GREEN COMPUTING: TOP TEN BEST PRACTICES FOR A GREEN DATA CENTER

HOW TO BUILD AND OPERATE AN ENERGY-EFFICIENT DATA CENTER AND REDUCE OPEX

PLANNING AND IMPLEMENTING AN ENERGY-EFFICIENT DATA CENTER IS EASILY ACCOMPLISHED AT SCALE

SUPERMICRO

SUMMARY OF DATA CENTERS IMPACT ON THE ENVIRONMENT

- Up to 400 TWh of Electricity Used to Power Data Centers Worldwide*
- Estimated 2+ Million Tons of E-Waste Generated By Data Centers Per Year**
- Performance/Watt Increasing Faster than CPUs and GPUs Power Requirements
- In Some Regions, 80% of Electricity Generated by Fossil Fuels***
- Up to 8 Billion Trees Will Not Have to be Planted for Carbon Offset if Data Centers Worldwide Used More Efficient Servers****

* https://www.akcp.com/blog/the-real-amount-of-energy-a-data-center-use/#:~:text=In%202020%2C%20the%20data%20center,annual%20data%20center%20energy%20consumption

** https://www.grcooling.com/blog/3-ways-to-manage-e-waste-from-data-centers/

*** https://www.eesi.org/topics/fossil-fuels/description#:~:text=Fossil%20fuels%E2%80%94including%20coal%2C%20oil,were%20compressed%20and%20heated%20underground

**** https://s25.q4cdn.com/632471818/files/doc_financials/2022/q4/Earnings-Deck-Q4FY22-V6.pdf

TEN ACTIONS TO REDUCE THE ENVIRONMENT IMPACT OF YOUR DATA CENTER

The following is a top 10 list of most impactful actions to be taken to reduce the environmental impact of data centers. Data centers adopting these suggestions are likely to experience lower OPEX and be able to provide measurable improvements in their carbon footprint.

As a quick summary, the actions that can be implemented are as follows:

1. **Right Size System Designs to Match Your Workload Requirements**
2. **Share Common Scalable Infrastructure (Multi-node, Blade Efficiency)**
3. **Operate at Higher Ambient Temperature (Thermostat, Free Air)**
4. **Capture Heat at the Source (Hot or Cold Aisle Containment, Liquid Cooling)**
5. **Select & Optimize Key Components for Workload Perf/Watt (CPU, GPU, SSD, ...)**
6. **Optimize Refresh Cycles at Component Level for Perf/Watt**
   (Disaggregated, Universal GPU, Long Life Chassis/PS)
7. **Optimize Power Delivery**
8. **Utilize System Consolidation, Virtualization and Power Management**
9. **Source Green (Renewable Energy, Green Manufacturing)**
10. **Rethink Site Selection Criteria with Climate – Location, Location, Location**

EXECUTIVE SUMMARY

Globally, data centers use at least two hundred Terawatts (TW) of electricity per year (about 2% of global energy use[1,2]), and growth models show it will continue to grow to 4% to 8% by 2030 depending on sources. With an expectation to grow between 2% and 8% by 2030. Reducing electricity usage in the data center results in less environmental impact and lower costs through OPEX reductions. With the expectation that climate change will continue to worsen, now is the time to take action on reducing the Total Cost to the Environment of data centers. Whether a hyperscale data center, co-location data center or a data center physically located within an enterprise, there are a number of opportunities to reduce the effect on the environment and increase sustainability. According to one study[3], there are over 700 hyperscale data centers worldwide at the end of 2021, with an estimate of about 1200 by 2026[4]. While the demand for digital services has been increasing at a much higher rate than the energy consumption due to increased server efficiencies, data center operators and designers need to implement practices that reduce energy demand to help governments and businesses reach environmental goals. From the location of a data center to the choice of microprocessors, there are numerous places where data center designers, operators, and end users can help reduce the data center's environmental impact.

In a recent survey with data center operators and decision-makers, worldwide, over 77% of the respondents stated that the environmental impacts of their data centers were very important to their organization (See Figure 2). This confirms earlier discussions in past years that a primary concern among these operators continues to be the enormous amounts of electricity that is needed for the operation of data centers, and needs to be reduced.

Worldwide, in some regions, about 80%[5] of electricity is currently produced by burning fossil fuels (coal, gas, oil), although this varies by geography. Data centers still use substantial amounts of grid power that burn fossil fuels to fulfill energy needs, even though some data centers—and corporations that use them—are transitioning to renewable sources or have targets to reduce fossil fuel consumption. While the world is transitioning to renewable energy, a significant amount of electricity will continue to be generated by fossil fuels.

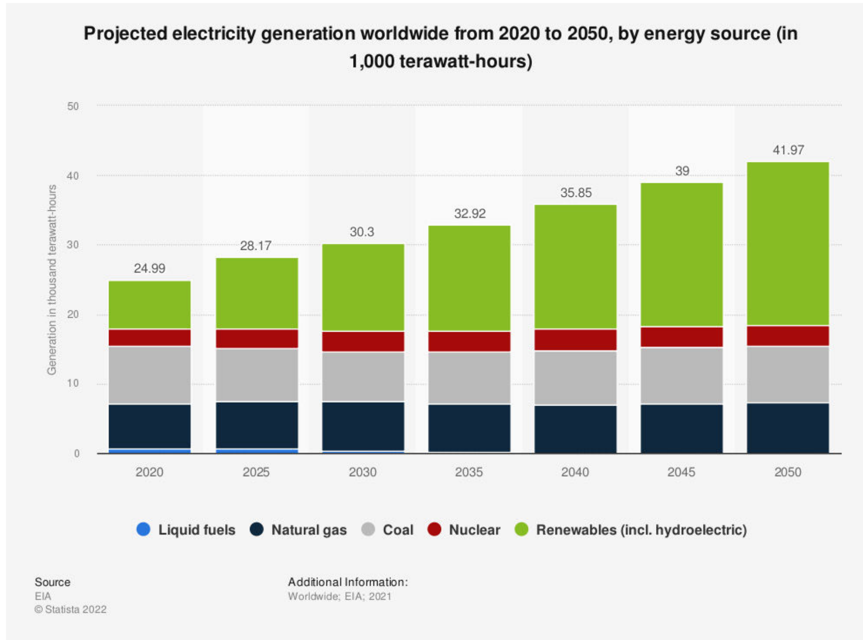


Figure 1 – Projected Electricity Generated Worldwide by Energy Source

To "green" a data center, a number of actions can be taken, even once a data center is up and running. These actions range from purchasing green (renewable) energy to planning ahead and purchasing the most efficient servers available for the SLAs of workloads required. This paper will explore 10 Best Practices that can be implemented and have already shown to be effective in reducing power usage and improving the Power Usage Effectiveness (PUE) of a data center.

Several considerations will reduce environmental impacts when creating or refreshing a data center. Worldwide, 93% of the survey* respondents stated that their data center's environmental impacts were either very important or of secondary importance.

From the survey, no region was below 76% for "Yes, very important" when looking at the importance of environmental impacts of data centers from different geographies.

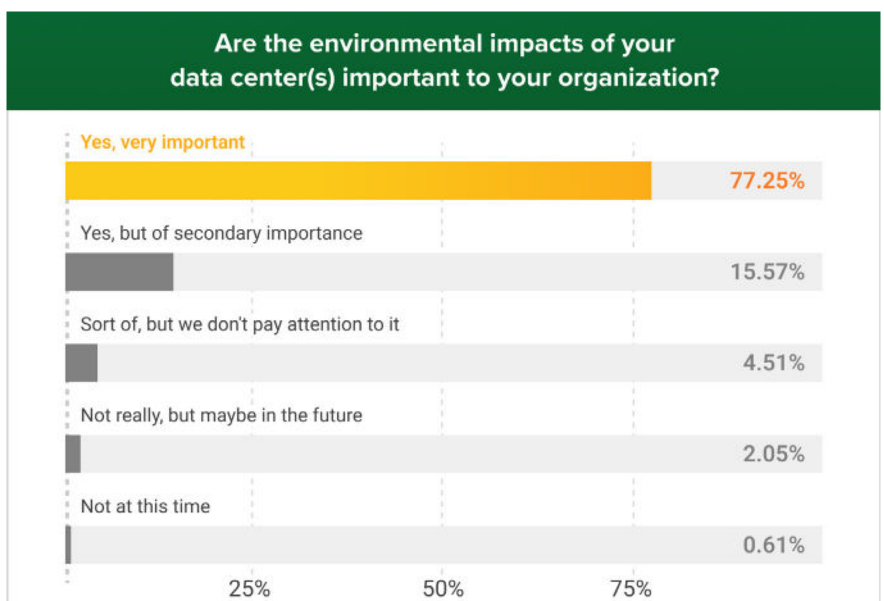


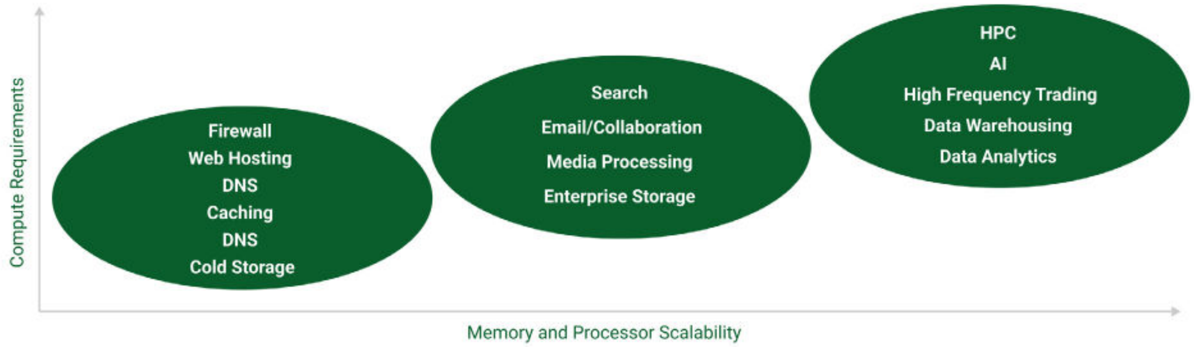
Figure 2 –
Environmental Impacts Importance

