WHITE PAPER
NOVEMBER 2017

IT@INTEL

Data Center Strategy Leading Intel's Business Transformation

Our refined data center strategy has generated savings exceeding USD 1.5 billion from 2010 to 2016.

Shesha Krishnapura
Intel Fellow and Intel IT CTO

Shaji Achuthan
Senior Staff Engineer, Intel IT

Pushpa Jahagirdar
Technical Program Manager, Intel IT

Vipul Lal
Senior Principal Engineer, Intel IT

Raju Nallapa
Principal Engineer, Intel IT

Sanjay Rungta
Senior Principal Engineer, Intel IT

Ty Tang
Senior Principal Engineer, Intel IT

Executive Overview

To better meet Intel's business requirements while providing our internal customers with optimal data center infrastructure capabilities and innovative business services, Intel IT has overhauled our data center strategy. Our data center transformation strategy is to run Intel data center services like a factory, affecting change in a disciplined manner and applying breakthrough technologies, solutions, and processes.

We use three key metrics to measure data center transformation success: meet growing customer demand (service-level agreements and quality of service) within constrained spending targets (cost-competitiveness) while optimally increasing the utilization of infrastructure assets (operational efficiency).

Building on previous investments and techniques, our refined data center strategy has generated savings exceeding USD 1.5 billion from 2010 to 2016. Our key achievements include the following:

- Our breakthrough disaggregated server design allows independent refresh of CPU and memory without replacing other server components, which results in faster data center innovation and a 44 percent cost savings compared to a full-acquisition refresh.
- We developed a system software capability called NUMA-Booster, which has delivered additional server capacity.
- We deployed more than 13,000 Intel® Solid State Drives as fast swap drives, which increased server capacity by 27 percent.
- Five generations of high-performance computing in our Design computing environment created a 121x capacity increase and a 161x quality improvement.
- We adopted new storage capabilities, accelerated storage refresh, and focused on increasing utilization, resulting in cost avoidance.
- We deployed more than 87,000 10 gigabit Ethernet (GbE) network ports and 3,100 40 GbE network ports.

Over the next three years, we plan to extend the data center strategy to continue to transform our data center infrastructure. We will do so by using disruptive server, storage, network, and data center facility technologies that can lead to unprecedented quality-of-service levels and total cost of ownership reduction for business applications—all while continuing to improve IT operational efficiency.

Our data center transformation strategy is key for intel IT to continuously deliver innovative, efficient and cost competitive data center capabilities. Implementing breakthrough solutions and pursuing aggressive goals are critical factors to success in this transformation.

Contents

1 Executive Overview

2 Background
– Meeting Compute Environment Challenges
– Aligning Data Center Investments with Business Needs

4 Unique Elements of Our Data Center Strategy
– Defining Key Performance Indicators and Goals
– Stimulating Bold Innovation through a New Investment Model
– Implementing a New Unit-Cost Financial Model

8 Results: Building on the Past, Building for the Future
– Cumulative Results from 2003 to 2016
– Results from 2010 to 2016
– Reduced TCO with Blade Server Technology

15 Summary of Data Center Best Practices

17 Plans for 2017 and Beyond

17 Conclusion

Acronyms

DOMES	Design, Office, Manufacturing, Enterprise, and Services
EDA-MIPS	electronic design automation-meaningful indicator of performance per system
GbE	gigabit Ethernet
HPC	high-performance computing
IaaS	Infrastructure as a Service
KPI	key performance indicator
MIPS	meaningful indicator of performance per system
MOR	model of record
MTTR	mean time to repair
NAS	network-attached storage
NIC	network interface card
NUMA	non-uniform memory access
PUE	power usage effectiveness
QoS	quality of service
RISC	reduced instruction set computing
ROI	return on investment
SAN	storage area network
SLA	service-level agreement
SSD	solid state drive
TCO	total cost of ownership
USD	U.S. dollar
VM	virtual machine

Background

Intel IT operates 56 data centers housing approximately 217,000 servers that underpin the computing needs of more than 106,000 employees.[1] To support the business needs of Intel's critical business functions—Design, Office, Manufacturing, Enterprise, and Services (DOMES)—while operating our data centers as efficiently as possible, Intel IT has engaged in a multi-year evolution of our data center strategy, as outlined in Table 1.

[1] Number of data centers and servers as of September, 2017. To define "data center," Intel uses IDC's data center size classification: "any room greater than 100 square feet that houses servers and other infrastructure components."

Table 1. Intel's Data Center Strategy Is a Continuous Improvement Process

Pre-2000	• No centralized strategy or ownership • Began RISC migration • Built data centers to support acquisitions • Decentralized procurement and management
2000-2006	**Standardization and Cost Control** • Formed data center team • Began data center consolidation efforts • Completed RISC to Intel® architecture migration in Design • Standardized data center designs
2006-2010	**Foundation for Efficient Growth** • Business-focused investments for DOMES • Proactive server and infrastructure refresh • Virtualization and enterprise private cloud • Storage optimization and IT sustainability
2010-2013	**Transform Business Capabilities** • Holistic TCO assessment of IaaS, including facilities, server, storage, network, OS, and middleware software capabilities • Introduction of data center MOR to drive practically achievable performance at peak efficiency and at the lowest cost • Unit costing model to plan improvement targets and benchmark among industry peers and external cloud providers • Data center pulse dashboard for comprehensive state of IaaS capacity and utilization to enable future planning and current improvements
2013+	**Focus on Resource and Energy Efficiency** • Centralize batch computing capacity in two mega-hubs • Combine high-frequency servers and optimal workloads for platform pairings • Centralized management of servers and resources • Convert older wafer fabrication facilities into data centers • Custom rack design to optimize space, compute and power density • Environmental sustainability—either free-air cooling or evaporative cooling-tower water to condition the data centers • State-of-the-art electrical density and distribution system • Breakthrough disaggregated server architecture

Meeting Compute Environment Challenges

In the past, we focused our data center investments on improving IT infrastructure to deliver a foundation for the efficient growth of Intel's business. Our primary goal was cost reduction through data center efficiency and infrastructure simplification while reducing energy consumption and our carbon dioxide footprint to improve IT sustainability.

Over the last several years, we have reduced data center energy consumption and greenhouse gas emissions, while at the same time meeting the constantly increasing demand for data center resources. We anticipate these growth rates to continue or even increase further:

- 30 to 40 percent annual growth in compute capacity requirements
- 35 to 40 percent annual growth in storage needs
- 30 to 40 percent annual growth in demand for network capacity

To address these challenges without negatively impacting service delivery, we developed and continue to rely on many established industry best practices in all areas of our data center investment portfolio—servers, storage, networking, and facility innovation. Since 2010, these techniques, which are described in detail later in this paper, have enabled us to realize 1.5 billion U.S. dollars in cost savings while supporting dramatic growth.

Aligning Data Center Investments with Business Needs

We have learned that a one-size-fits-all architecture is not the best approach for our unique business functions. After working closely with business leaders to understand their requirements, we chose to invest in vertically integrated architecture solutions that meet the specific needs of individual business functions.

Design

Design engineers run more than 140 million compute-intensive batch jobs every week. Each job can take a few seconds to several days to complete. In addition, interactive Design applications are sensitive to high latencies caused by hosting these applications on remote servers. We have used several approaches in our Design computing data centers to provide enough compute capacity and performance to support requirements, including high-performance computing (HPC), grid computing, clustered local workstation computing, using solid state drives (SSDs) as fast-swap drives, single-socket servers, and a specialized algorithm that increases the performance of the heaviest Design workloads.[2] Together, these investments enable Design engineers to run up to 49 percent more jobs on the same compute capacity—which equates to faster design and time to market.

Our breakthrough disaggregated server architecture allows us to independently refresh servers' CPU and memory without replacing other server components—resulting in faster technology adoption, which in turn puts new technology at our Design engineers' fingertips. The disaggregated servers also save 44 percent over a full-acquisition refresh.

