WHITE PAPER

Data Center

August 2019

IT@INTEL

Data Center Strategy Leading Intel's Business Transformation

By applying breakthrough technologies, solutions, and processes, we have optimally served the acceleration of Intel's business

Intel IT Authors

Shesha Krishnapura
Intel Fellow and Intel IT CTO

Shaji Kootaal Achuthan
Senior Staff Engineer

Pushpa Jahagirdar
Technical Program Manager

Vipul Lal
Senior Principal Engineer

Raju Nallapa
Principal Engineer

Sanjay Rungta
Senior Principal Engineer

Ty Tang
Senior Principal Engineer

Table of Contents

Executive Summary

Intel IT runs Intel data center services like a factory, affecting change in a disciplined manner and applying breakthrough technologies, solutions, and processes. This enables us to optimally meet Intel's business requirements while providing our internal customers with effective data center infrastructure capabilities and innovative business services.

Building on previous investments and techniques, our data center strategy has generated savings exceeding USD 2.8 billion from 2010 to 2018.

Over the next three years, we plan to extend the data center strategy to continue our data center infrastructure transformation. We will accomplish this by using disruptive server, storage, network, infrastructure software, and data center facility technologies that can lead to unprecedented quality-of-service levels and reduction in total cost of ownership (TCO) for business applications—all while continuing to improve IT operational efficiency and being environmentally responsible.

USD 2.8 BILLION IN SAVINGS

THROUGH

44% SAVINGS WITH DISAGGREGATED SERVER DESIGN
compared to a full-acquisition refresh

40 PB FAST LOCAL SSD CACHE TO REDUCE NETWORK TRAFFIC
using Intel® SSDs

1 DAY

ONE-DAY DEPLOYMENT WITH OUR PROCESS TRANSFORMATION
for new physical server deployment

236x CAPACITY INCREASE IN OUR HPC ENVIRONMENT
and an 80x quality improvement

Background

Intel IT operates 56 data centers with a total capacity of 86 megawatts, housing approximately 289,000 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, and Enterprise (DOME)—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 Figure 1.

[1] Number of data centers and servers as of May, 2019. 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."

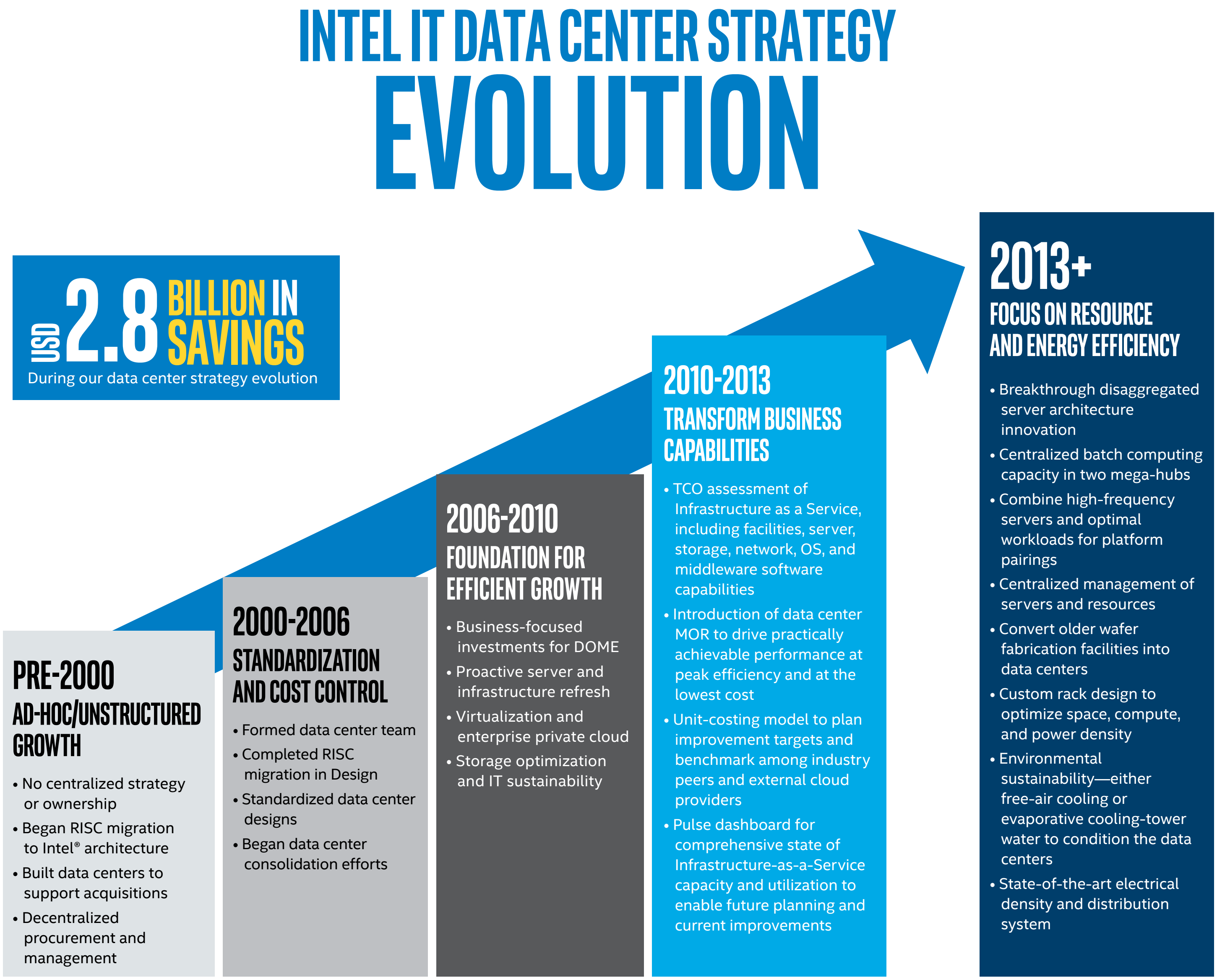


Figure 1. Intel's data center strategy is a continuous improvement process.

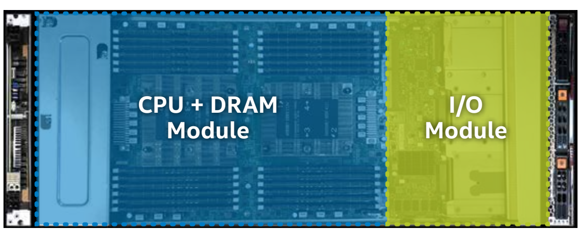
Meeting Compute Environment Challenges

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.

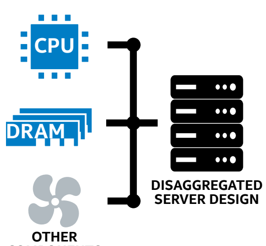
Over the last several years, we have reduced data center energy consumption and greenhouse gas emissions, while at the same time meeting the constantly increasing demand for data center resources. We anticipate these annual growth rates to continue or even increase further:

- 30 to 40 percent in compute capacity requirements
- 35 to 50 percent in storage needs
- 30 to 40 percent 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 USD 2.8 billion in cost savings while supporting significant growth.



Breakthrough Disaggregated Server Architecture



By decoupling the CPU/DRAM and NIC/Drives modules from other server components, we can 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.

Learn More:

- In this Document: Disaggregated Server Innovation Reduces TCO and TCE
- White Paper: Disaggregated Servers Drive Data Center Efficiency and Innovation
- Blog: Disaggregated Servers
- Video: Mission - Green Computing

> "Since 2010, these techniques have enabled us to realize USD 2.8 billion in cost savings while supporting significant growth."