* To understand the thoughts and actions of data center managers worldwide, Supermicro commissioned a survey in the summer of 2022. Some of the results of the study are mentioned in this document. See the details of the survey demographics in Appendix A.

BEST PRACTICE #1: Rightsized Systems to Match Your Workload Requirements

There are many component choices and configuration options on a server. Traditional general-purpose servers are designed to work for any typical workload, which leads to over-provisioning resources to ensure the system works for the widest range of applications. A workload-optimized system, instead, optimizes component choices and configuration options to exactly match the requirements for a target set of workloads. These optimizations reduce unnecessary functionality, which reduces cost, but also reduces power consumption and heat generation. When the optimized solution is scaled to 100s, 1000s, or 100,000s of systems, the savings are significant.

Different product lines are optimized for different workloads, for example, servers with more CPUs, cooling capacity, memory capacity, I/O capacity, or networking performance. HPC applications require fast CPUs, while content delivery networks need massive I/O capabilities Using the server type that is designed for the workload reduces the excess and unused capacity and, thus, a cost reduction.



BEST PRACTICE #2: Share Common Server Infrastructure

Systems can be designed in a way to share resources, which can lead to better overall efficiency. For example, sharing power supplies or fans among several nodes reduce the need to duplicate these components for each node. For instance, the Twin family of servers (TwinPro, BigTwin, FatTwin, and GrandTwin) all share power supplies and fans. The result is that larger fans and more efficient power supplies can be used, reducing the electricity use when all of the nodes are running applications.

Another example of saving resources on a Supermicro system is the Supermicro Superblade®, an innovative system that allows multiple blades to share networking, fans, and power supplies. With this system, significant densities per rack can be installed, reducing the square footage needed in a data center. In addition, from the recent survey, almost 95% of the respondents look for servers that reduce electricity usage.

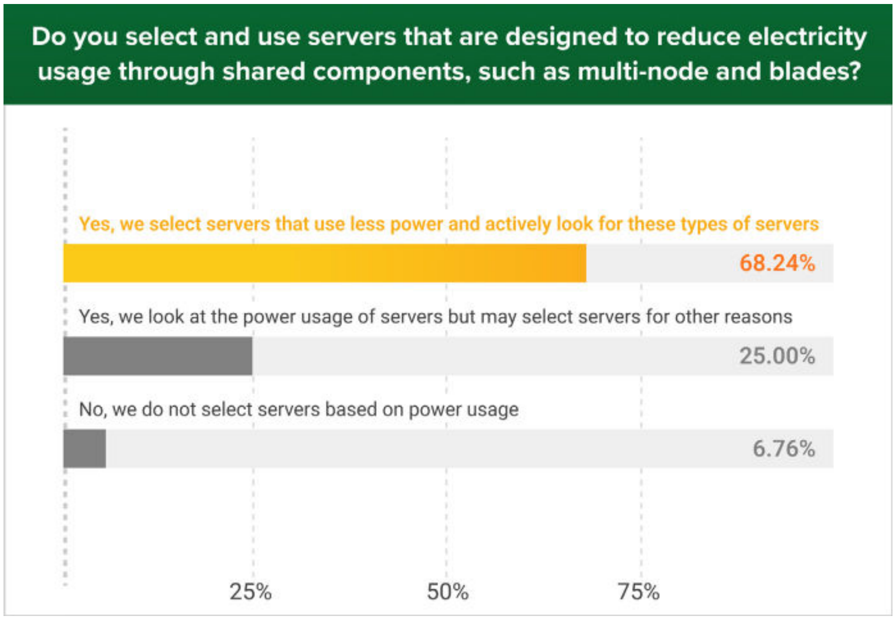


Figure 3 – Selection Of Servers with Shared Components

MULTINODE AND BLADE SERVERS

Supermicro BigTwin® (2U 4 node server)

High-density, multinode servers deliver massive computing power in minimal space for workloads such as High Performance Computing (HPC), Artificial Intelligence (AI), cloud, grid, and analytics, while reducing costs, energy, and space requirements.

Multinode servers provide significant power and space savings compared to standard industry rackmount servers. These servers integrate two or more independent nodes into a single enclosure, thus increasing density.

The main advantage for green data centers is saving power due to shared components including power supplies, fans, enclosures, and cabling.

- **Blade Servers:** optimized to use less space and energy, minimizing power consumption.
- **Multinode Servers:** designed with two or more independent server nodes in a single enclosure; ideal for enclosures with limited space.
- **Hyperconverged Infrastructure (HCI) Servers:** combines storage, compute, and networking into a single system. This decreases data center complexity and increases stability.

Another method that decreases energy usage when air cooling a server is to be aware of and reduce cabling issues. Power and network cables that block airflow require the fans to operate at a higher RPM, using more electricity. Careful placement of these cables within the chassis and external to the chassis reduces this possible issue. In addition, a server that consists of blades with integrated switching will typically have fewer cables connecting systems, as this is done through a backplane.

Comparing 20 1U rackmount servers to a Supermicro SuperBlade 8U chassis with 20 blades. There is a 95% reduction in the number of cables, which can contribute to improved airflow and less electricity used by the system in terms of fan speed.

Type	1U Rackx20	8U20 blades
Data (25 GbE)	2x20	None/ Internal
Data (10GbE)	2x20	
Power	2x20	6
IPMI/Mgmt	1x20	1
Total	140	7
%Reduction	95.0%	

Figure 4 – Cable Reduction with SuperBlades vs. Rackmount Servers

BEST PRACTICE #3: Operate at Higher Temperatures

When using traditional air cooling mechanisms, the air entering the server (inlet temperature) is maintained by Computer Room Air Conditioning (CRAC).The amount of air conditioning used in a data center contributes the most to the PUE calculation. Reducing the amount of air conditioning significantly lowers the PUE and, thus, OPEX costs. Around the world, many data centers are keeping inlet temperatures too low. Data center operators can reduce power usage by increasing the inlet temperatures to the manufacturers recommended maximum value. Looking at the results from the survey, there is a wide range of inlet temperatures, which also shows that most IT administrators are limiting the inlet temperature to less than the manufacturer sets as the high limit. For example, the Supermicro BigTwin® engineering specifications (https://www.supermicro.com/en/products/system/bigtwin/2u/sys-220bt-hntr) allow the operating temperature to be as high as 95° F (35° C), and only about 15% of the responses show that the inlet temperatures are above 84° F (29° C).

“Lower OPEX with a lower PUE”

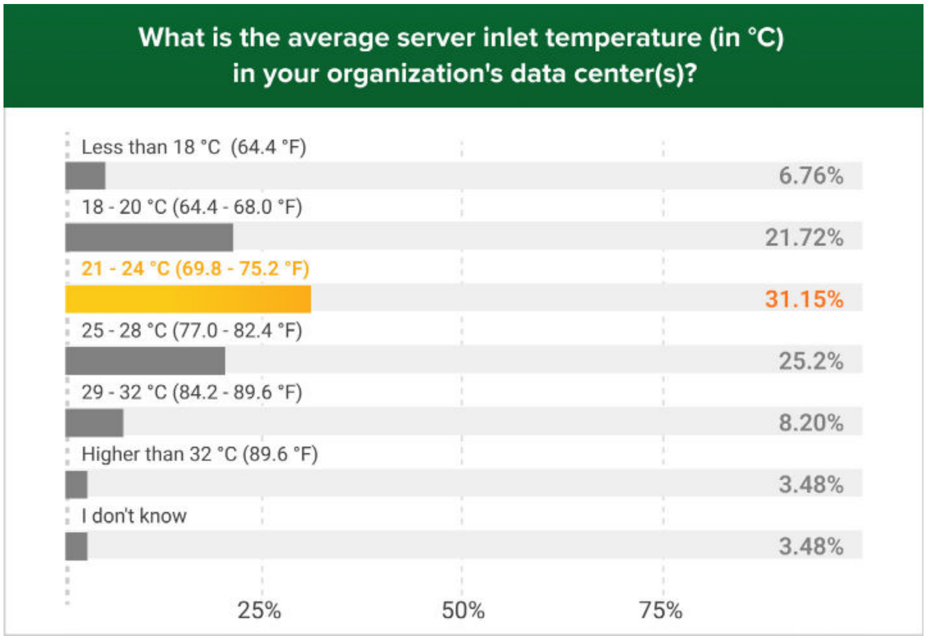


Figure 5 – Average Server Inlet Temperatures

Cooling with "free air," can be defined as using external air and fans which is filtered and the humidity adjusted, can be a significant contributor to a green data center by lowering the need and use of a computer room air conditioner (CRAC). The use of outside air is only possible in certain climates and geographies and may be part of a decision-making process regarding where a new data center is