Because Design engineers need to access Design data frequently and quickly, we did not simply choose the least expensive storage method for this environment. Instead, we have invested in clustered and higher performance scale-out, network-attached storage (NAS) in combination with highly scalable parallel storage for our HPC needs. We use storage area networks (SANs) for specific storage needs such as databases.

[2] Intel uses grid computing for silicon design and tapeout functions. Intel's compute grid represents thousands of interconnected compute servers, accessed through clustering and job scheduling software. Additionally, Intel's tapeout environment uses an HPC approach, which optimizes all key components such as servers, storage, network, OS, applications, and monitoring capabilities cohesively for overall performance, reliability, and throughput benefits. For more information on HPC at Intel, refer to "High-Performance Computing for Silicon Design," Intel Corp., December 2015.

Manufacturing

IT systems must be available 24/7 in Intel's Manufacturing environment, so we use dedicated data centers co-located with the factories for Manufacturing. We have invested heavily over the last few years to develop a robust business continuity plan that keeps factories running even in the case of a catastrophic data center failure. These efforts have paid off, and we have not experienced factory downtime related to data center facilities since 2009.

In our Manufacturing environment, we pursue a methodical, proven infrastructure deployment approach to support high reliability and rapid implementation. This "copy-exact" approach deploys new solutions in a single factory first and, once successfully deployed, we copy that implementation across other factory environments. This approach reduces the time needed to upgrade the infrastructure that supports new process technologies—thereby accelerating time to market for Intel® products. The copy-exact methodology allows for rapid deployment of new platforms and applications throughout the Manufacturing environment, enabling us to meet a 13-week infrastructure deployment goal 95 percent of the time—compared to less than 50 percent without using copy-exact methodology.

Office, Enterprise, and Services

To improve IT agility and the business velocity of our private enterprise cloud, we have implemented an on-demand self-service model, which has reduced the time to provision servers from three months to on-demand provisioning. We have achieved virtualization of 90 percent of OS instances in our Office and Enterprise environment, up from 12 percent in 2010.

In contrast to the Design environment, in the Office, Enterprise, and Services environments we rely primarily on SAN storage, with limited NAS storage for file-based data sharing.

Unique Elements of Our Data Center Strategy

Our transformational data center strategy is to run Intel data centers and all underlying infrastructure as if they were factories, with a disciplined approach to change management. By applying breakthrough technologies, solutions, and processes, we can lead the industry and keep up with the accelerating pace of Intel's business.

We have realized substantial cost savings since 2006 by proactively refreshing our infrastructure, adopting cloud computing, updating our network, pursuing IT sustainability, and consolidating data centers. In addition, we have supported business growth and capability improvements by deploying unique solutions that benefit Intel's critical business functions—DOMES. We have enhanced our strategy to include several new elements (as detailed in subsequent sections):

- **Key performance indicators (KPIs).** We have implemented three KPIs and have established goals for each of them:
  - Quality of service (QoS), using a tiered approach to service-level agreements (SLAs)
  - Cost efficiency
  - Effective utilization of assets and capacity

  Based on improvements each year in technologies, solutions, and processes, we identify the best achievable SLA, the lowest achievable cost, and the highest achievable resource utilization. We call this combination the model of record (MOR) for that year. We set investment priorities based on these KPIs to move toward the MOR goal; each year we are getting closer to the MOR while at the same time balancing the three vectors.

- **Investment decision model.** Focusing on the MOR and comparing current data center capabilities to the best achievable KPIs enables us to prioritize our investment decisions. This approach seeks to remove the conventional improvement mindset, which focuses only on incremental improvements. Instead, we are transforming our capabilities by identifying further groundbreaking innovations—like those already used to implement our private cloud and our highly efficient silicon design computing grid.
- **Unit-costing financial model.** By identifying metrics for improvements in each DOMES area, we can benchmark ourselves and further prioritize our investments.

We believe our new approach to data center costing and investment evaluation, along with a continued focus on meeting business needs, has stimulated a bolder approach to continuous innovation. Our efforts have improved the quality, velocity, and efficiency of Intel IT's business services, creating a sustained competitive advantage for Intel's business. For details, see "Results: Building on the Past, Building for the Future."

Defining Key Performance Indicators and Goals

The KPIs provide a means to measure the effectiveness of data center investments. Because the service output for each business function is different, we evaluate each business function separately. In our data center investment decisions, we seek to balance and meet all business requirements while optimizing the KPIs.

Quality of Service

We use a tiered approach to SLAs, tailored to each business function's sensitivity to performance, uptime, mean time to repair (MTTR), and cost. Our goal for this KPI is to meet specific performance-to-SLA requirements for defined tiering levels. For example, for our most mission-critical applications, we aim for a higher performance to SLA than for second-tier applications, which are less critical. The end goal and true measure of IT QoS is zero business impact from IT issues.

Effective Resource Utilization

Our refined data center strategy represents a dramatic shift in how we view resource utilization. Historically, we measured utilization of IT assets—compute, storage, network, and facilities—by simply determining how busy or loaded an asset was. For example, if a server was working at peak capacity 90 percent of the time, we considered it 90 percent utilized. If 80 percent of available storage was allocated, we considered that 80 percent utilization.

In contrast, we now focus on the actual output of an asset—that is, *effective* utilization. For example, if Intel's Design engineers start one million design jobs—thereby keeping the servers very busy—but a third of those jobs terminate before completion because there was not enough storage available,

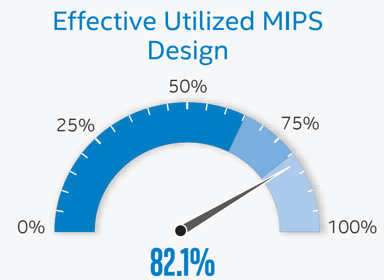
Intel IT Data Center Dashboard

To better monitor and manage our worldwide network of data centers, we developed and deployed an integrated business intelligence dashboard. This BI tool is modeled on a dashboard used in Intel's Manufacturing environment.

This dashboard will help us monitor our key performance indicators (KPIs) by highlighting the current state and opportunities for optimization, thereby enabling overall improvements that align with our data center strategy goals.

For example, the dashboard can report on effective utilization of several data center resources, including electronic design automation-meaningful indicator of performance per system (EDA-MIPS); raw and utilized storage capacity; and facilities space, power, and cooling.

This data can report statistics by business function or by data center, and can be used to compare KPIs and metrics across several data centers. The figure below shows a sample of the dashboard.



The Intel IT Data Center Dashboard provides a holistic view of data center resources to help us track our KPIs and identify opportunities for optimization and improvement.

that is low effective utilization of compute capacity—only 66 percent. Or, if a customer consumes only 4 GB of a 10-GB storage allocation, the remaining 6 GB is essentially wasted storage—even though it is allocated—and does not represent effective utilization of this asset. Our goal for the effective utilization KPI is to achieve 80 percent effective utilization of all IT assets.

Cost per Service Unit

As shown in Table 2, different business functions have a different service unit that we can measure. This unit represents the capacity we enable for our business users.

Our goal for this KPI is to achieve a 10 percent improvement in data center cost efficiency every year. This goal does not necessarily mean we will spend less each year, but rather that we will get more for each dollar we spend. For example, we may spend less for the same number of service units, or we may spend the same amount but get more service output.

Table 2. Service Unit for Each Business Function

Function	Service Unit
Design	Cost per EDA-MIPS
Office, Enterprise, and Services	Cost per OS instance
Manufacturing	Cost per integrated factory compute environment

Stimulating Bold Innovation through a New Investment Model

Building on a time-tested methodology that has proven successful in Intel's Manufacturing environment over multiple process technology generations, we adopted a new data center investment decision model that compares current data center capabilities to a "best achievable model" that guides us to make investments with the highest impact.

Previously, Intel data center planning teams looked at existing capabilities and funding to establish a plan of record. This plan drove incremental improvements in our existing capabilities; our goal was to minimize total cost of ownership (TCO) and deliver positive return on investment (ROI).