Aligning Data Center Investments with Business Needs

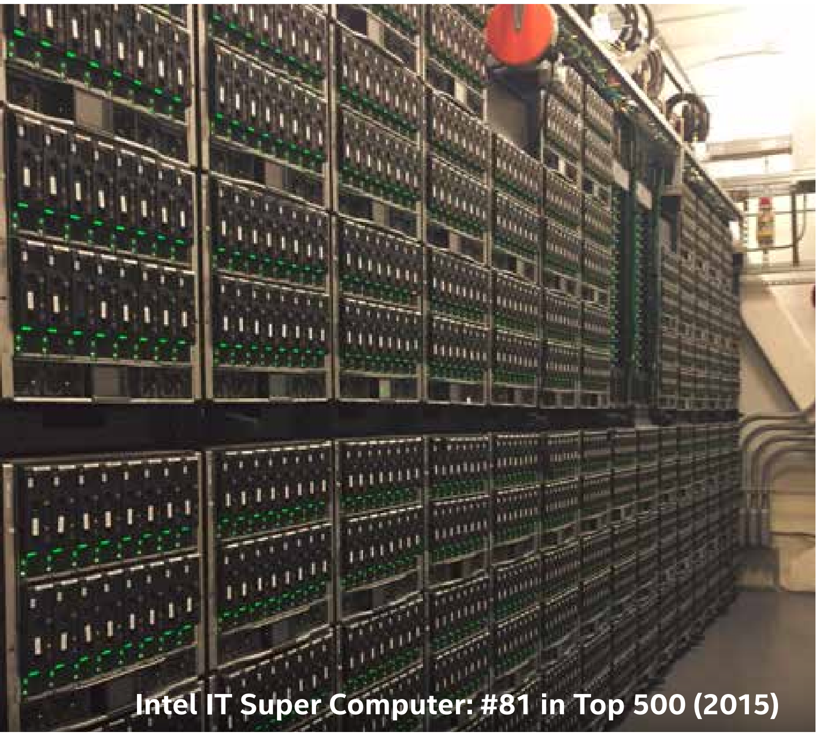
We have learned that a one-size-fits-all architecture is not the best approach for Intel's unique business functions. After working closely with business leaders to understand their requirements, we chose to invest in vertically integrated architecture solutions that meet the specific needs of individual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

Design

Design engineers run more than 166 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, and clustered local workstation computing.[2] We used Intel® solid state drives (SSDs) as fast local data cache drives, single-socket servers, and a specialized algorithm that increases the performance of the heaviest Design workloads. 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.

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 in combination with caching on local storage for our HPC needs. We use storage area networks 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.



Intel IT Super Computer: #81 in Top 500 (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 and Enterprise

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 a mature level of virtualization in our Office and Enterprise computing environment and have started the deployment of containers technology to further improve the agility in managing infrastructure and application, in software development and testing, and in scalable services deliveries.

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

Defining a Model of Record

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. Applying breakthrough technologies, solutions, and processes in an effective controlled manner can help us be an industry leader and to keep up with the accelerating pace of Intel's business.

Based on improvements each year in technologies, solutions, and processes, we use three key performance indicators (KPIs) to define a model of record (MOR) for the year. These KPIs—which are discussed in more detail in subsequent sections—include best achievable quality of service (QoS) and service-level agreements (SLAs); the lowest achievable unit cost; and the highest achievable resource utilization. We set investment priorities based on the KPIs to move toward the MOR goal. As shown in Figure 2 on the next page, each year we get closer to the MOR while at the same time balancing the KPIs.

We use five primary tactics to achieve our MOR goals:
- Embrace disruptive servers
- Adopt tiered storage
- Increase facilities efficiency
- Drive network efficiency
- Improve operational efficiency

More information is provided about each of these tactics in subsequent sections.

Learn More:
- Press Article: Inside Intel's 610K Core EDA System
- Podcast: Intel IT: Update on Intel's Data Center Strategy

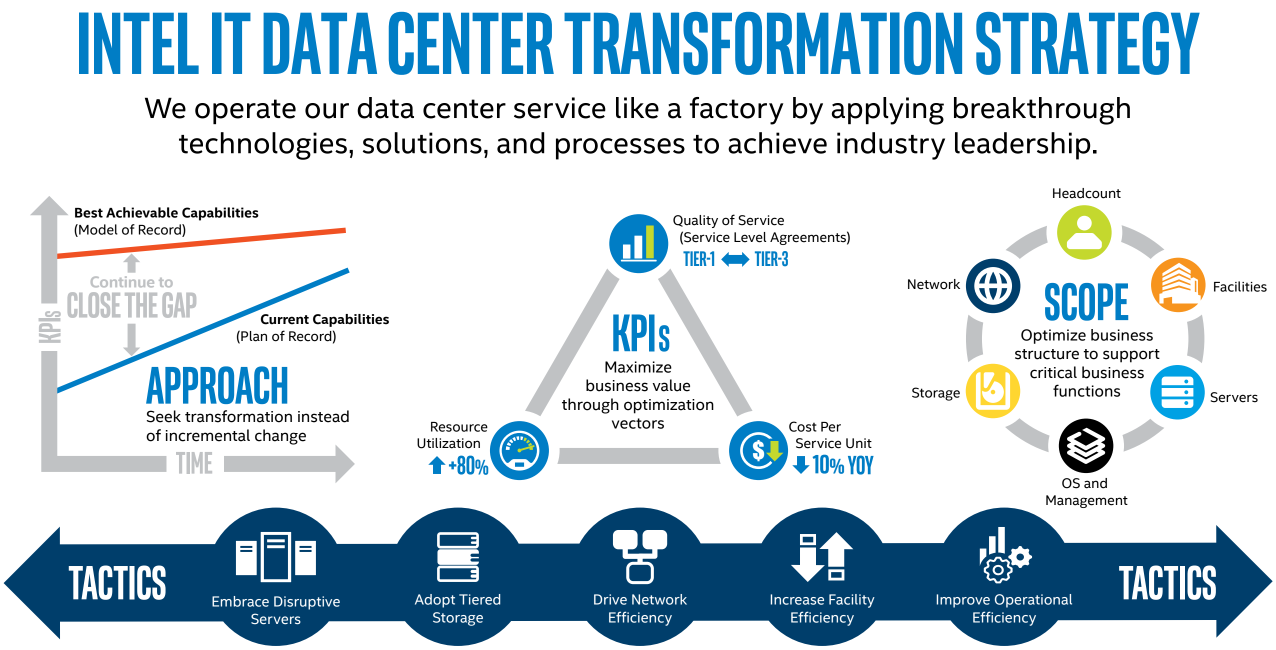


Figure 2. Maximizing the business value of Intel's data center infrastructure requires continued business-driven innovation in the areas of compute, storage, network, and facilities, while balancing KPIs to achieve the MOR.

> "Our new data center investment model encourages innovation and provides significant business results."

Achieving Economic Value

Our new data center investment model encourages innovation and provides significant business results. We have realized substantial cost savings since 2006 by proactively refreshing our infrastructure. For example, Intel® Xeon® processor-based servers have contributed significant economic value—our total server and data center infrastructure capital and operational costs have remained relatively flat from mid-2006 to mid-2019, while delivering substantially higher computational throughput as measured by a practical electronic design automation (EDA) workload. Further costs savings result from adopting cloud computing-like technologies, 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—DOME.

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 KPIs 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 them 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, 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.

Cost per Service Unit

As shown in Table 1, different business functions have a different service unit that we can measure. This unit represents the capacity we enable for our business users.

Table 1. Service Unit for Each Business Function

FUNCTION	SERVICE UNIT
Design	Cost per EDA-MIPS
Office and Enterprise	Cost per OS instance
Manufacturing	Cost per integrated factory compute environment

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.

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, 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.