“Free air cooling reduces the need for CRAC”

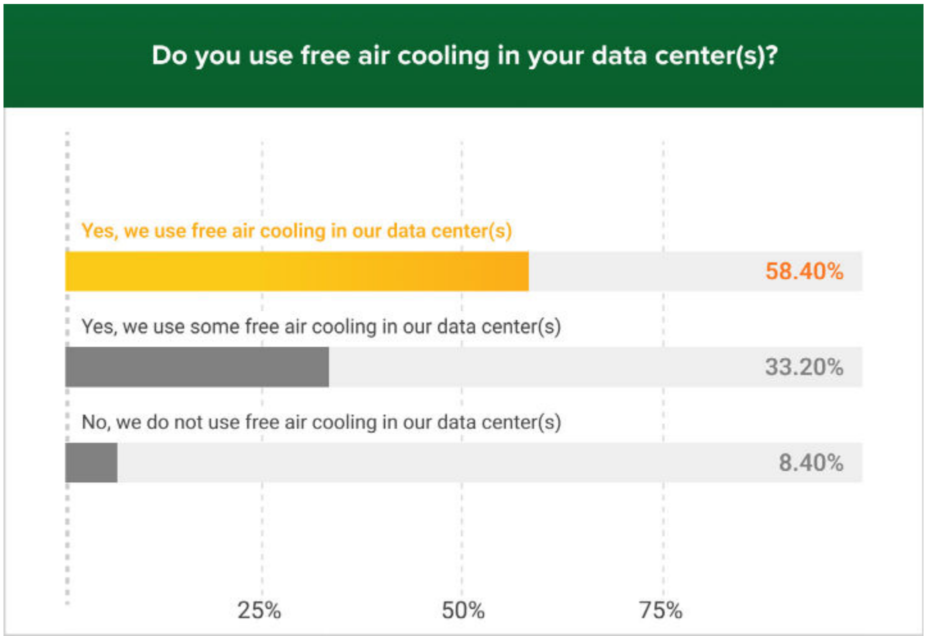
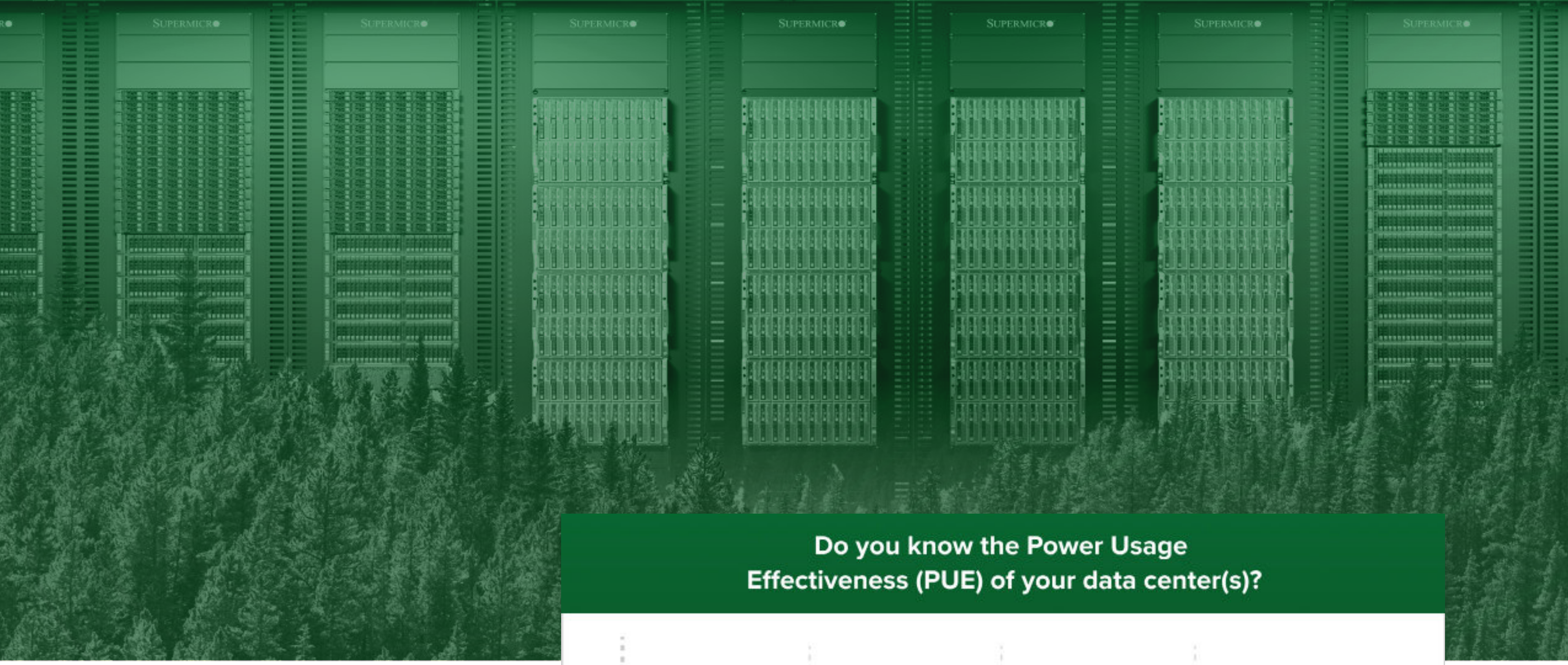


Figure 6 – Free Air Cooling Use



to be located. Careful consideration should be given to understanding the yearly climate norms and extremes.

In the recent survey, over 90% of the respondents stated that they use some amount of free air cooling.

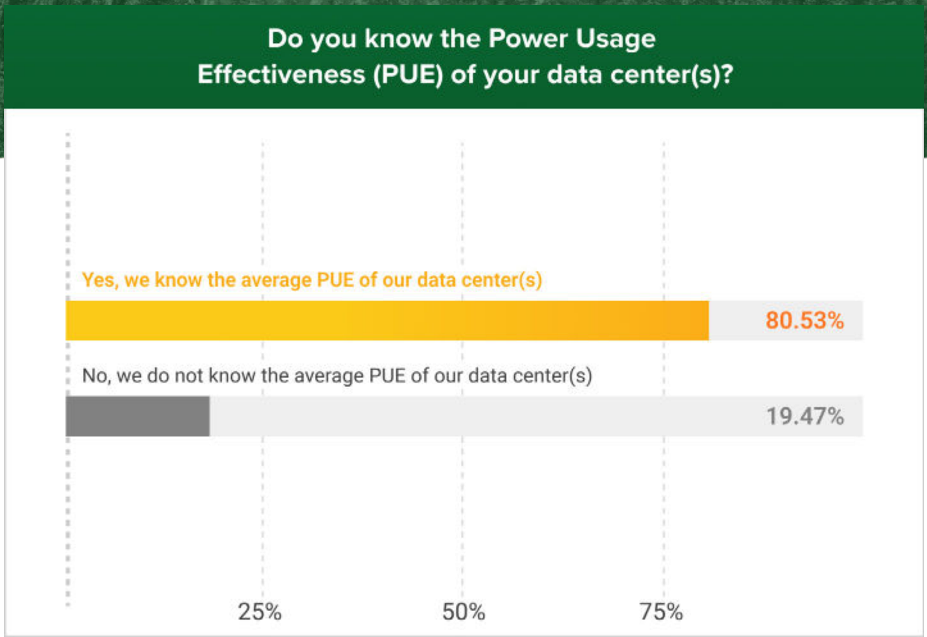


Figure 7 – Knowledge of PUE In Respondent's Data Center

BEST PRACTICE #4: Capture Heat At The Source

Computer room air conditioning is the most significant variable to optimize in order to lower overall PUE. The PUE of a data center is defined as the total amount of power delivered to the data center divided by the amount of power used by the IT components. The lower the value, the more energy efficient the data center is. In the recent survey, about 80% of the respondents knew the PUE of their data centers, and the most frequent PUEs were in the 1.11 to 1.40 range.