In contrast, the MOR ignores the constraints imposed by what we have today. Instead, it identifies the minimum amount of resources we should *ideally* have to support business objectives—thereby establishing an optimal state with available technology.

By setting a standard of maximum achievable performance, the new model enables us to:

- Determine which investments will have the highest ROI.
- Identify the benefits of using disruptive infrastructure technologies and breakthrough approaches that deliver more optimal data center solutions across all aspects of our infrastructure.
- Make data center location decisions, including identifying potential data centers to consolidate, upgrade, or close.

The new model focuses limited available resources in specific areas for maximum holistic gain.

As shown in Figure 1, because technology is always changing, peak performance also changes—the maximum achievable performance

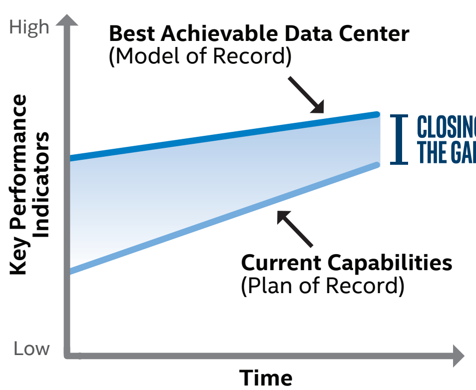


Figure 1. Our new data center investment model encourages innovation and provides significant business results.

keeps improving through innovation. We know that resource constraints make it impossible to ever actually achieve the standard set by the new investment model—although our HPC environment comes very close to that goal. However, the model enables us to identify gaps between where we are and where we would like to be. We can then identify the biggest gaps in capability to prioritize our budget allocation toward the highest value investments first.

Implementing a New Unit-Cost Financial Model

We evolved our financial model from project- and component-based accounting to a more holistic unit-costing model. For example, we previously used a "break/fix" approach to data center retrofits. We would upgrade a data center facility or a portion of the facility in isolation, looking only at the project costs and the expected ROI of that investment, with no holistic view as to the impact of service unit output. In contrast, today we focus on TCO per service unit—using the entire data center cost stack per unit of service delivered. This cost stack includes all cost elements associated with delivering business services and now considers the worldwide view of all data centers in the assessment of our investments.

As shown in Figure 2, there are six major categories of cost to consider: network, headcount, servers, facilities, OS and management, and storage and backup and recovery. By adding these costs and then dividing by the total number of appropriate service units for the environment, we arrive at a cost per service unit.

Service-based unit costing enables us to benchmark ourselves and prioritize data center investments. Determining service-based unit costs also allows us to measure and compare the performance of individual data centers to each other, identifying which are underperforming and giving us the tools to decide whether to upgrade or consolidate underperforming data centers.

To show how the new unit-based costing model works, Figure 3 compares Design cost data and Office, Enterprise, and Services cost data. The headcount category accounts for a greater percentage of total cost in Office, Enterprise, and Services than it does in Design; in contrast, servers are more of a cost factor in Design than they are in Office, Enterprise, and Services. Knowing our exact unit cost in each environment, as well as the breakdown of that cost, enables us to develop optimized solutions for each environment that will have the greatest effect on cost efficiency and ROI.

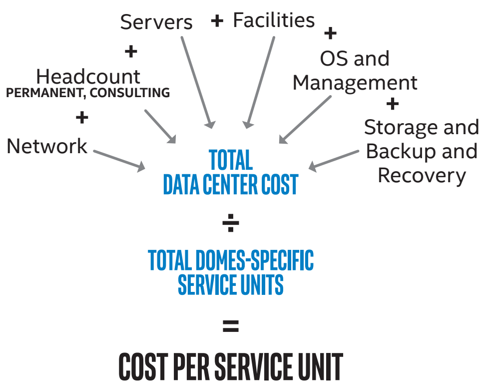


Figure 2. We arrive at a data center unit cost by considering all categories of cost and dividing by the number of units for that environment, such as EDA-MIPS of performance per system in Design and OS instances in Office, Enterprise, and Services.

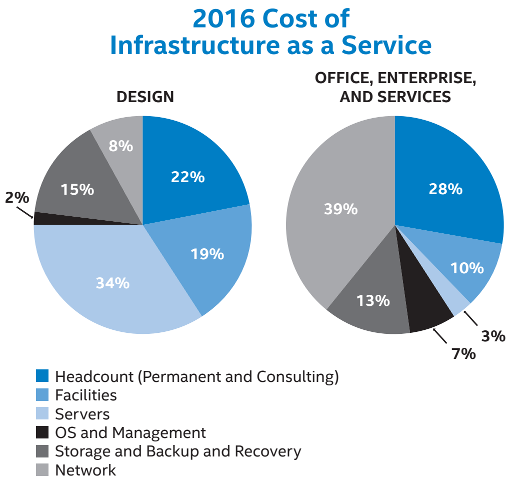


Figure 3. Knowing the total unit cost, as well as the individual cost category figures for each business environment, enables us to better choose IT investments that will lower costs the most.

Results: Building on the Past, Building for the Future

This section provides details on some of the improvements and cost savings our data center strategy has enabled over the years. We are building on the success we have already achieved through our data center initiatives over the last decade. Therefore, some of the results shown here are cumulative; others have been achieved over the last three years and are a direct result of our new MOR strategy. Our refined data center strategy will enable us to support the growth of Intel's customers, products, and acquisitions, as well as enhance the quality, velocity, and efficiency of the services we offer to Intel business groups.

Cumulative Results from 2003 to 2016

During the last decade, we have dramatically improved performance and reduced costs for our data centers (Table 3).

Table 3. Data Center Improvements from 2003-2016

Data Center-Wide	• Smaller total data center footprint • Improved overall storage and network practices
Design Environment	• More efficient Design compute and storage • 5th generation of high-performance computing • Increased Design throughput using NUMA-Booster • Faster Design throughput using Intel® SSDs
Office and Enterprise Environment	• More efficient Office and Enterprise compute and storage

Data Center-Wide Improvements

We have improved the performance and cost efficiency of our data centers overall.

Smaller Total Data Center Footprint

Figure 4 shows how we have consolidated our data center facilities during the past 15 years. We have reduced the total square footage by 27 percent and reduced the number of data center facilities from 152 to 56. See "Continued Data Center Consolidation" for a discussion of how the MOR strategy has directly enabled some of these achievements.

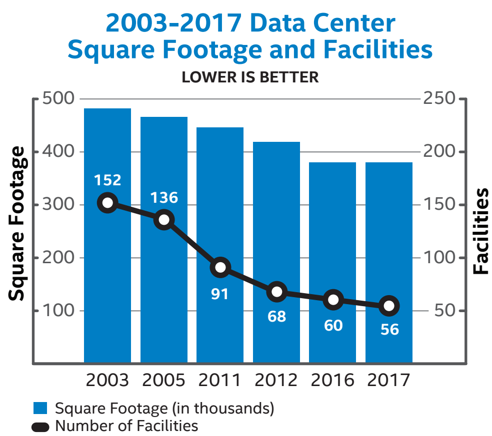


Figure 4. Over the last decade, even as we have met increasing demands for compute and storage resources, we have reduced our data center footprint by 27 percent.

Improved Overall Storage Practices

A significant focus on effective utilization in our Design environment has enabled us to improve resource utilization from below 45 percent to more than 70 percent—our goal has been updated to reach 80 percent.

We have applied several storage techniques to enhance storage efficiency and reduce costs:

- **Tiered storage.** A five-tier approach to storage has helped us increase effective utilization of storage resources, improve our performance to SLAs, and reduce the TCO for Design storage. The tiers of Design storage servers are based on performance, capacity, and cost. Tier-1 servers have the highest performance and the least storage capacity. Tier-2 servers offer medium performance but greater storage capacity. Tier-3 servers provide lower performance but emphasize capacity, while Tier-4 and Tier-5 servers have the highest capacity but are used for low-frequency access and read-only archive data. We updated our strategy to account for computational scale of the site to determine the appropriate performance level required for each tier; this enabled us to improve our ability to meet the quality, SLA, and cost targets.