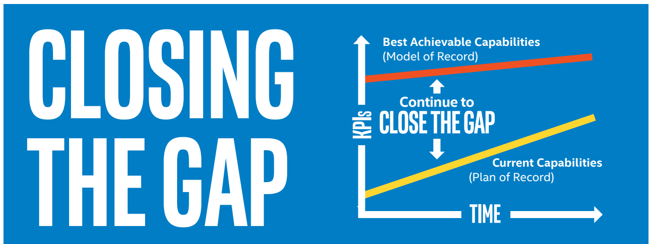
"Our goal for the effective utilization KPI is to achieve 80 percent effective utilization of all IT assets."

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.



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

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.

Because technology is always changing, peak performance also changes—the maximum achievable performance keeps improving through innovation. We know that resource constraints make it difficult to 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 return on 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 3, there are six major categories of cost to consider: headcount, facilities, servers, OS and manageability, storage and backup/recovery, and network. 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, which helps us identify the ones that are not performing optimally and decide whether to upgrade or consolidate underperforming data centers.

To show how the new unit-based costing model works, Figure 4 compares Design cost data and Office and Enterprise cost data. The headcount category shows equal percentage of total cost in Office and Enterprise and in Design; in contrast, servers are more of a cost factor in Design than they are in Office and Enterprise. 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.

Determining the Cost per Service Unit

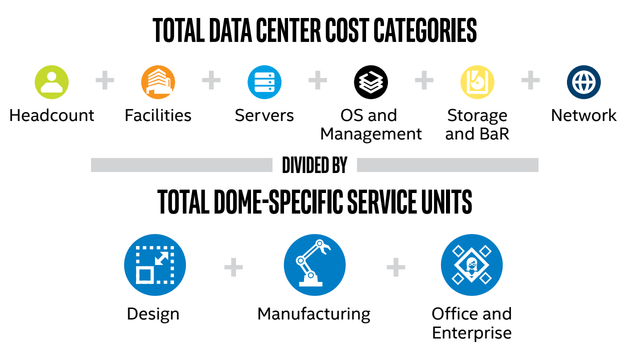


Figure 3. We arrive at a data center unit cost by considering all categories of cost and dividing by the number of units for that environment, such as EDA-MIPS in Design and OS instances in Office and Enterprise.

2018 Unit-based Costing of IaaS

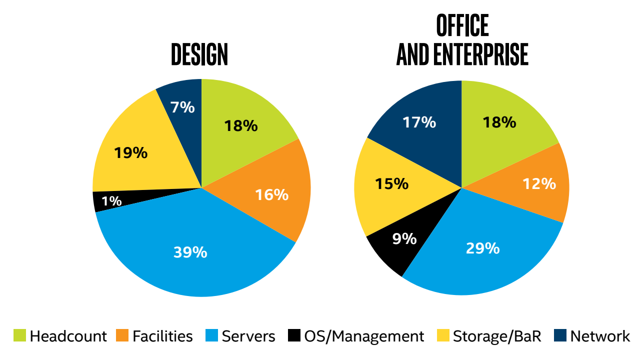


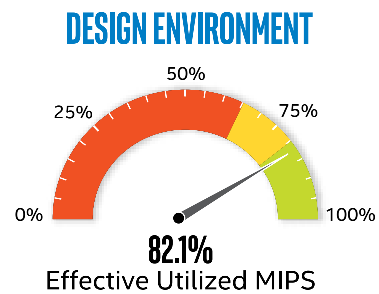
Figure 4. Knowing the total unit cost, as well as the individual cost category figures for each business environment, enables us to better choose IT investments that will lower costs the most.

Intel IT Data Center Dashboard

To better monitor and manage our worldwide network of data centers, we developed and deployed an integrated Intel IT Data Center Dashboard. This dashboard 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.

Results: Building on the Past, Building for the Future

This section details some of the improvements and cost savings our data center strategy has enabled over the years, using our five primary tactics of embracing disruptive servers, adopting tiered storage, increasing facilities efficiency, driving network efficiency, and improving operational efficiency. We are building on previous successes. Therefore, some of the results shown here are cumulative; others have been achieved over the last three years as a direct result of our MOR strategy. Our refined data center strategy enables 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.

We have dramatically improved performance and reduced costs for our data centers (Table 2).

Table 2. Data Center Improvements from 2003-2018

DATA CENTER-WIDE

- Smaller total data center footprint
- Improved overall storage and network practices
- Increased data center facilities efficiency

DESIGN ENVIRONMENT

- Deployed disaggregated servers
- More efficient Design compute and storage
- 5th generation of high-performance computing (HPC)
- Increased Design throughput using NUMA-Booster
- Faster Design throughput using Intel® SSDs

OFFICE AND ENTERPRISE ENVIRONMENT

- More efficient Office and Enterprise compute and storage

Understanding Disaggregated Server Architecture

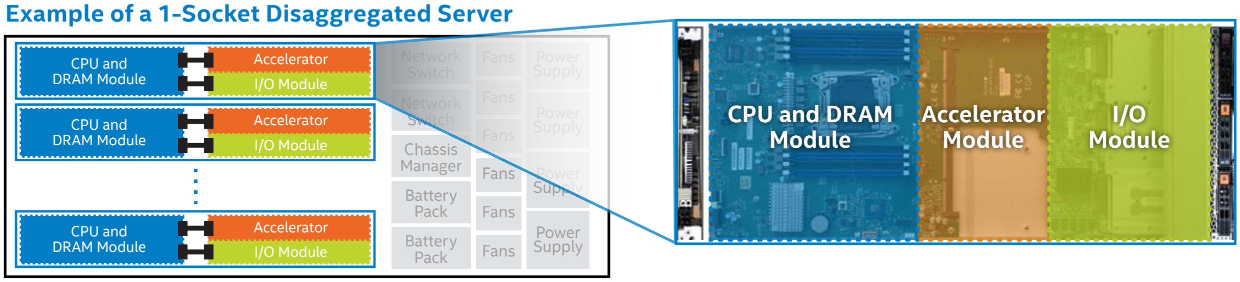
As shown below, Intel IT has developed a disaggregated server architecture—the first major server innovation since the introduction of blade servers in 2005—that separates the CPU/DRAM module and the NIC/Drives module on the motherboard. Redesigning the server to be modular enables us to upgrade the CPU/DRAM module while retaining the other components (such as fans, power supplies, cables, network switches, drives, add-on module/accelerator, and chassis) that are not ready for end-of-life.

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

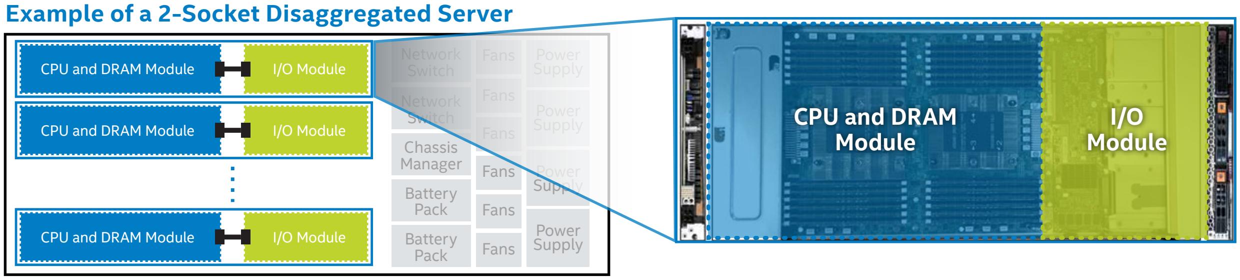
We have found that the disaggregated design offers the following benefits:

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

Example of a 1-Socket Disaggregated Server



Example of a 2-Socket Disaggregated Server



Disaggregated Server Innovation Reduces TCO and TCE

One of our leading tactics to achieve our MOR goals is to adopt disruptive server technology. To this end, we are deploying disaggregated servers throughout our data centers. Just as it makes little sense to replace an entire light fixture when all that is needed is a more energy-efficient or powerful light bulb, replacing an entire server is not necessary when all that is needed is a more advanced CPU and DRAM.