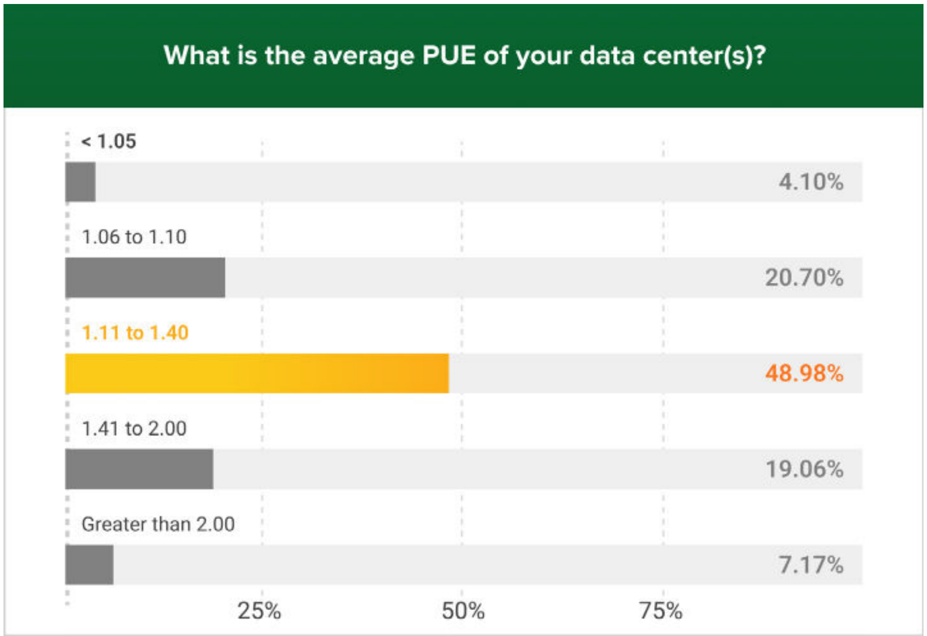


Figure 8 – Average PUE In Respondents Data Center

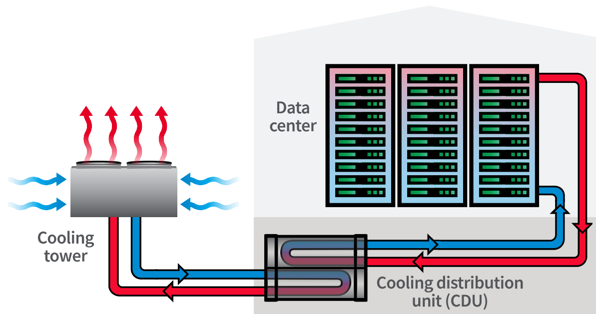
Liquid Cooling

Liquid cooling of the CPUs and GPUs can significantly reduce the need for having CRAC units in data centers and the need to push air around. There are several different methods to use liquid cooling to reduce the need for forced air cooling and the potential to reduce the time to completion of a workload. Using liquid cooling can significantly lower the PUE of the data center.

"Liquid Cooling Dramatically Improves the PUE of a Data Center"

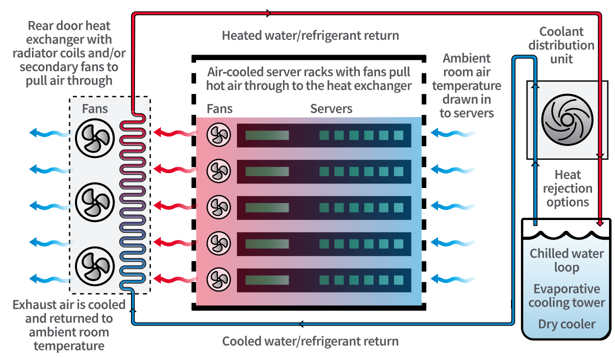
Direct To Chip (DTC or D2C) Cooling

In this method, a cold liquid is passed over the hot CPU or GPU. Since liquid is much more efficient at removing and transporting heat than air is, the CPU or GPU can be kept within its thermal design power (TDP) envelope. As CPUs and GPUs are drawing more power, each creates more heat that must be removed. With more heat in an air cooled data center, the CRAC units must work harder to cool the hot air before it is returned to the inlet face of the server. When D2C liquid cooling is implemented, a significant reduction in CRAC and a reduction in the fan speed are needed to keep the system cool. The hot liquid is then sent to an external cooling apparatus, which may be external to the data center. A recent experiment showed that when two Supermicro systems were compared, one using air cooling, and one using liquid cooling (D2C), the liquid cooled system used about 10% less power (due to slower fan speeds required) than the air cooled system. This difference can lead to significant savings when looking at thousands of systems in a medium to a large data center.



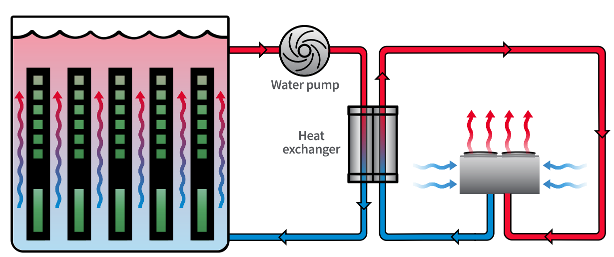
Rear Door Heat Exchanger (RDHx)

The rear door of the rack contains liquid and fans which cools the hot server exhaust air before the air enters the data center. The hot liquid needs to be cooled before it is returned to the data center CRAC. This method of using liquid keeps the air at a lower temperature in the data center, allow the reduction of cooling demands on the CRAC, which will lessen the data center electricity needed.



Immersion Cooling

The entire server or groups of servers are immersed in a dielectric liquid. The close contact of the liquid molecules with the hot CPUs, GPUs, and other components are an efficient way to cool the servers, as fans will be removed from the servers. There are some minor modifications that must be done to the server before immersion. This type of cooling is considered a closed loop system, where the cooler liquid is pumped into the immersion tank. As the hot electronics heat the liquid, the liquid rises and is then pumped out of the tub and cooled elsewhere. An entire rack of servers can be cooled in this manner.



Hot and Cold Aisles

A significant amount of electricity can be saved in the usage of the Computer Room Air Conditioning (CRAC) if the hot and cold aisles are separated in the data center. When designed with hot and cold aisles, the inlet and exhaust air should not mix, allowing the data center cooling to operate more efficiently. For adequate cooling, the rows of racks need to be installed so that the rear of the racks face each other, creating a hot aisle. Therefore, one of the more important best practices when designing an energy efficient data center is to have hot and cold aisles. The benefit is that the warm/hot air is kept away and separated from the cooler air.

Traditional methods to remove heat from a server include moving cooler air over the electronic components, where hot air rises and is expelled from the chassis. Typically, the airflow is from front to back (as inserted in a rack), with fans pulling the air through the server located at the server's rear. With multiple servers in a rack, the hot air from all servers creates a very warm/hot zone behind the rack. As multiple racks are lined up side-by-side, the total amount of hot air expelled from the back of the servers increases, resulting in an aisle full of hot air that should be contained.

- Containment

  Creating a cold aisle and hot aisle will keep the cold aisles cold and allow just the hot air to be returned to the CRAC.

- Systems to remove the hot air to be cooled

  Below are examples of an isolated hot aisle where the hot air is returned to the CRAC unit and a cold aisle containment where the cooler air is delivered beneath the floor as intake to the servers. The hot air is then chilled with computer room air conditioners or other methods.

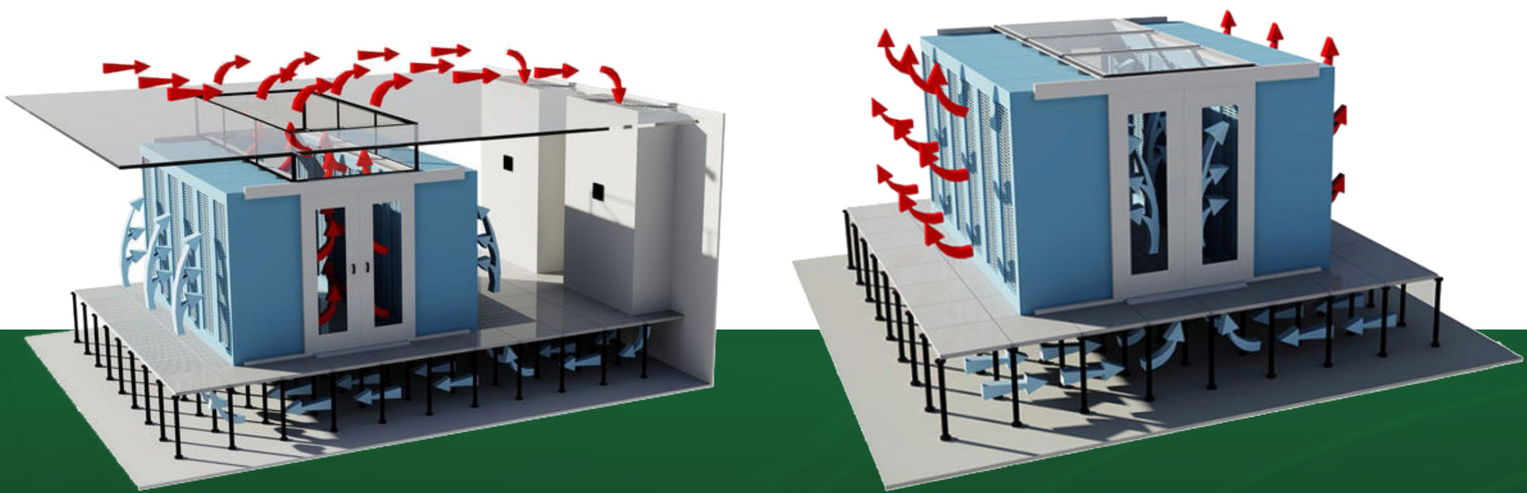


Figure 9 – Hot Aisle and Cold Aisle Containment Options
Image Source – https://www.colocationamerica.com/blog/hot-vs-cold-aisle-containment

Designing a data center with a hot aisle or cold aisles depends on the initial investment. Cold aisle containment is generally less expensive because you may only require doors and a roof for cold aisle containment, making it more affordable. Additionally, it's a more straightforward setup, and expansion costs are less if additional growth is needed.