- **Scale-out storage.** We are making a strategic shift from a fragmented scale-up storage model to a pooled scale-out storage model. Scale-out storage better supports on-demand requests for performance and capacity. In addition, scale-out storage enables transparent data migration capabilities and increases the effective utilization of space freed using efficiency technologies such as deduplication and compression.
- **Storage refresh cycle.** To improve performance and reduce cost, we are implementing an efficiency-based refresh cycle. This refresh cycle enables us to take advantage of storage servers with better performance and more efficient energy use, thereby reducing both capital and expense costs. For example, a more energy-efficient server can reduce data center power usage; a more powerful server that replaces several older servers can reduce our data center footprint while delivering better performance for our customers at a similar or lower cost per TB.
- **Data deduplication.** The introduction of new storage to support company growth and our commitment to timely refresh are enabling us to use the latest generation of Intel® Xeon® processors. These processors provide us with the processing power to handle data deduplication and compression on our primary storage servers—freeing greater than 15 PB of capacity, which we are making available for our users.

Improved Overall Network Practices

To accommodate the increasing demands that data center growth places on Intel's network, Intel IT converted our data center network architecture from multiple 100 megabits per second and 1 gigabit Ethernet (GbE) connections to 10 GbE connections. The older, slower connections no longer supported Intel's growing business requirements. The conversion to 10 GbE started in 2010; we currently have deployed more than 87,000 10 GbE ports. Our new 10 GbE data center fabric design accommodates our current annual 40 percent network capacity growth (see Figure 5). We have started deploying 40 GbE ports in specific use cases where they provide business value.

In addition to increasing the network capacity, we have also increased the effective utilization of network ports over the last six years, from 40 percent to 68 percent (see Figure 6). Higher utilization means we do not have to purchase additional ports to meet network capacity demand growth.

We are also focusing on improving data center stability. In the past, we used a large installation of layer 2-based technology. We are

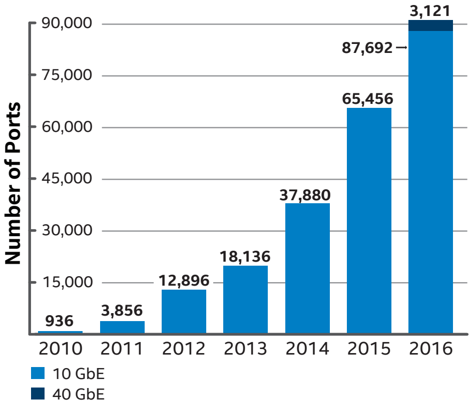


Figure 5. Implementing 10 gigabit Ethernet (GbE) data center fabric design accommodates current capacity growth as well as meets increasing network demand in the future.

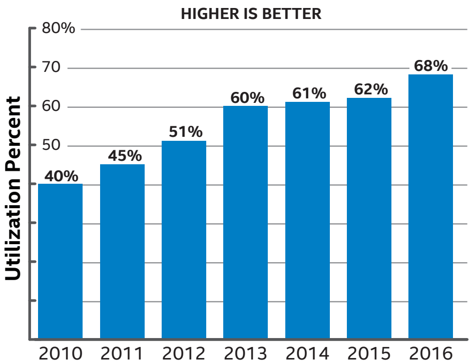


Figure 6. We have continued to increase effective utilization of our network.

now migrating to a layer 3-based network. This new architecture is enabling us to use all available bandwidth on primary and secondary paths at the same time. Therefore, we can use our network capacity more effectively. We are also able to eliminate the spanning-tree protocol within our data centers; this protocol does not scale well for large networks. Because the Internet uses layer 3-based, scalable architecture, using this concept within Intel's data center will make our data center network more scalable and resilient. Also, we are using other technologies such as overlay, multi-chassis link aggregation (MLAG), and tunneling to extend layer 2 across data centers, over the layer-3 topology.

As shown in Figure 7, we tend to adopt higher-speed network technology almost as soon as it is available in the market. We started adoption of 40 GbE in data centers in 2015 and we expect to adopt 100 GbE technology within some data center environments by the end of 2017, to keep pace with network demand.

In 2015 we also made two key architecture changes within Design data centers to reduce the oversubscription through the infrastructure and shift from chassis-based switches to fixed form factor switches for better cost and upgrade efficiency. Over the next three years we will reduce the oversubscription from 8:1 to 6:1 on the compute side and 8:1 to 3:1 on the file server side. Over the same period we plan to transition Design data centers by 70 percent using fixed form factor switches in a modular design.

Improvements in the Design Environment

Because silicon chip design represents a significant portion of Intel's business, we have applied our data center strategy to several aspects of Design computing.

More Efficient Design Compute and Storage

One of the major challenges in our Design environment is that server and storage growth is occurring at a high rate. Compute demand is growing 30 to 40 percent year over year, while storage capacity demand is increasing 35 to 40 percent annually (see Figure 8).

We expect the number of cores to continue to increase. We plan to measure data center performance based on number of cores, number of racks, power consumed, and the extent to which we meet the meaningful indicator of performance per system (MIPS) demand.

5th Generation of HPC

Designing Intel® microprocessors is compute intensive. Tapeout is a final step in silicon design, and its computation demand is growing exponentially for each generation of silicon process technology.

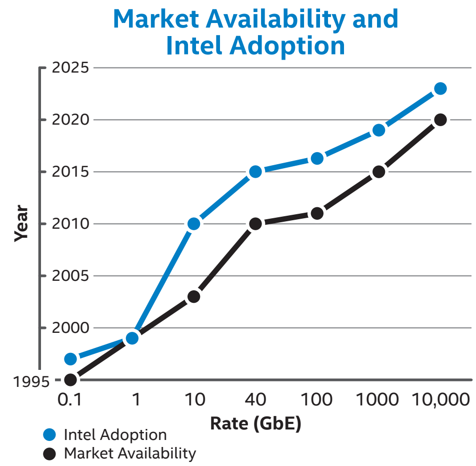


Figure 7. Intel IT adopts higher-speed network technology almost as soon as it is available.

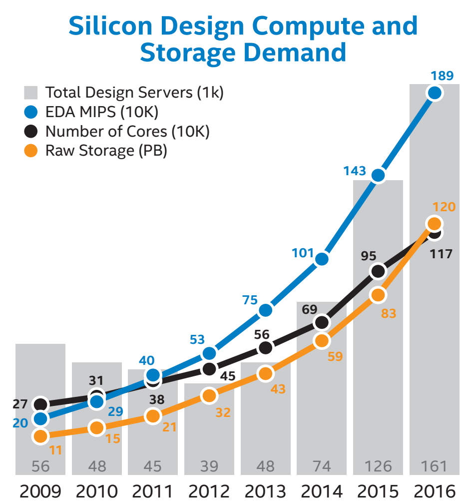


Figure 8. Despite continuing growth in compute and storage demand, our Design data centers are using powerful Intel® technology to meet demand. In 2013, we shifted to a new server form factor which shows server count increasing; however, it is highly dense, allowing us to place 140-180 servers per rack. In 2016, we again shifted to a new form factor—the disaggregated server—which allows us to place 280 servers per rack.

Intel IT adopted HPC to address this large computational scale and realized significant improvements in computing performance, reliability, and cost.

As shown in Figure 9, our HPC solution has enabled a 121x growth in tapeout compute capacity from 2005 to 2016. We are now using the 5th generation of our HPC solution and will continue to develop new HPC generations as Intel process technology advances. The figure also shows our commitment to quality. Through a disciplined approach to change management (basically running our data centers as if they are factories), we have reduced the number of compute issues that impact tapeout by 161x.