Our disaggregated server architecture has the potential to dramatically change how data centers around the world perform server refreshes—leading to significant refresh savings (see Figure 5) 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 have the world's best 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. We have deployed over 126,000 disaggregated servers so far, based on multiple generations of Intel Xeon processors.

In addition to the TCO benefits of 44 percent lower refresh cost over a full acquisition (rip-and-replace) refresh, reduced provisioning time of 77 percent, and reduced shipping costs, disaggregated servers have total cost to environment (TCE) benefits of 82 percent reduction in material shipping weight and significantly reduced e-waste.

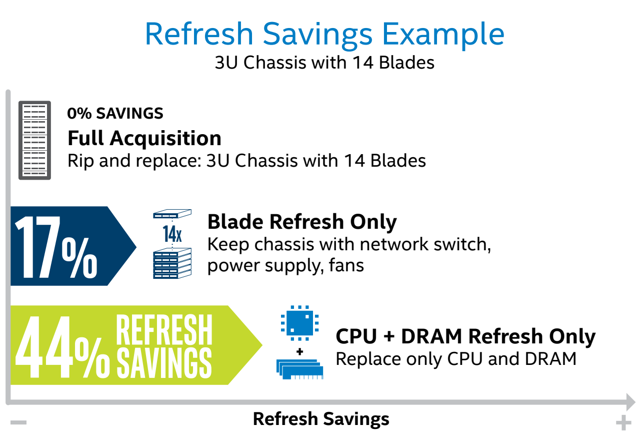


Figure 5. Refreshing the CPU/DRAM module in a disaggregated server saves at least 44 percent compared to a full-acquisition server refresh. Based on Intel internal testing, March 2017.

Adopting Tiered Storage and Other Storage Techniques

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.

Tiered storage is foundational to meeting our MOR goals. A four-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, and Tier-4 servers have the highest capacity but are used for low-frequency access and read-only archived 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.

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

- **Scale-out storage.** We have executed 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 by using efficiency technologies such as deduplication and compression. We are performing storage scaling on-demand for read-only storage areas, which require extremely high access rates. We have also enabled high-performance shared scratch spaces to meet the demand from our hyperscale EDA compute environment.
- **Storage refresh cycle.** To improve performance and reduce cost, we implemented an efficiency-based refresh cycle that 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 also reduce our data center footprint while delivering better performance for our customers at a similar or lower cost per TB.
- **Data reduction.** 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,

compaction, and compression on our primary storage servers—freeing more than 61 PB of capacity, which we are making available for our users. We continue to work closely with our internal design teams to optimize their design flows to reduce the growth rate of their data and IOPS requirements, dynamically adjust the allocations based on usage, and over-allocate capacity.

Increasing Facilities Efficiency

We used our new investment model to evaluate the number of data centers we currently have and the number we should have. The new investment model identified opportunities to reduce the number of data centers 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 since 2011, eliminating 61,770 square feet and converting 23,609 square feet of data center space to low-cost infrastructure rooms, saving Intel USD 25.45 million annually.

Figure 6 shows how we have consolidated our data center facilities from 2003-2019. We have reduced the total square footage by 30 percent and reduced the number of data centers from 152 to 56, while at the same time increasing our data center compute capacity and commissioned power by 72 percent from 50 MW to 86 MW over the last seven years. From 2012-2018, we have saved over 546 million KW hours compared to industry-standard data centers.

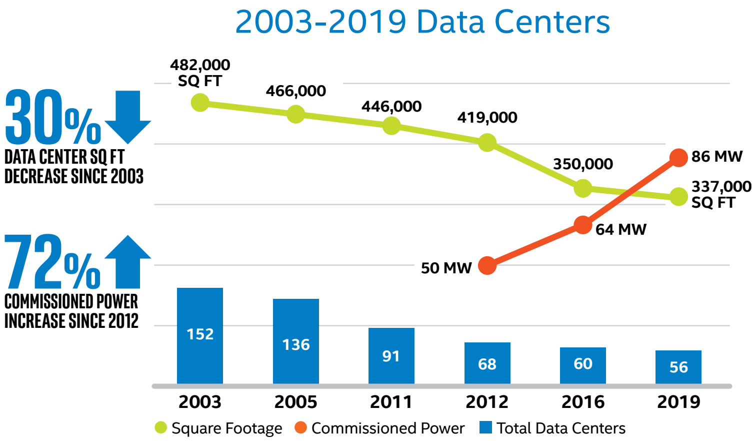
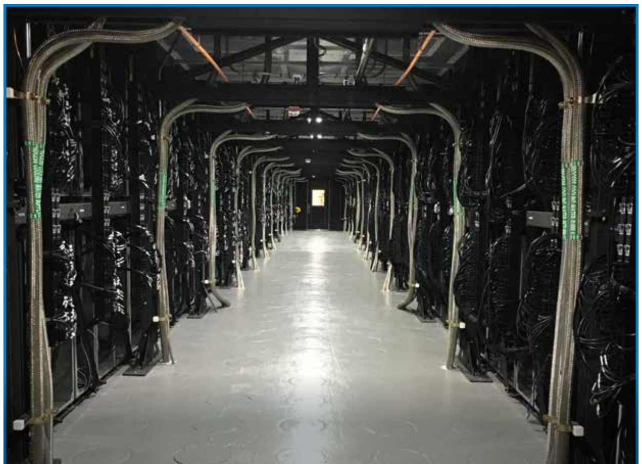


Figure 6. Even as we have met increasing demands for compute and storage resources over the years, we have reduced our data center space footprint by 30 percent, while increasing the power density and capacity.

"From 2012-2018, we have saved over 546 million KW hours compared to industry-standard data centers."



Gen 3: Close-coupled evaporative cooled data center hot aisle with 60U rack, up to 43 KW/rack, and up to 280 disaggregated servers/rack.

Data Center Evolution at Intel

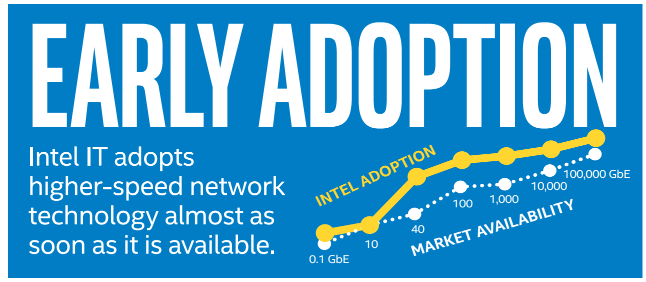
Driving up density while driving down PUE

Intel IT is continually honing data center design to increase density and efficiency. Since the 1990s, our data centers have evolved through three generations.

- **Gen 1 (1990s).** Characterized by forced chilled air from the ceiling, with no hot/cold air segregation, these early data centers could accommodate 42U racks with a power consumption of 5 KW—resulting in a power usage effectiveness (PUE) of more than 2.0. Data centers that used chilled air from the row end had a PUE of about 1.40.
- **Gen 2 (early to mid-2000s).** With improvements such as raised-floor forced chilled air or hot/cold air segregation including chimney racks, density stayed at 42U, but power consumption delivered to the racks increased to as much as 30 KW, resulting in a lower PUE of about 1.18.
- **Gen 3 (2013 and beyond).** Our modern data centers use free air cooling or close-coupled evaporative cooling to achieve an industry-leading PUE of 1.06, with an extreme rack density of 60U and up to 43 KW/rack.