Many hot aisle setups are raised floor configurations, which are more efficient with the advantage of reducing costly use of CRAC units. The disadvantage is the initial cost of building a data center from the ground up, raised floors, and ductwork. Still, the long-term cost-benefit is more significant than a cold aisle configuration if optimized by utilizing best practices.

Additional Benefits to Heat Removal

Supermicro offers a wide range of products that enable front-accessible-service design (FASD) for cold-aisle serviceability. This design allows for Field Engineers (FE) to service from the cold aisle, which eliminates the need to enter the hot aisle where temperatures can reach 120° F. The FASD also allows the FE to reduce service time because they don't have to deal with cabling, power cords, liquid cooling tubes, power cords etc.

Figure 10 – GleSYS Facilities in Vastberga, Stockholm, Sweden

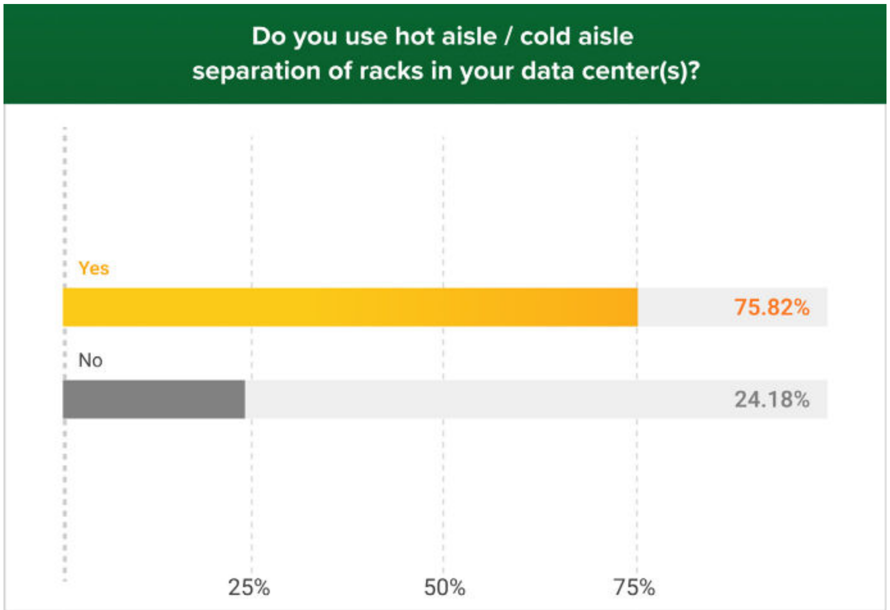


Figure 11 – Use of Hot and Cold Aisles in Respondents Data Centers

Thermal energy benefits can extend beyond the server room. Thermoelectric devices can capture server heat and turn it into electricity for other operations. For example, an innovative approach is currently operating at GLeSYS[6] in the town of Vastberga, Stockholm, Sweden, where the hot air is pumped to a nearby heating plant, reducing the nearby towns heating requirements. Thermal capture of night-time cooling can also benefit data centers. Thermal Energy Storage (TES) allows this method to capture cooling in various mediums, stored and used during the day to meet air conditioning load demand or process cooling loads. Maximizing thermal energy benefits can create a PUE approaching 1.0 when the server's heat goes back into the facility or for reuse.

BEST PRACTICE #5: Select & Optimize Key Components for Workload Perf/Watt

Performance/Watt

As CPU technology constantly improves, one of the most critical gains is that more work per watt is accomplished with each generation of CPUs and GPUs. The most recent offerings by Intel and AMD are up to three times more performant in terms of the work produced per watt consumed. This technological marvel has enormous benefits for data centers that wish to offer more services with constant or reduced power requirements.

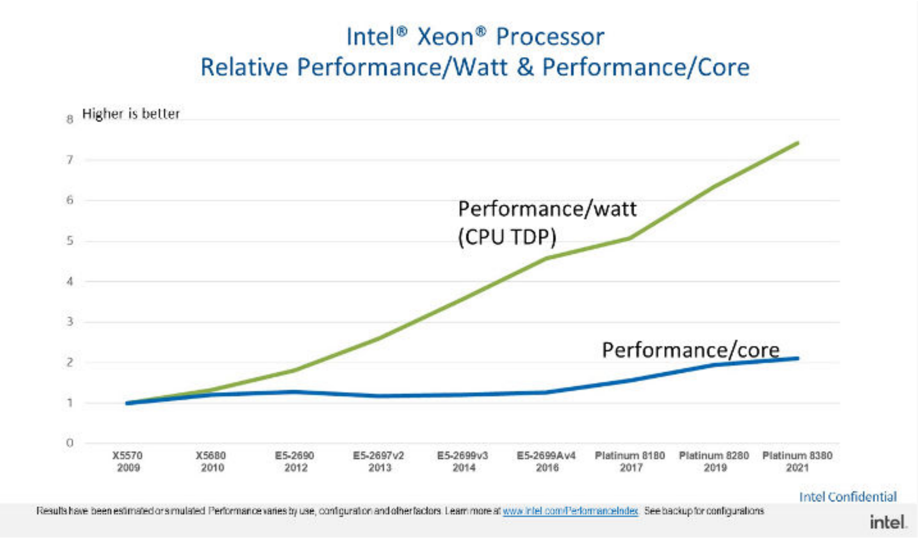


Figure 12 – Performance per Watt Over Time (Normalized)

CPU Differences

CPUs for servers and workstations are available in many different configurations. CPUs are generally categorized by the number of cores per CPU, the processor clock rate, the power draw, the boost clock speed, and the amount of cache. The number of cores and clock rate are generally related to the amount of electricity used. Higher numbers of cores and clock rates will usually require more electricity delivered and will run hotter. Conversely, the lower number of cores and clock rates will use less power and run cooler.

For example, suppose workload A is not required to be completed in a defined amount of time. In that case, a server with lower powered CPUs (generally related to performance) can be used compared to a higher powered system when the SLA may be more stringent. An email server is an example. The response time to view or download emails to a client device needs to be interactive, but a slower and less power demanding CPU could be used since the bottlenecks would be storage and networking. On the other hand, a higher performing and a higher energy CPU would be appropriate for a database environment where data may need to be analyzed quickly. While putting an email server onto a high performing system would not cause harm, the system will not be used for its intended purpose.

Accelerators

Today's computing environments are becoming more heterogeneous. Accelerators are available to increase the performance of specific tasks, even while CPU performance has increased exponentially over the past few years.

Type	CPU	GPU
Cores	10-100 per socket	Thousands
Context Switching	Excellent	Limited
Parallel Operations	Limited	Excellent
Flexibility	Excellent	Limited to Parallel Tasks

The most popular and visible type of accelerators are GPUs, which can be used for massively parallel tasks. New GPUs contain thousands of "cores" compared to tens to the low hundreds in CPUs. With HPC and AI applications, GPUs deliver tremendous performance increases but come with an increased electricity requirement. CPUs (late 2022) are topping out at 350W, but GPUs are up to 700W. However, since the performance for HPC and AI applications is significantly improved when the application has been designed to take advantage of the massive parallelism of GPUs, the time to run the application will be decreased as well as electricity. For example:

Type	Dual Socket CPU after Only Server	Dual Socket CPU + GPU Server (8 HGX GPU)
Watts (for Entire System)	1,000	4,000
Execution Time	1 hour	.2 hour
Total Kilowatt-Hour	1 KWh	.8 KWh

Thus, the combined CPU + GPU system uses 40% less power for a given task than a CPU only system. In the recent survey, almost 80% of the respondents use GPUs.

SSD vs. HDD

Hard Disk Drives (HDD) have been the primary storage method for over 50 years. While the capacity of HDDs in recent years has increased dramatically, the access time has remained relatively constant. Throughput has increased over time, as has the capacities of HDDs. However, Solid State Drives (SDD) are faster for data retrieval and use less power than HDDs, although HDDs are suitable for longer term storage within an enterprise. M.2 NVMe drives, SSD drives currently transfer about 3GB/sec, which can significantly reduce the time required to complete a heavily I/O application. Overall, this performance will result in lower energy consumption for the time to complete a task compared to other I/O technologies.