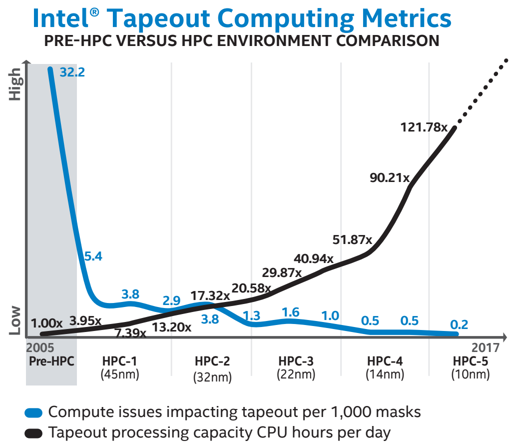


Figure 9. Our HPC solution, combined with disciplined change management, has steadily increased compute capacity and improved quality of service.

Breakthrough Disaggregated Server Innovation Reducing Refresh Costs

Just as it makes little sense to replace an entire light fixture when all that is needed is a more energy-efficient and powerful light bulb, replacing an entire server does not make sense if all that is needed is a more advanced CPU and DRAM. As shown in Figure 10, Intel IT has developed a disaggregated server architecture—the first major server innovation since the introduction of blade servers in 2005—that separates the CPU/DRAM module and the NIC/Drives module on the motherboard. Redesigning the server to be modular enables us to upgrade the CPU/DRAM module while retaining the other components (such as fans, power supplies, cables, network switches, drives, and chassis) that are not ready for end-of-life.

We have found that the disaggregated design offers the following benefits:

- No need to replace perfectly good components.
- No need to reinstall the OS.
- Cuts refresh costs by a minimum of 44 percent (see Figure 11).
- Reduces technician time spent on refresh by 77 percent.
- Decreases refresh materials' shipping weight by 82 percent.

Our disaggregated server architecture has the potential to dramatically change how data centers around the world perform server refreshes—leading to significant refresh savings and the opportunity to quickly take advantage of the latest compute technology. This technology is already being used in Intel's data centers in Santa Clara, California, which feature the world's lowest power usage effectiveness (PUE) rating of 1.06.

The ability to spend less time and money on refreshing servers means Intel IT can afford to refresh faster, bringing the most advanced Intel Xeon processor-based technology into Intel's data centers. We are excited about the resulting opportunities to boost data center efficiency and more effectively power Intel's silicon design jobs.

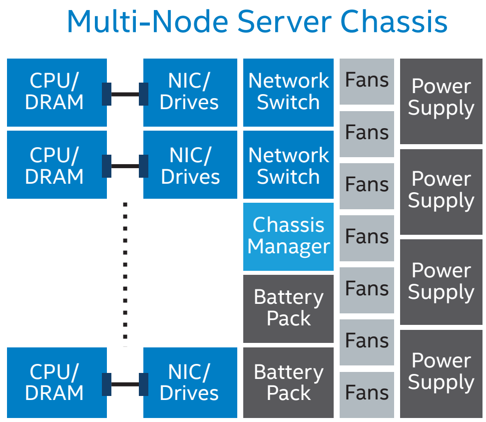


Figure 10. The disaggregated server architecture is characterized by a CPU/DRAM module and a NIC/Drives module that can be refreshed independently of each other and of the rest of the server components.

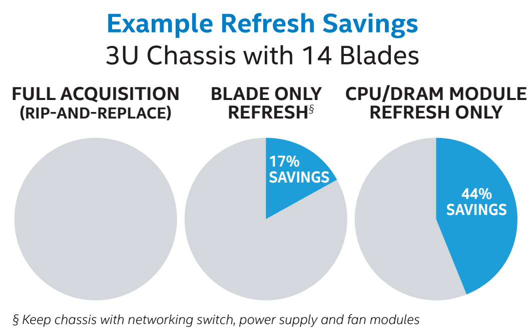


Figure 11. Refreshing the CPU/DRAM module in a disaggregated server saves at least 44 percent compared to a full-acquisition server refresh. Based on Intel internal testing, March 2017.

Increased Design Throughput Using NUMA-Booster

Overall data center optimization includes more than simply looking at server performance and facility efficiency. Application performance and workload optimization can also be contributing factors. We developed a system software capability called NUMA-Booster, which automatically and transparently intercepts our Design workloads and performs workload scheduling better than the default OS scheduling capability. This is implemented on all our two-socket servers.

We have achieved the following specific results without any system downtime or end-user impact:

- **Performance.** Our tests showed an average 17 percent improvement in design performance (see Figure 12).
- **Data center space and procurement costs.** We have deployed NUMA-Booster on approximately 40,000 servers, thereby reducing the footprint needed to meet demand by 4,451 servers (representing 92 racks of data center space).
- **Carbon footprint.** These 4,451 servers represent a savings of approximately 15.98 million kWh annually, which equals about 9,472 metric tons of CO_2.

We expect to reap even greater results from NUMA-Booster as we retire older servers and deploy newer multicore servers with NUMA-Booster capability.

Increased Design Throughput Using Intel® Solid State Drives as Fast-Swap Drives

Silicon chip Design engineers at Intel face the challenge of integrating more features into ever-shrinking silicon chips, resulting in more complex designs. The increasing design complexity creates large electronic design automation workloads that have considerable memory and compute requirements. We typically run the workloads on servers that need to be configured to meet these requirements in the most cost-effective way.

In Intel IT tests with large silicon design workloads, substituting lower-cost Intel® Solid State Drives (Intel® SSDs) for part of a server's physical memory resulted in a 1.63x performance-normalized cost advantage. Using SSDs as fast-swap drives increased the Design throughput of more than 13,000 servers by 27 percent.

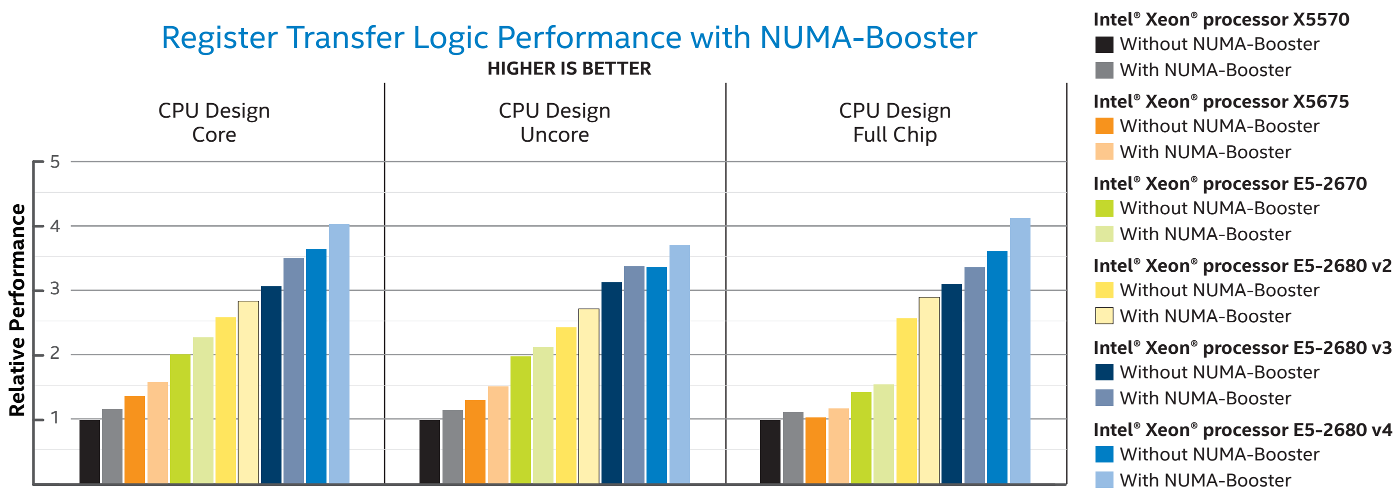


Figure 12. NUMA-Booster has increased Design compute performance by 17 percent.