Learn More:

- White Paper: Extremely Energy-Efficient High-Density Data Centers



Driving Network Efficiency

To accommodate the increasing demands that data center growth places on Intel's network, in 2010 Intel IT began to convert our data center network architecture from multiple 100 Mb/s and 1 gigabit Ethernet (GbE) connections to 10 GbE connections because the older, slower connections no longer supported Intel's growing business requirements. To meet today's scale and capacity demand, we are now migrating to 40 GbE and 100 GbE, depending on the use case.

We currently have deployed more than 128,756 10 GbE ports. Our new 10 GbE data center fabric design accommodates our current annual network capacity growth of more than 30 percent. We have already deployed 40 GbE and 100 GbE ports on the hosts in specific use cases where they provide business value. However, we are now migrating to multiple 100 GbE inter-switch links to keep up with the traffic growth.

In addition to increasing the network capacity, we have also increased the effective utilization of network ports over the last nine years from 40 percent to 70 percent (1.75x increase). Higher utilization means we do not have to purchase additional ports to meet network capacity demand growth. Figure 7 illustrates the growth in data center network port deployments, and Figure 8 shows the continual increase in port utilization.

We are also focusing on improving data center stability. In the past, we used a large installation of layer 2-based technology. We are 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. Using layer 3-based, scalable architecture within Intel's data center lets us plan for scale and resiliency. Also, we are using other technologies such as overlay, multi-chassis link aggregation, and tunneling to extend layer 2 across data centers, over the layer 3 topology.

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 adoption of 100 GbE technology in 2017, to keep pace with network demand. Due to the scale of the data center and new landing, we are focusing on zero-touch provisioning and automation to keep the design consistent. Our focus is on day-zero and day-one automation to expediate the landing and support.

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 70 percent of our Design data centers to use fixed form-factor switches using a modular design.

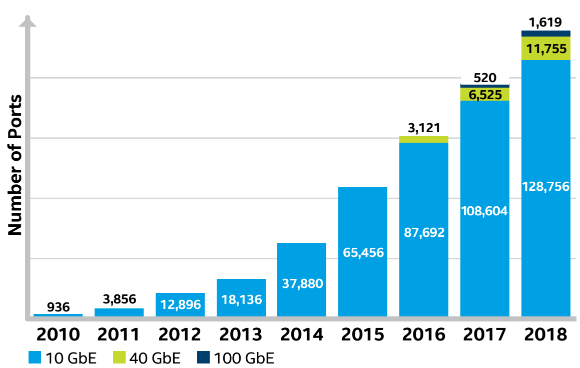


Figure 7. Implementing 10/40/100 GbE data center fabric design accommodates current capacity growth.

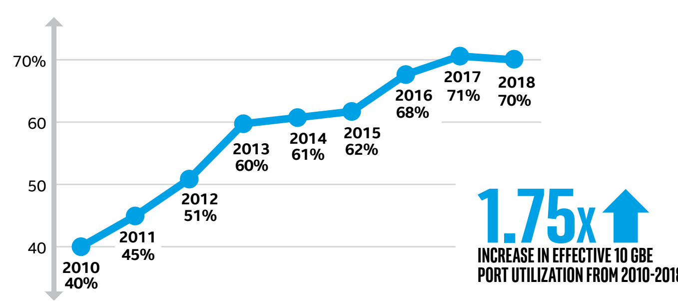


Figure 8. Effective utilization of network ports has increased by 1.75x between 2010-2018.

Achieving 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. Average annual growth rate of compute capacity demand over last 10 years is ~34%, while storage has grown annually at ~35% (see Figure 9).

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.

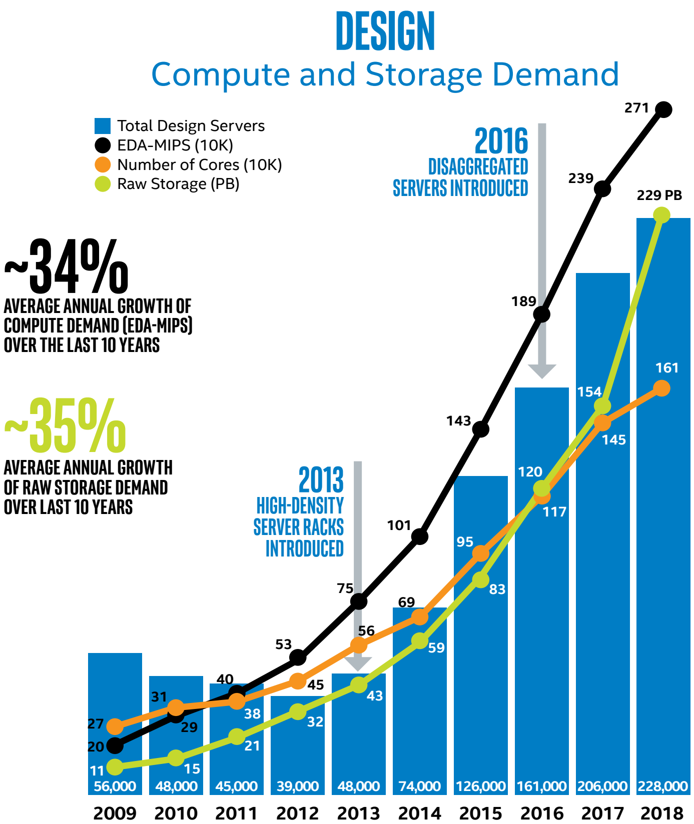
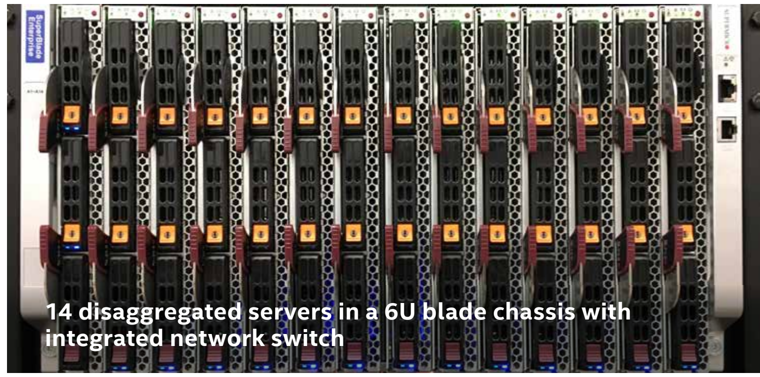


Figure 9. Despite continuing growth in compute and storage demand, our Design data centers are using powerful Intel® technology to meet demand. In 2013, we shifted to a new server form factor which shows server count increasing; however, it is highly dense, allowing us to place 140-180 servers per rack. In 2016, we again shifted to a new form factor—the disaggregated server—which allows us to place 280 servers per rack.

Learn More:
White Paper: Hyperscale HPC for Silicon Design



14 disaggregated servers in a 6U blade chassis with integrated network switch

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

As shown in Figure 10, our HPC solution has enabled a 236x growth in tapeout compute capacity from 2005 to 2018. 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 80x.

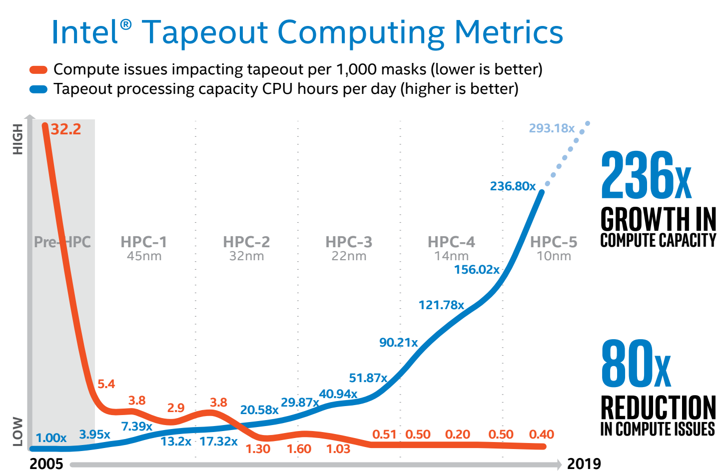


Figure 10. Our HPC solution, combined with disciplined change management, has steadily increased compute capacity and improved QoS.