	Hard Disk Drive	Solid State Drive
Performance	*	***
Latency	*	***
Capacity	***	**
Lifespan	***	**
Cost/(GB,TB)	***	*
Shock Resistant	*	***
Use Cases	Long term storage	OS, Frequently Accessed Files

*** Advantage
**
* Disadvantage

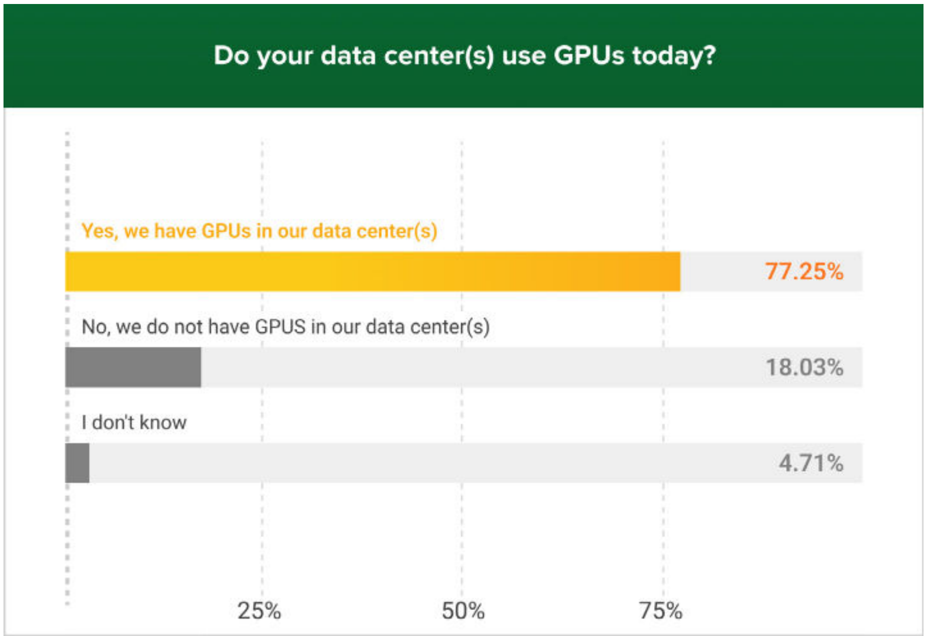


Figure 13 – Data Center GPU Use

BEST PRACTICE #6:

Optimize Refresh Cycles at Component Level for Perf/Watt(Disaggregated Servers)

The major components of servers are continually improving in terms of price and performance. As applications continue to use more data, whether for AI training, increased resolution, or more I/O (as with content delivery networks), the latest servers that contain the most advanced CPUs, memory, accelerators, and networking may be required. However, each of the sub-systems evolves at a different rate. With refresh cycles decreased from five and three years, according to some estimates, entire servers do not need to be discarded, contributing to E-waste. With a disaggregated approach, the components or sub-systems of a server can be replaced when newer technology is deployed. A well designed chassis will be able to accommodate a number of electronic component technology cycles, which allows for component replacement. By designing a chassis for future increases in the power required for CPUs or GPUs, the chassis will not have to be discarded as new CPUs are made available.

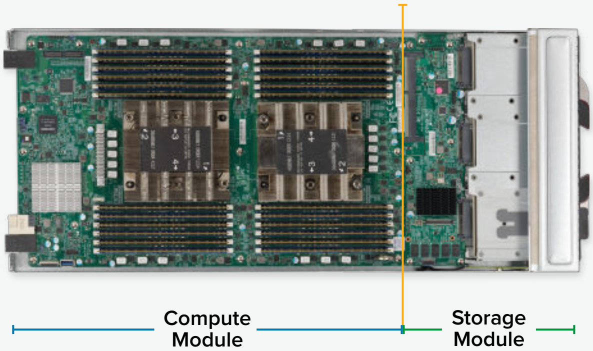
"Disaggregated Architectures can reduce need for full system replacement during refresh and help reduce E-Waste"

For example, in a recent white paper, Intel discusses how disaggregated servers allow Intel to reduce capital expenditures while upgrading certain technology as needed. (https://www.intel.com/content/dam/www/central-libraries/us/en/documents/intel-it-green-computing-at-scale-paper.pdf).

DISAGGREGATED SERVERS

The 6U SuperBlade utilizes a disaggregated architecture that enables the independent upgrades of system components. Each blade is composed of a storage module and a separable compute module with CPUs from Intel®, for example, and memory that can be refreshed at faster rates than the rest of the system.

Supermicro 6U Disaggregated SuperBlade®



An example of a system that is designed for a choice of CPU and GPU technology, is the Supermicro Universal GPU server (https://www.supermicro.com/en/products/universal-gpu). The Supermicro Universal GPU server is designed to house either AMD or Intel CPUs and a variety of GPUs, both in terms of type (PCI-E or HGX), and the number of GPUs. The Supermicro's Universal GPU System is the industry's most advanced and flexible GPU server platform in the market today. Designed for large-scale AI deep learning and HPC workloads, this modular, standards-based platform supports the best available GPU technologies in a variety of form factors and combinations both today and into the future .

BEST PRACTICE #7: Optimize Power Delivery

Power conversion from AC to DC entails some amount of heat generated. With AC being delivered to the data center, the power must be converted to DC for the system. With each conversion of AC to DC, power is lost, contributing to the inefficiency of the data center. More efficient conversion will result is less wasted power during the conversion, with heat being the by-product which must be removed from the system.

Titanium power supplies are the most efficient option, offering 96% power efficiency. Platinum power supplies are slightly less efficient at 94%. Gold power supplies offer a lower efficiency of 92%. The efficiency of a power supply isn't linear or flat when it comes to the supply's output range. Most power supplies operate at their maximum efficiency when they're running in the upper ranges of their rate capacity. This means that an 800-watt power supply providing 400 watts of power (50% capacity) will be less efficient than a 500-watt power supply providing that same 400 watts of output power (80% capacity).

Additionally, the power distribution for a rack of servers is best accomplished by using multi-node and blade systems. The higher the AC input voltage, the more efficient the entire power conversion process. Multi-node and blade systems share the AC power supplies among a number of independent servers. This results in a more efficient process in the AC to DC power conversion process.

One method to reduce the PUE of a data center is to optimize the different power conversion steps. In our survey, 66% of the respondents said they optimized the power conversion steps.

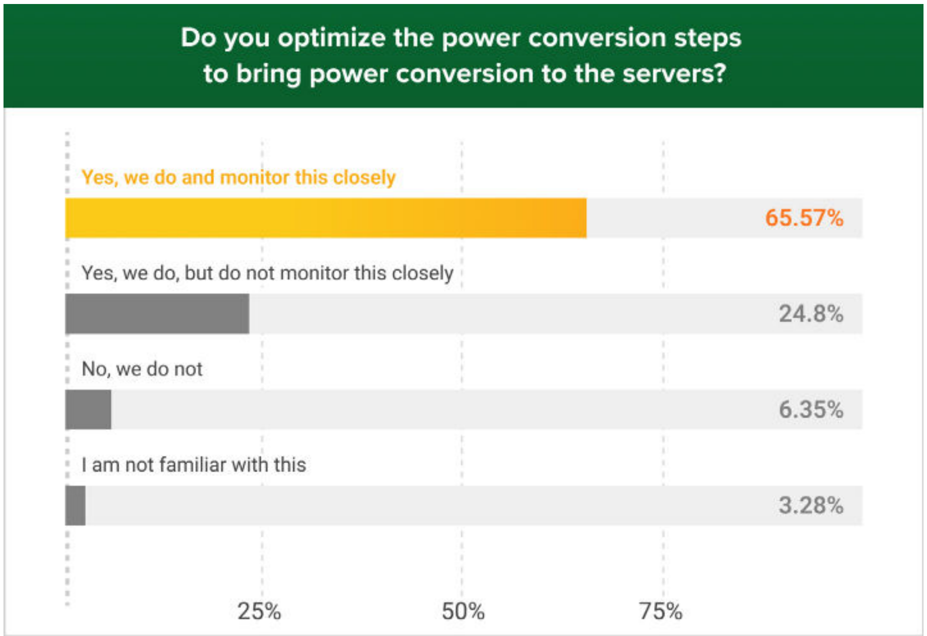


Figure 14 – Optimizing Power Conversion Steps

Ultra–1U: 84 CPUs, 252 TB, 504 NVMe drives per rack

BigTwin®–2U: 168 CPUs, 504 TB, 504 NVMe drives per rack

SuperBlade®–8U: 200 CPUs, 600 TB, 400 NVMe drives per rack

BEST PRACTICE #8: Utilize System Consolidation, Virtualization, and Power Management

System consolidation, Virtualization and power management tools are all great examples of ways to improve power utilization and increase flexibility.