System with 2x Intel® Xeon® processor X5570, 72 GB DDR3-1333 RAM, 1x 900 GB 10K RPM SAS hard drive, with Linux* 2.6 OS, running Intel silicon design simulation workload, Intel IT measurement.
System with 2x Intel® Xeon® processor X5675, 96 GB DDR3-1333 RAM, 1x 900 GB 10K RPM SAS hard drive, with Linux 2.6 OS, running Intel silicon design simulation workload, Intel IT measurement.
System with 2x Intel® Xeon® processor E5-2670, 128 GB DDR3-1333 RAM, 1x 900 GB 10K RPM SAS hard drive, with Linux 2.6 OS, running Intel silicon design simulation workload, Intel IT measurement.
System with 2x Intel® Xeon® processor E5-2680 v2, 256 GB DDR3-1600 RAM, 1x 900 GB 10K RPM SAS hard drive, with Linux 2.6 OS, running Intel silicon design simulation workload, Intel IT measurement.
System with 2x Intel® Xeon® processor E5-2680 v3, 256 GB DDR4-2133 RAM, 1x 900 GB 10K RPM SAS hard drive, with Linux 3.0 OS, running Intel silicon design simulation workload, Intel IT measurement.
System with 2x Intel® Xeon® processor E5-2680 v4, 256 GB DDR4-2400 RAM, 1x 1.2 TB 10K RPM SAS hard drive, with Linux 3.0 OS, running Intel silicon design simulation workload, Intel IT measurement.

Optimizing Servers to Meet Increasing Compute Demand

To achieve continually faster time-to-market improvements, given the ever-growing complexities in Intel® silicon design, Intel IT provides a global framework for parallel hardware and software design of numerous System on a Chip platforms and IP blocks.

Matching single-socket servers and highly scalable server configurations in our data centers yields 25 to 30 percent faster product design and architecture validation processes. By leveraging a global scheduling mechanism pooling compute capacity of over 160,000 servers at multiple sites around the world, our design hub provides scalable capacity and delivers optimal memory and compute capability in a shorter amount of time.

More Efficient Office and Enterprise Compute and Storage

Like our Design environment, the compute and storage demands in our Office and Enterprise environment are also growing quickly. Nevertheless, as shown in Figure 13, we continue to meet that demand while maintaining the number of physical servers over the last three years. From 2009 to 2016, we achieved an approximate 6.1x increase in the number of virtual machines (VMs). We also greatly increased average VM density per physical server—from 11 VMs in 2009 to 26 VMs in 2016 due to improved server platforms.

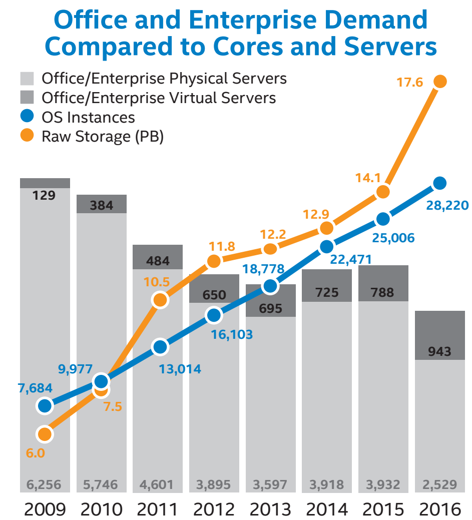


Figure 13. A high rate of virtualization combined with Intel® architecture has enabled us to meet growing Office and Enterprise compute and storage demand while significantly decreasing the number of required physical servers.

Results from 2010 to 2016

Our strategic approach has enabled us to deliver data center infrastructure best suited to meet our complex and ever-increasing compute needs while transforming our cost structure. By applying the innovative data center techniques listed in this paper, we have achieved unit cost levels that are significantly lower than if we were to host our workloads using public cloud infrastructure (Figure 14). Our workloads and our ability to achieve a high server utilization are particularly well suited towards private cloud investment. Over the seven-year period, 2010-2016, we have garnered combined capital and operational savings in excess of USD 1.5 billion, which help fuel our continuous innovation cycle.

Here are some examples of the efficiency improvements and cost savings we have achieved in the Design environment from 2010 through 2016:

- **Design computing.** Intel IT innovations in the Design computing data center include disaggregated server innovation (44 percent savings during refresh); the NUMA-Booster solution (17 percent higher performance); Intel SSDs (27 percent higher capacity at lower cost); faster servers (35 percent higher performance); and procurement efficiency. Together, these innovations have resulted in USD 284 million in savings.
- **Design storage.** We have implemented Design computing data center storage efficiency improvements by adopting new technology capabilities and increasing utilization—generating USD 45 million in savings.
- **Design network.** The adoption of a multi-vendor strategy for our Design computing data center network, combined with a focus on reduction of expensive maintenance costs associated with older equipment, has generated USD 13.32 million in savings.

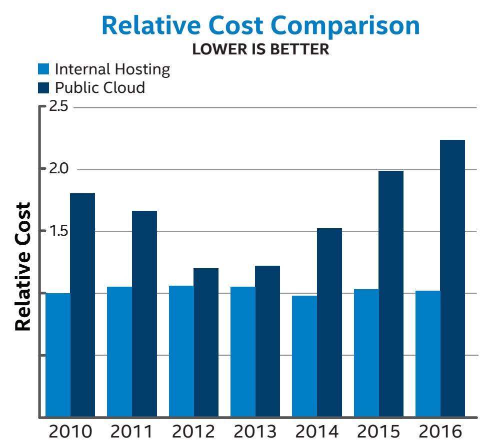


Figure 14. Total Cost of Ownership analysis including: Servers, storage, network and operational costs shows internal hosting of our data center workloads is significantly less expensive than if we had used public cloud services.

In addition to the contributions to the Design-specific environment, our new investment model has enabled us to identify other actionable gaps between the best achievable performance and our current plan. These actions include the following:

- Continuing to consolidate data centers
- Reducing unit cost for both the Design and Office and Enterprise environments
- Extending our use of blade servers in the Office, Enterprise, and Services environments

Continued Data Center Consolidation

We used our new investment model to look at the number of data centers we have and the number we should have. The new investment model identified opportunities to reduce the number of Intel data centers by as much as 35 percent, using techniques such as the following:

- Closing, retrofitting, or reclassifying data centers and improving efficiency.
- Co-locating local infrastructure with Design and Manufacturing data centers or providing services from a server closet.
- Managing local infrastructure sites remotely.
- Improving facility power efficiency through strategic investments.

We have targeted 32 inefficient data centers, eliminating 61,770 square feet of data center space and converting 23,609 square feet of data center space to low-cost infrastructure rooms, saving Intel USD 25.45 million annually.

Reduced Unit Cost for the Design Environment

Figure 15 details how our budget has remained relatively flat (white background of figure) while unit growth (grey background of figure) has continued to rise in both the Design and Office and Enterprise environments. Our investment model has enabled us to reduce unit costs in both environments (blue background of figure)—reducing Design unit cost by 85 percent and Office and Enterprise unit cost by 77 percent. Figure 16, on the next page, shows the same 85 percent unit cost reduction in Design and 77 percent unit cost reduction in Office and Enterprise by relational pie charts.