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 batch 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 11).
- **Data center space and procurement costs.** We have deployed NUMA-Booster on approximately 31,300 servers, reducing the footprint needed to meet demand by 4,073 servers (representing 84 racks of data center space).
- **Carbon footprint.** These 4,073 servers represent a savings of approximately 14.62 million kWh annually, which equals about 10,339 metric tons of CO_2.

Increased Design Throughput Using Intel® SSDs as Fast Local Data Cache Drives

Intel® 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.

Intel IT has deployed over 40 PB of Intel SSD storage in over 20,000 servers as fast local data cache drives, improving workload performance due to reduced network traffic and storage demand.

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 228,000 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.

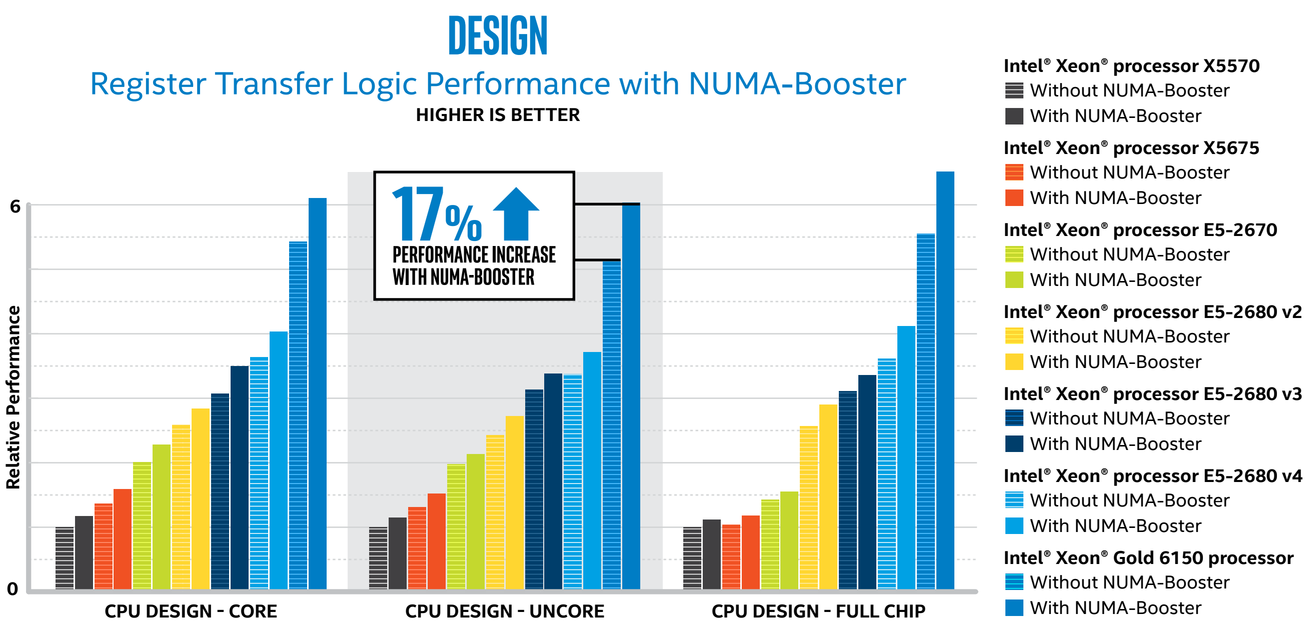


Figure 11. NUMA-Booster has increased Design compute performance by 17 percent. Intel IT measurement

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.
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.
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.
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.
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.
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.
System with 2x Intel® Xeon® Gold 6150 processor, 768 GB DDR4-2666 RAM, 2x 1.2TB 10K RPM SAS hard drive, with Linux 3.0 OS, running Intel silicon design simulation workload.

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 12, we continue to meet that demand while maintaining the number of physical servers over the last three years. From 2009 to 2017, we achieved an approximate 19x increase in the number of virtual OS instances. We also greatly increased average VM density per physical server—from 11 VMs in 2009 to 30 VMs in 2017 due to improved server platforms. In 2018, we implemented an aggressive VM reclamation strategy that led to a reduction of about 5,400 VMs. New workloads that were more cost effective to deploy on cheaper physical platforms than on a virtualized platform led to an increase in physical server counts. Process improvements and enhanced automation led to additional savings, and we are now deploying performance-based VMs.

OFFICE AND ENTERPRISE
Compute and Storage Demand

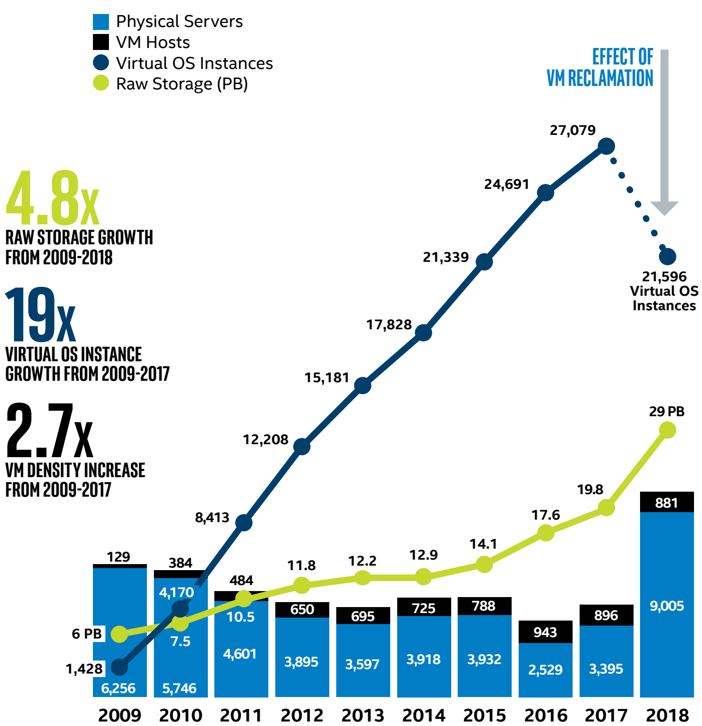


Figure 12. A high rate of virtualization combined with Intel® architecture has enabled us to meet growing Office and Enterprise compute and storage demand while significantly decreasing the number of required virtualization host servers.

Results from 2010 to 2018

Our strategic approach has enabled us to deliver a 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 13). Our workloads and our ability to achieve a high server utilization are particularly well suited towards private cloud investment.

Over the nine-year period, 2010-2018, we have garnered combined capital and operational savings in excess of USD 2.8 billion, which help fuel our continuous innovation cycle.

DESIGN
Relative Unit-Cost Comparison

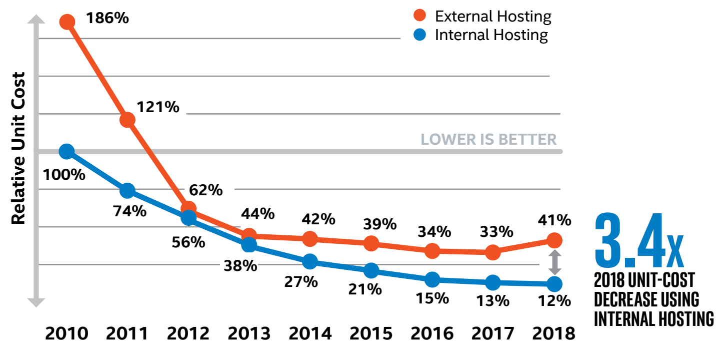


Figure 13. Unit cost including servers, storage, network, and operational costs shows private cloud hosting of our data center workloads is significantly less expensive than if we use public cloud services.