IT administrators who can control power utilization can significantly affect the overall power consumed in a data center. Empowering the administrators to monitor and then regulate this critical aspect of the entire data center can lead to a more efficient operation and less expense. For example, by analyzing logs of power usage over time, IT administrators can move specific applications (jobs) to use the systems when more renewable power is available or when power costs are lower (time of day pricing). Capping the power a server or a cluster of servers can use through the Supermicro SuperCloud Composer can reduce costs by limiting how much energy a server can use. Less power delivered to the server will result in lower performance, but it can still meet SLAs for specific workloads. Virtualization and container management systems are also critical for reducing power usage in the data center. Virtual servers allow for a higher utilization of the CPU and memory resources, which can decrease the number of servers in the data center.

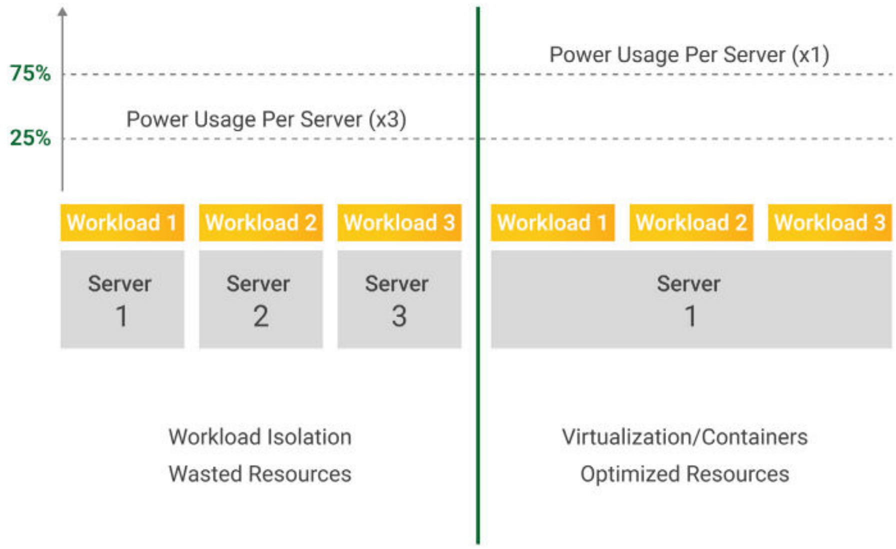


Figure 15 – Comparison of Non-Virtualized and Virtualized Environments

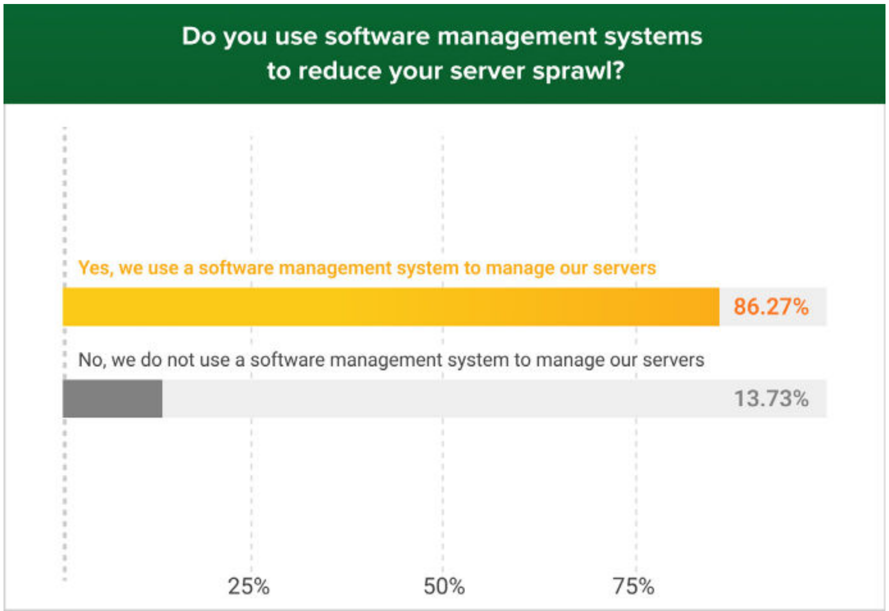


Figure 16 – Use of software management systems

BEST PRACTICE #9: Source Green Energy

A data center's energy source has the most significant impact on its carbon footprint and poses the most substantial opportunity to benefit the environment. Typically, a data center operator decides the energy source for all facility users. In addition, many data centers, including many colocation data centers, publicize the movement toward energy generation from 100% renewable sources.

Renewable energy programs for commercial customers include generation through utility, third-party power purchase agreements (PPA), or renewable energy credits (REC). Distributed renewable energy production owned or controlled by data centers is optimal. But on-site renewable energy sources do not always satisfy data center energy demands. Fortunately, clean grid energy can augment this. There are also increasingly effective energy storage solutions for deployment on-site, coming down in cost as battery technology improves and scales.

When our survey asked how renewable energy or RECs were used, the results showed that about 86% of the data centers using renewable energy, some more than others.

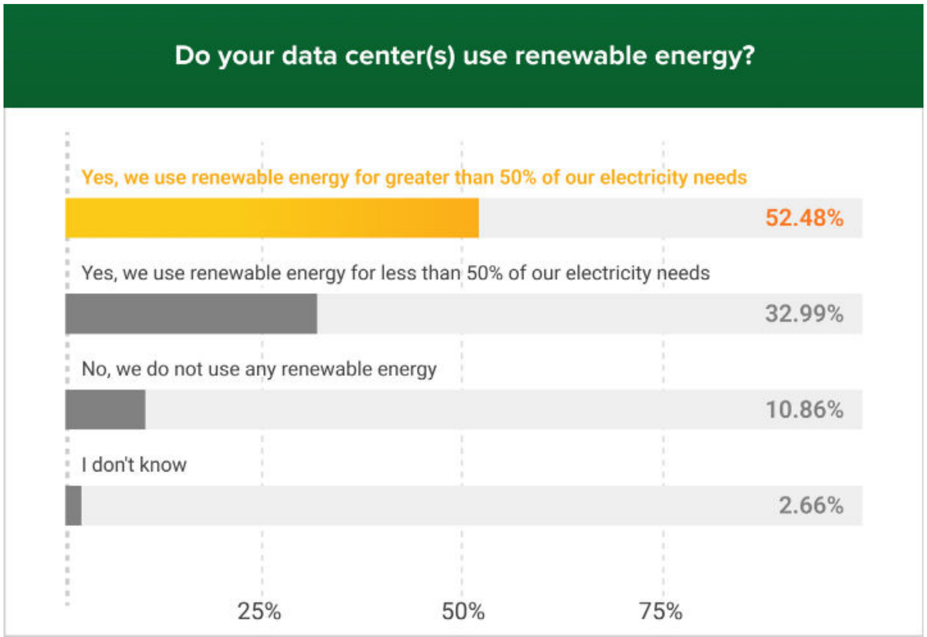


Figure 17 – Renewal Energy Use in Data Centers

In addition, about 72% of the respondents stated that they use Renewable Energy Credits or RECs.