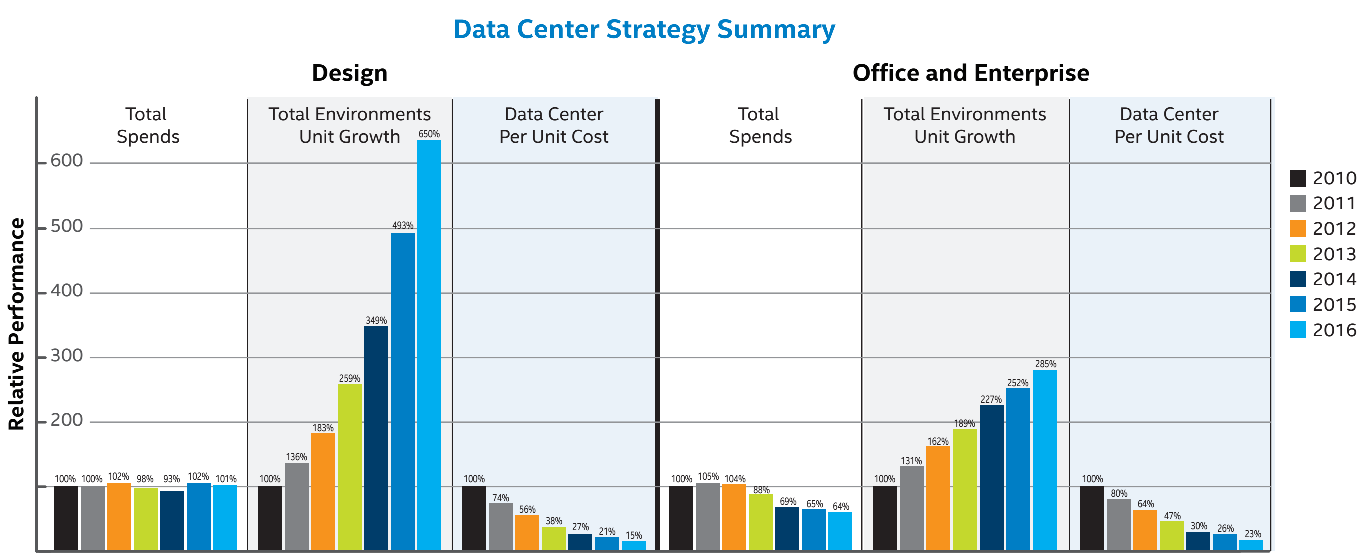


Figure 15. Our new strategy has enabled us to meet increasing growth and reduce unit cost without increasing our budget.

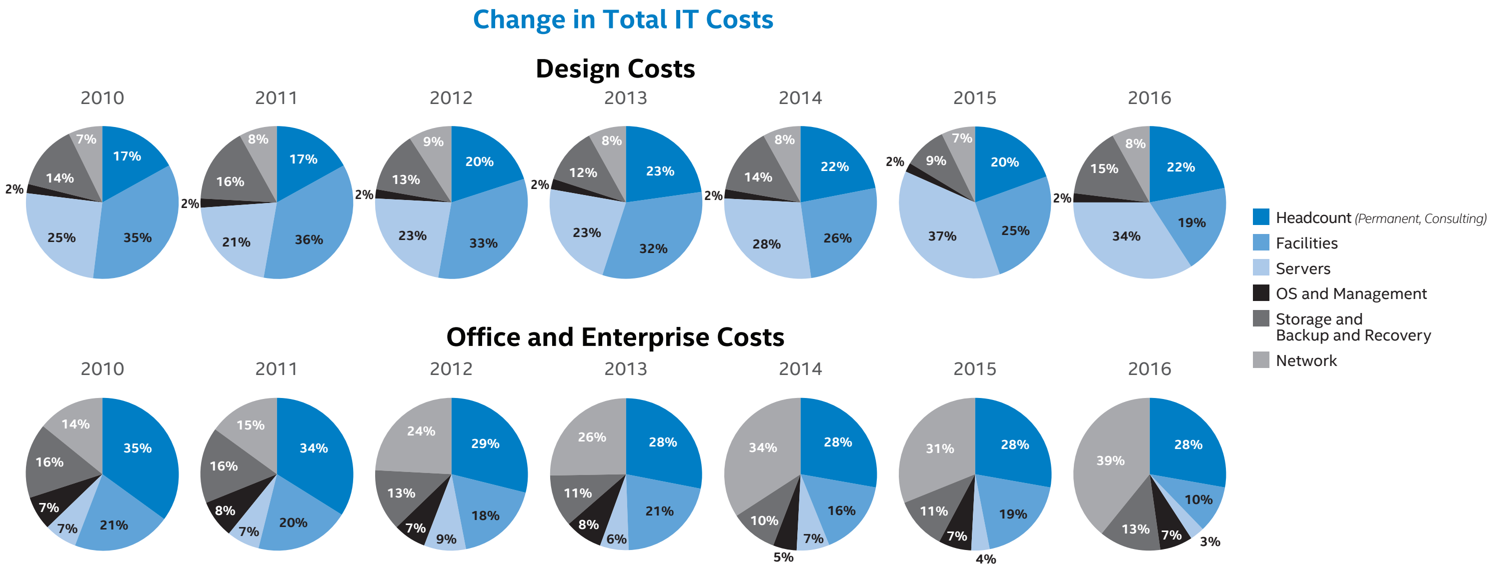


Figure 16. The pie charts reflect how our total IT costs have changed with the new data center strategy.

Reduced TCO with Blade Server Technology

As shown in Figure 17, our new investment model has shown us that moving from rack-mount servers to blade servers can reduce TCO in our cloud computing environment by about 39 percent. This reduction results from reduced port, network, and cable costs. For example, a group of 16 blade servers compared to 16 rack-mount servers requires only 8 Ethernet interfaces instead of 128, and only four Fibre Channel interfaces instead of 32. Deploying a newer generation of blade-server technology with converged network fabric within the blade chassis (labeled "Gen-2" in Figure 17) allowed us to reduce the cost even further.

Based on this data, we are actively deploying blade servers to support further virtualization efforts in the Office, Enterprise, and Services environments.

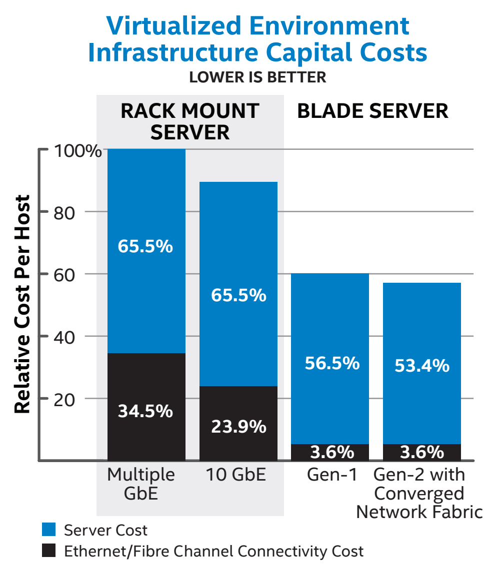


Figure 17. Blade servers with integrated network/data fabrics have reduced our virtualized environment infrastructure costs by 39 percent.

Summary of Data Center Best Practices

Over the last decade, we have made many strategic investments and developed solutions to make our data centers more efficient and better serve the needs of Intel's business. We are now applying our MOR approach across our entire infrastructure stack—compute, storage, networking, and facilities. Table 4, on the next page, provides a summary of the best practices we have developed and the business value they have generated.

Table 4. Intel IT Data Center Best Practices and a Few Examples of Business Value