Reducing Unit Cost

Figure 14 (see page 16) details how our budget has remained relatively flat while unit growth 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 by 88 percent.

> "Our investment model has enabled us to reduce unit costs in both environments by 88 percent."

Design Improvement Examples

Below are some examples of the efficiency improvements and cost savings we have achieved in the Design environment from 2010 through 2018:

Computing. Intel IT innovations in the Design computing data center include disaggregated server innovation (44% savings during refresh); the NUMA-Booster solution (17% higher performance); Intel® SSDs (27% higher capacity at lower cost); faster servers (35% higher performance); single-day dock-to-production deployment and procurement efficiency.

Storage. We have implemented Design computing data center storage efficiency improvements by adopting new technology capabilities and increasing utilization.

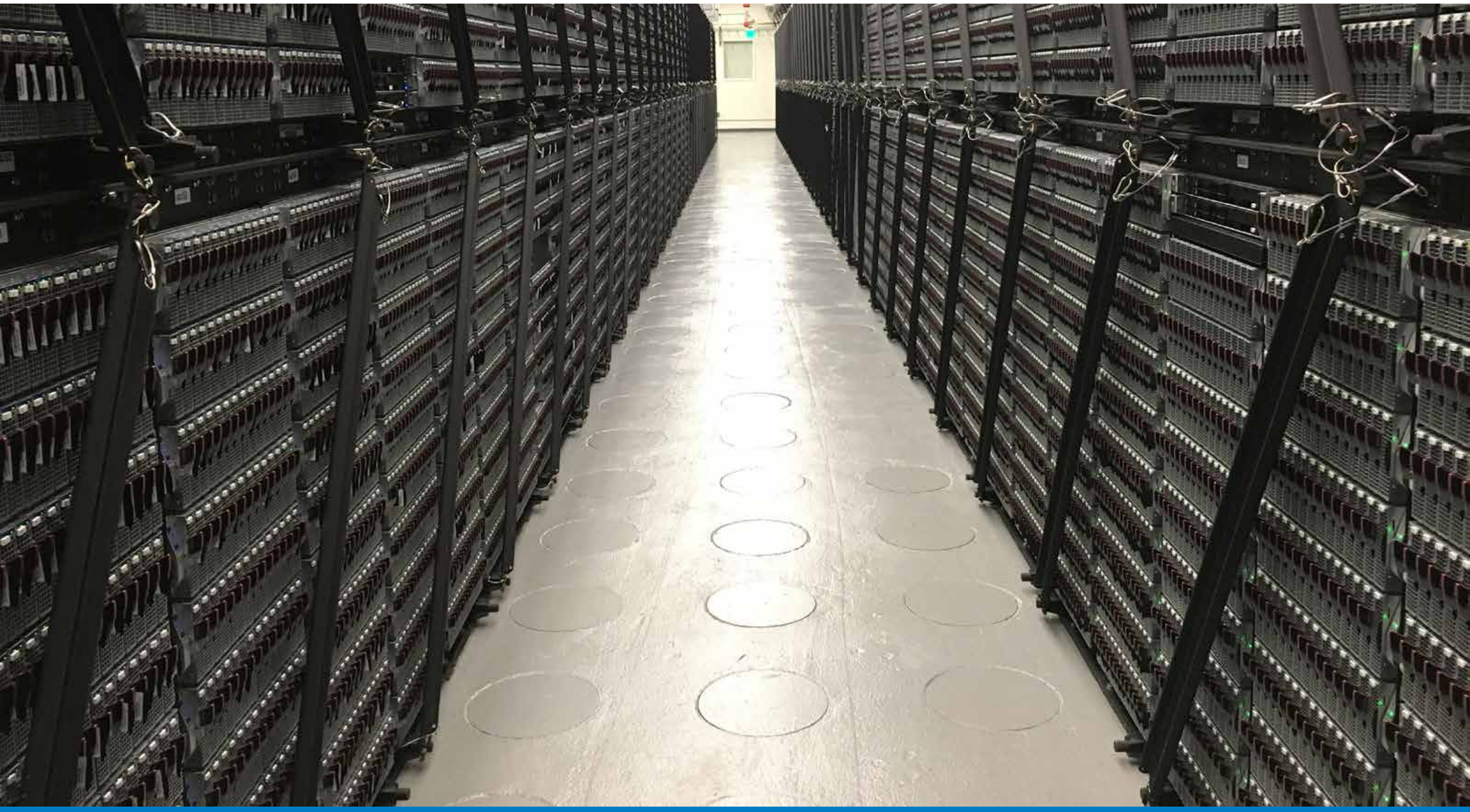
Network. We adopted a multi-vendor strategy for our Design computing data center network, combined with a focus on reduction of expensive maintenance costs associated with older equipment. As we adopt 100 GbE we are focusing on Intel® Silicon Photonics-based optics because that technology has significant cost advantage over laser-based optics.

Before implementing our data center strategy, we spent a third or more of our Design environment budget on facilities, and only a quarter or less on servers—but the servers are what power Intel's business success. Our new investment model has enabled us to reverse that ratio, now spending only 16 percent on facilities and almost 40 percent on servers. A similar transformation has occurred in the Office and Enterprise environment, with a lot of the growth driven by newer analytics and security workloads.

Summary of Best Practices

Over the last decade, we have made many strategic investments and developed solutions to enable our data centers to be more efficient and to 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 3 on page 17 provides a summary of the best practices we have developed and the business value they have generated.

"We are now applying our MOR approach across our entire infrastructure stack."



