their compute environments.

Transparency also ties directly back to service levels, so we can publish a catalog of services to the application owners and even move to a chargeback model if that makes the most sense.

Application and Workflow Characterization

A full, formally described understanding of our enterprise application environment is critical to our plans for data center consolidation. We have tools to analyze the behavior of applications at system and network level, and we are developing processes to fully understand and document application behavior.

Once we have profiled our individual applications, our next step will be to characterize the overall workflow. Our goal is to develop a full understanding of an entire set of applications and its dependencies so that we know exactly how migrating applications to a different data center might affect workflow.

Conclusion

Consolidation and overall streamlining of data centers will result in a significant transformation of Intel's data center environment. This improvement includes substantial projected savings through efficiencies in areas such as data center consolidation, grid computing, server virtualization, and storage optimization.

Our transformation has already enabled us to consolidate services and reduce the number of data centers by 13 percent in 2007, while reducing square footage by 4 percent. Over time, we plan to reduce our data center space requirements by about 35 percent, from 460,000 square feet to about 300,000 square feet total.

We have already begun to achieve benefits in several other areas, including grid computing and server consolidation.

Our use of virtualization and grid computing reduces energy consumption and boosts utilization, enabling us to share compute servers by removing physical, geographical, and organizational boundaries, as well as allowing Intel to avoid more than USD 30 million in 2007 capital purchases.

Since 2006, we have seen an 11 percent increase in server utilization, which will allow Intel to save approximately USD 77 million in 2008 capital purchases.

After achieving savings and creating a more efficient infrastructure during the initial part of our plan, we will expand our focus to other areas. Ultimately, our goal is to run our data centers much as we run our factories, with similar efficiencies, economies of scale, and responsiveness to business needs.

Authors

Data Center Core Team: Alan Ross, Brently Davis, Christian Petersen, Craig Bolton, David Anzelc, David Schlick, Diane Boyington, Erika Edgerly, Gerald Seamans, Ila Patel, Jodie Vanderburg, John Vicente, Ken Dehoff, Kenny Sng, Naomi Masterson, Mike Budzowski, Paul Vaccaro, Rick Kraemer, Sally Wellsandt, Sanjay Rungta, Travis Broughton, Uttam Shetty

Acronyms

BC	business continuity
DOME	design, office, manufacturing, and enterprise
ILM	information lifecycle management
SAN	storage area network
SLA	service level agreement
SOA	service-oriented architecture
TCO	total cost of ownership
UPS	uninterruptible power supply

www.intel.com/IT

Printed in USA
0108/IPKA/RDA/PDF

♻ Please Recycle
ITAI Number: 07-2203w