BEST PRACTICE	BUSINESS VALUE
Compute (Servers)	
Migrate applications from RISC to Intel® architecture[i]	• Enabled significant savings and IT efficiencies • Allowed us to realize the benefits of industry-standard OSs and hardware
Adopt cloud computing	• Virtualized more than 90 percent of Office and Enterprise servers • Reduced the time it takes to provision a server from 90 days to on-demand provisioning • Adopt Value based Multi-Cloud Strategy
Regularly refresh servers using the latest generations of Intel® Xeon® processors	• Virtualization ratios of up to 60:1 • Reduced Design environment energy consumption by 10 percent annually between 2008 and 2013 • ~10x increase in performance between 2005 and 2016
Deploy high-performance computing[ii]	• 121x increase in capacity, with a 161x increase in stability • Saved USD 44.72 million net present value with HPC-1
Enhance server performance through software optimization	• Increased Design job throughput up to 49 percent • Delivered USD 284 million from 2010 to 2016 from various optimizations including disaggregated servers, NUMA-Booster, fast-swap based on Intel® Solid State Drives, and high-frequency servers and optimal workload to platform pairing
Storage	
Refresh and modernize storage using the latest generations of Intel Xeon processors	• Take advantage of new technology to increase storage capacity, quality, velocity, and efficiency at a lower cost • More than twice the I/O throughput than older systems • Reduced our data center storage hardware footprint by more than 50 percent in 2011-2012 • Reduced backup infrastructure cost due to greater sharing of resources • Tiered backup solutions to optimize backup costs and improve reliability
Right-size storage solutions using a tiered model[iii]	• Provide storage resources based on business needs: performance, reliability, capacity, and cost • Better management of storage costs while still enabling easy access to necessary data • Transition to scale-out storage to reduce operational complexity in tiering data • Automated policy-based data migration between tiers
Continuously monitor and reclaim disk space consumed by aged data	• More than USD 1 million in capital expenditure avoidance in 2011
Implement thin provisioning and deduplication for storage resources	• Helps control costs and increase resource utilization without adversely affecting performance • Increased storage effective utilization in Design from 46 percent in 2011 to more than 70 percent now
Network	
Upgrade the data center LAN network architecture to 10 gigabit Ethernet[iv]	• Increased data center network bandwidth by 400 percent over three years, enabling us to respond faster to business needs and accommodate growth • Increased the network utilization from 40 percent to 68 percent between 2010 to 2016 • Eliminated spanning tree with multi-chassis link aggregation (MLAG) and Layer 3 protocol • Reduced network complexity due to fewer network interface cards (NICs) and LAN ports • Reduced network cost in our virtualized environment by 18 to 25 percent
Open the data center network to multiple suppliers	• Generated more than USD 60 million in cost avoidance over five years with new network technology
Facilities	
Increase cooling efficiency	• Saved close to 16 million kilowatt-hours over 18 months, which is equivalent to reducing our carbon dioxide emissions by 6,800 metric tons
Use a tiered approach to redundancy, availability, and physical hardening	• Better matching of data center redundancy and availability features to business requirements • Reduced wasted power by more than 7 percent by eliminating redundant power distribution systems within a data center
Retrofit and consolidate data centers using a modular design	• Retrofitted old wafer fabrication plant to high-density, high-efficiency data center modules with industry-leading PUE of 1.06 • Utilized free-air cooling and environmentally efficient evaporative cooling for maximum energy efficiency • Avoided significant capital expenditures by not equipping the entire facility with generators • Quickly respond to changing data center needs with minimal effort and cost

[i] For more information see "Migrating Mission-Critical Environments to Intel® Architecture."
[ii] For more information see "High-Performance Computing for Silicon Design."
[iii] For more information see "Implementing Cloud Storage Metrics to Improve IT Efficiency and Capacity Management."
[iv] For more information see "Upgrading Data Center Network Architecture to 10 Gigabit Ethernet."

Plans for 2017 and Beyond

Our data center strategy is a continuous improvement process—we are always striving to close the gap between current achievements and the best possible scenario. To that end, we plan to explore the following areas and apply our MOR approach to them.

- **Embrace disruptive servers.** Deploy ultra-dense, power-optimized disaggregated server nodes to reduce data center space and power consumption for computing needs.
- **Adopt standards-based storage.** Use industry-standard hardware and software for scale-up and scale-out storage, to take advantage of the latest hardware more quickly—enabling higher throughput.
- **Drive network efficiency.** Continue to drive LAN utilization toward 75 percent and pursue top-of-rack architecture to support ultra-high-density data center designs. Implement 40 GbE and 100 GbE where appropriate and cost effective, to meet network capacity demands.
- **Increase facilities efficiency.** Use techniques such as higher ambient temperature for specific data center locations to take advantage of newer equipment specifications, which will reduce cooling needs.

Conclusion

To provide a foundation for continuous innovation that will improve the quality, velocity, and efficiency of Intel IT's business services, we have refined our data center strategy, building on the practices established over the last decade. Our refined data center strategy has created new business value exceeding USD 1.5 billion from 2010 to 2016. The data center transformation strategy (Figure 18) is a key for Intel IT to stay competitive.

Key achievements include the following:

- Our breakthrough disaggregated server design allows independent refresh of CPU and memory without replacing other server components, which results in faster data center innovation and a 44 percent cost savings compared to a full-acquisition refresh.
- We developed a system software capability called NUMA-Booster, which has saved millions while delivering additional usable server capacity.
- We deployed more than 13,000 Intel SSDs as fast-swap drives, which increased server capacity by 27 percent.
- Five generations of HPC in our design computing environment created a 121x capacity increase and a 161x quality improvement.
- We adopted new storage capabilities, accelerated storage refresh, and focused on increasing utilization, resulting in additional storage capacity.
- We deployed more than 87,000 10 GbE network ports and 3,100 40 GbE network ports.

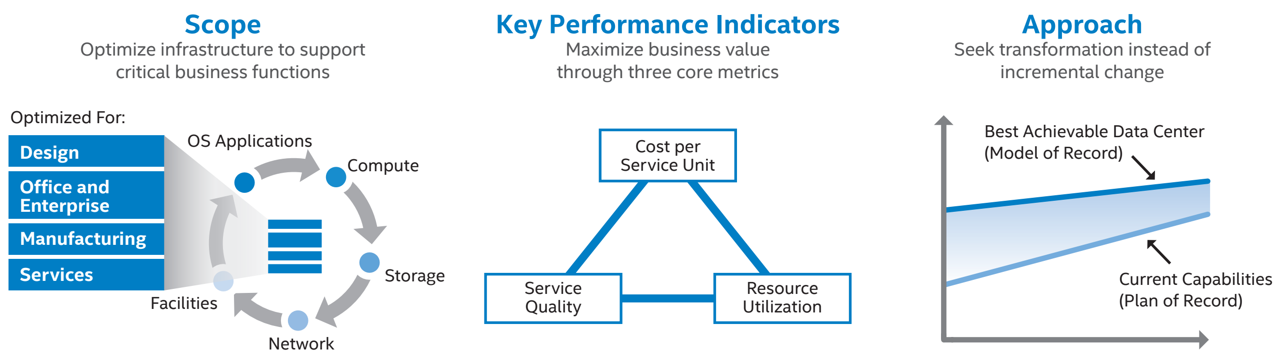


Figure 18. Maximizing the business value of Intel's data center infrastructure requires continued business-driven innovation in the areas of compute, storage, network, and facilities.

Share:

We have achieved these results by running Intel data centers like a factory, implementing change in a disciplined manner and applying breakthrough technologies, solutions, and processes. Transformational elements of our data center strategy include the following:

- **A focus on three primary KPIs.** These metrics enable us to measure the success of data center transformation: Meet growing customer demand (SLAs and QoS) within constrained spending targets (remaining cost-competitive) while optimally increasing infrastructure asset utilization (asset efficiency).
- **Stimulating bolder innovation by changing our investment model.** Comparing our current capabilities to a "best achievable model" encourages us to strive for innovation that will transform our infrastructure at a faster rate than if we sought only incremental change.
- **New unit-costing financial model.** This model enables us to better assess our data center TCO based on the business capabilities our infrastructure is supporting. The model measures the cost of a unit of service output and enables us to compare investments and make informed trade-off decisions across business functions—thereby maximizing ROI and business value.

For more information on Intel IT best practices, visit **intel.com/IT**.

IT@Intel

We connect IT professionals with their IT peers inside Intel. Our IT department solves some of today's most demanding and complex technology issues, and we want to share these lessons directly with our fellow IT professionals in an open peer-to-peer forum.

Our goal is simple: improve efficiency throughout the organization and enhance the business value of IT investments.

Follow us and join the conversation:

- Twitter
- #IntelIT
- LinkedIn
- IT Center Community

Visit us today at **intel.com/IT** or contact your local Intel representative if you would like to learn more.

Related Content

If you liked this paper, you may also be interested in these related stories:

- Disaggregated Servers Drive Data Center Efficiency and Innovation paper
- High-Performance Computing for Silicon Design paper
- Developing a Highly Available, Dynamic Hybrid Cloud Environment paper
- Collaborative Visual Data Analysis Enables Faster, Better Decisions paper
- Extremely Energy-Efficient, High-Density Data Centers paper
- How Software-Defined Infrastructure Is Evolving at Intel paper
- Improving Business Continuity with Data Center Capacity Planning paper
- Integrated Analytics Platform Helps Sales and Marketing paper
- Making Private-Public Cloud Decisions on the Way to a Hybrid Cloud paper