INTEL IT DATA CENTER STRATEGY
2010-2018 RESULTS

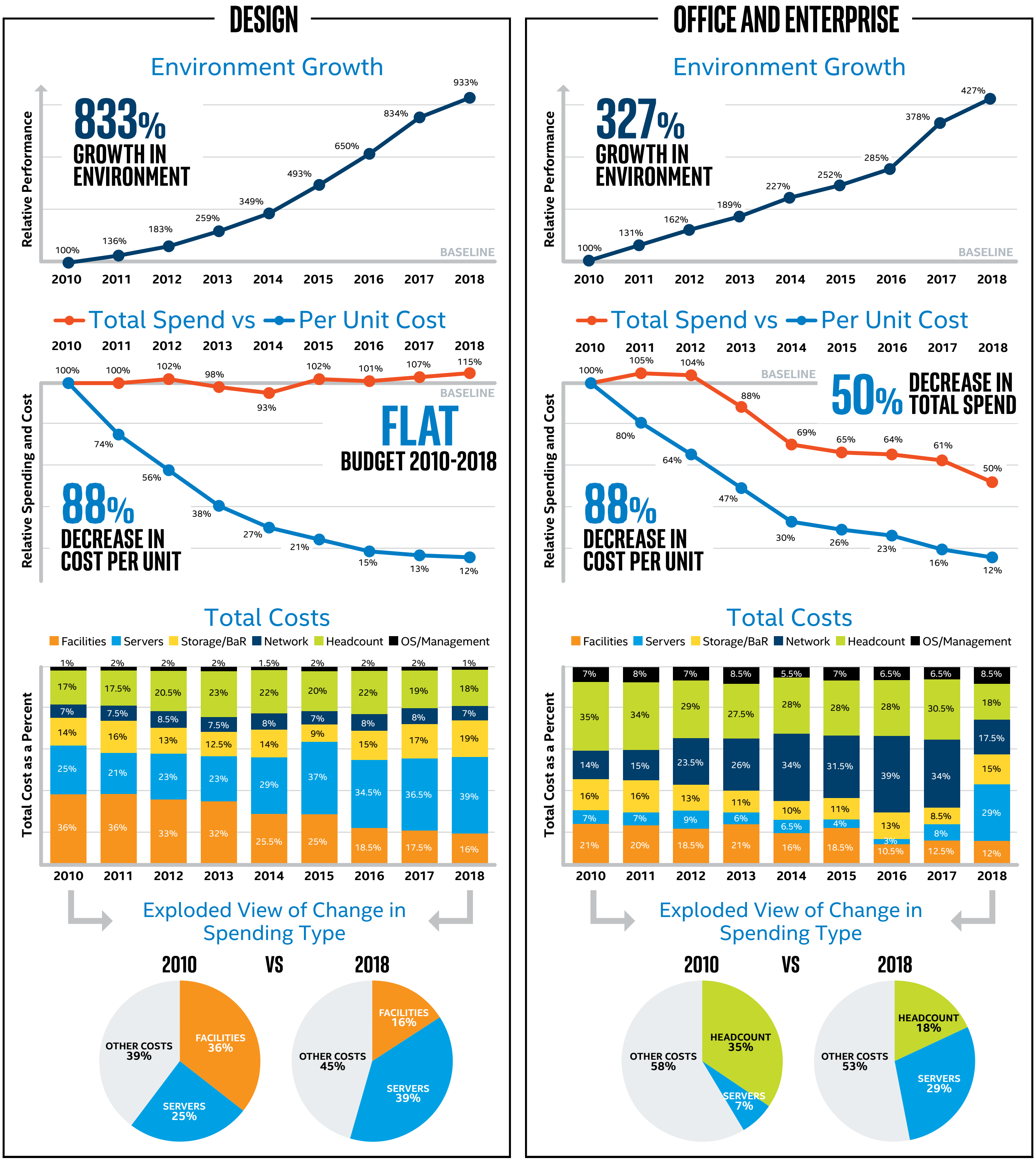


Figure 14. Our new strategy has enabled us to meet increasing growth and reduce unit cost without increasing our budget. As well, the makeup of our IT spending changed dramatically with the new strategy.

INTEL IT DATA CENTER STRATEGY BEST PRACTICES

Table 3. Intel IT Data Center Best Practices and a Few Examples of Business Value

Servers

BEST PRACTICE	BUSINESS VALUE
Adopt disaggregated servers	• Saves at least 44% over a full acquisition (rip-and-replace) refresh • Reduces provisioning time (IT technician labor) by as much as 77% • Decreases shipping weight of refreshed server material by 82%
Adopt elastic computing services and technologies	• Virtualized the majority of Office and Enterprise servers • Reduced the time it takes to provision a server from 90 days to on-demand provisioning using virtualization • Enabled containers as a service
Enable one-day dock-to-production for physical servers	• Upfront planning and process enhancement to order long-lead time items and rack readiness, reducing the dock-to-production release from 10+ days to one day
Deploy Intel® SSDs as the standard for local disk in all new servers	• Improved performance for I/O-intensive workloads and expected reduction of disk failure rates
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% annually between 2008 and 2013 • ~17x increase in throughput between 2005 and 2018
Migrate applications from RISC to Intel® architecture[i]	• Enabled significant savings and IT efficiencies • Allowed us to realize the benefits of industry-standard operating systems and hardware
Deploy HPC	• 236x increase in capacity during HPC-5, with an 80x increase in stability • Saved USD 44.72 million net present value during HPC-1 itself[ii]
Enhance server performance through software optimization	• Increased Design job throughput up to 49% • Delivered various optimizations including disaggregated servers, NUMA-Booster, fast local data cache based on Intel® SSDs, and high-frequency servers and optimal workload to platform pairing

Network

BEST PRACTICE	BUSINESS VALUE
Upgrade data center LAN architecture to support 10/40/100 GbE[iv]	• Increased data center network bandwidth by 400% over three years, enabling us to respond faster to business needs and accommodate growth • Increased the network utilization from 40 to 70% between 2010 to 2018 • Eliminated spanning tree with multi-chassis link aggregation and Layer 3 protocol • Reduced network complexity due to fewer network interface cards and LAN ports • Reduced network cost in our virtualized environment by 18 to 25%
Open the data center network to multiple suppliers	• Generated more than USD 60 million in cost avoidance over five years with new network technology
Deploy Intel® Silicon Photonics Optical Transceivers	• For large-scale 100 GbE deployment, leveraged Intel Silicon Photonics to significantly reduce the per-port cost

Storage

BEST PRACTICE	BUSINESS VALUE
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% 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 effective storage utilization in Design from 46% in 2011 to more than 70% now
Scale storage on demand and provide high-performance shared scratch spaces	• Enables higher workload throughput for read-only storage areas that require high access
Facilities	
Increase cooling efficiency	• Saved close to 16 million kilowatt-hours over 18 months, which is equivalent to reducing our carbon dioxide emissions by 11,314 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% 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 responded to changing data center needs with minimal effort and cost

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

Disaggregated server at scale

Plans for 2019 and Beyond

Our data center strategy is continuously improving—we are always striving to close the gap between current achievements and the best possible scenario. To that end, we plan to continue to apply the MOR approach to our primary enabling tactics:

- **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. Use strategic planning and storage protection technologies to deliver both backup and disaster-recovery coverage while reducing backup cost.
- **Increase facilities efficiency.** Use techniques such as higher ambient temperature for specific data center locations to take advantage of newer equipment specifications, which will help reduce cooling needs.
- **Drive network efficiency.** Continue to drive LAN utilization toward 75 percent and pursue software-defined networking to support agile, ultra-high-density data center designs. Implement 100 GbE with Intel Silicon Photonics optics where appropriate and cost-effective, to meet network capacity demands.
- **Improve operational efficiency.** Increase the telemetry within the data center to improve the operational efficiency.

"Our data center transformation strategy is critical for Intel IT to stay competitive."

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 2.8 billion from 2010 to 2018. Our data center transformation strategy is critical 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 minimum of 44 percent cost savings compared to a full-acquisition refresh. Along with this TCO reduction, the disaggregated server innovation enables significant TCE reduction (82 percent of material weight in a new server is removed with just a CPU-complex upgrade).
- One-day dock-to-production for new physical server deployment in our data center hub.
- We developed a system software capability called NUMA-Booster, which has saved millions while delivering additional usable server capacity.
- We deployed more than 40 PB of Intel SSDs as fast local data cache drives, which increased workload performance due to lower network traffic and storage demand.
- Five generations of HPC in our design computing environment created a 236x capacity increase and an 80x quality improvement.
- We adopted new storage capabilities like deduplication and compression, accelerated storage refresh, focused on increasing utilization, removed unneeded data, and implemented policy-based tiering—all of which have resulted in getting additional usable capacity out of storage.
- We deployed more than 128,756 10 GbE network ports, 11,755 40 GbE network ports, and 1,619 100 GbE network ports.

INTEL IT MOR TACTICS

Embrace Disruptive Servers
Adopt Tiered Storage
Drive Network Efficiency
Increase Facility Efficiency
Improve Operational Efficiency

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 Peer Network

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

Acronyms

DOME	Design, Office, Manufacturing, and Enterprise
EDA-MIPS	electronic design automation MIPS
HPC	high-performance computing
KPI	key performance indicator
MIPS	meaningful indicator of performance per system
MOR	model of record
NIC	network interface card
NUMA	non-uniform memory access
PUE	power usage effectiveness
QoS	quality of service
ROI	return on investment
SLA	service-level agreement
TCE	total cost to environment
TCO	total cost of ownership
VM	virtual machine