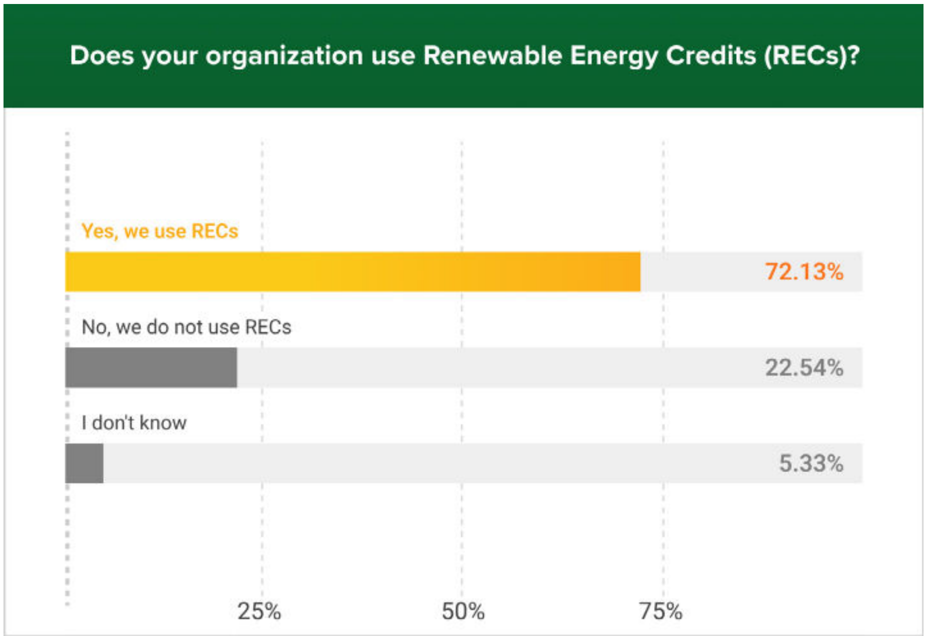


Figure 18 – Renewable Energy Credit Use

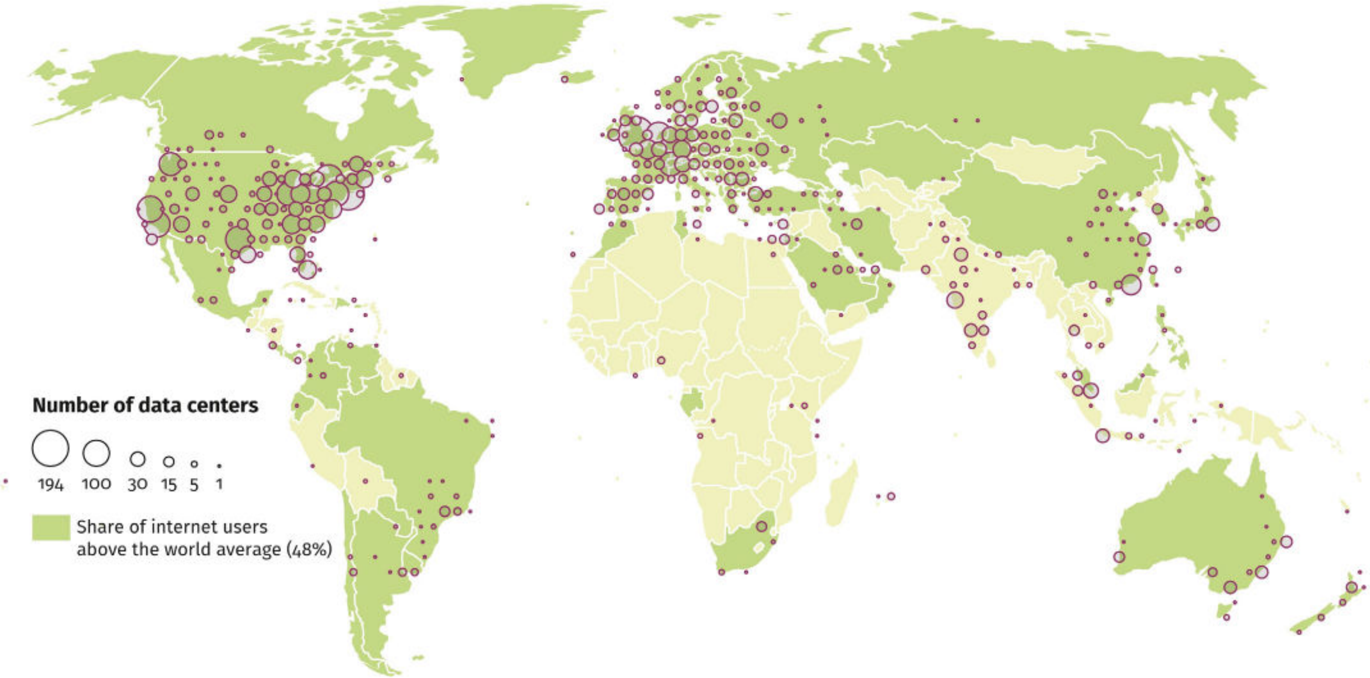
Additionally, it is important to understand and document the sourcing of components when manufacturing servers (Scope 3[7] emissions). By monitoring and using organizations that are part of the supply chain whose emissions are low, the overall contribution to the environmental impact of IT equipment will be lowered.

BEST PRACTICE #10: Rethink Site Selection Criteria

Large-scale data centers cost a lot of money to operate. For example, a single hyper-scale data center can demand 100 MW of power to keep servers, storage, and networking infrastructure performing as expected (enough to power 80,000 US households)[8]. In addition, while electronics use most of the energy consumed in a data center, cooling those electronics to maintain operating temperatures can consume 40% of facility energy. For example, there may be locations where a data center could be located to use free-air cooling for the entire year, or use 100% renewable energy, but latencies to users around the world may be impacted.

Building costs consist of the land value as well as the cost of construction. Construction prices vary depending on the region of the country and the world. Unlike building a home or an office building, a data center's location has some unique requirements to be considered "green" and deliver agreed-upon Service Level Agreements (SLAs). In addition, factors such as the climate, energy pricing, risk of natural disasters, water costs, and the cost of network bandwidth all contribute to the choice of data center locations.

Distributed digital infrastructure means an organization's IT infrastructure may have several locations, some needing greater proximity to business operations than others. Many data centers remain in hot or temperate climates. Although real estate prices may be lower in certain climates, the overall costs may be higher since free-air cooling may not be an option and CRAC utilization may need to be very high. Data centers in cooler climates can use outside air to cool the systems, reducing the need for Computer Room Air Conditioners (CRAC). Outside air can be brought into the data center directly (after passing through some filtering mechanism) to provide cool air for the cold aisles in the data center. However, poor air quality can affect the efficiency of bringing in outside air. Although this may reduce power costs, remote locations where real estate is less expensive may also be relatively distant from internet exchanges, resulting in higher latencies for customers. Connecting a data center to multiple trunk providers may be necessary. Before locating a data center in a colder climate, be aware of where energy comes from.



Sources: www.datacentermap.com ; International Telecommunications Union (ITU), www.itu.int

Some locations with outside cooler climates can provide natural cooling benefits. For example, suppose a data center is located near a moderate to cold body of water. In that case, systems can be constructed to allow the cooling system to tap into that nearby cool water. For example, naturally cooler river water can be used in a data center heat exchanger to significantly reduce the amount of power needed to cool the facility. As a result, much less power is needed to chill standard water delivery.

In our recent survey, 84% of the respondents stated that they do have the ability to locate their data centers where electricity prices are lower.

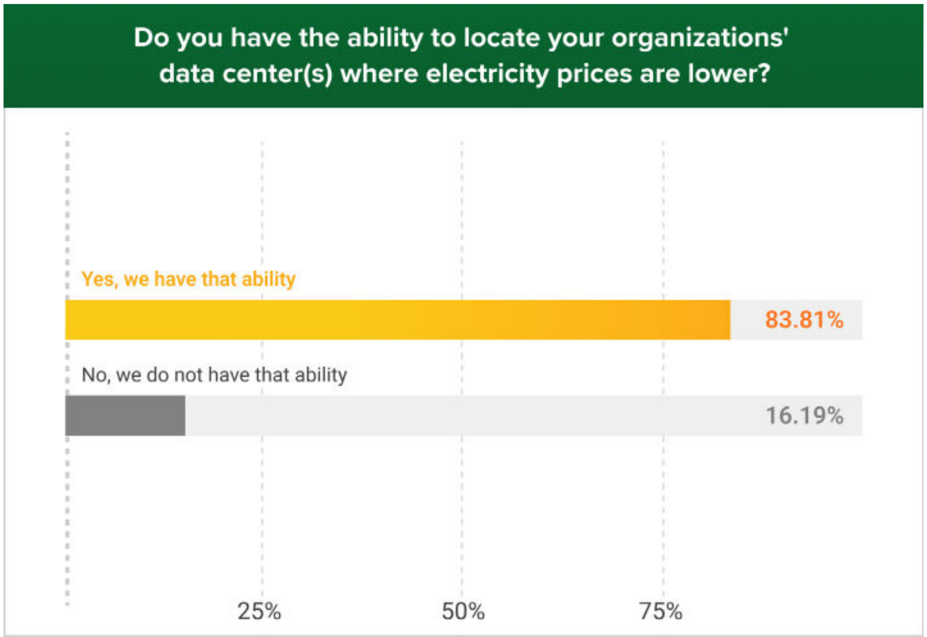


Figure 19 – Ability to Locate Data Center to Where Lower Electricity Prices Are Available

SUMMARY – ENVIRONMENTAL DATA CENTER DESIGNS REDUCE CLIMATE CHANGE

The environmental impact of a large data center is significant in terms of energy usage and E-waste. Data center owners can take a number of straightforward steps to reduce the electricity required for operating a data center. The result of these actions is both a reduction in the costs to operate the data center and doing what is better for the environment as well. Several choices can be made when designing, retrofitting, or upgrading a data center. There are many ways to make data centers more efficient and use less power, from the type and source of electricity generated to the refresh rate of new technologies. For example, looking at the location of a data center, designing the data center, and choosing suitable servers for the data center can significantly reduce the power needed for operation, which reduces the environmental impact and costs.

Appendix A

Geographical Distribution

North America 29%
Europe 25%
APAC (Not including China) 24%
China 1%
Africa and the Middle East 5%
Latin America 6%
South America 9%

Current Role within Organization

C-Level (CEO, CIO, CTS, COO) 36%
IT Management 37%
Engineering 7%
Data Center Facilities Management..15%
Other 5%

Organization Type

Software and/or Cloud Services 28%
Manufacturing 11%
Financial / Insurance 10%
Retail / Wholesale Distribution 10%
Healthcare / Pharmaceuticals 7%
Consulting / Professional Services ... 6%
Co-location / Multi-Tenant
Data Centers 4%
Education 4%
Other 18%

Organization Size

Small Office 5%
Small Business 8%
Medium Sized Business 16%
Large Sized Business 19%
Very Large Sized Business 22%
Enterprise Sized Business 30%

References

1. The Amount of Data Center Energy Use - AKCP Monitoring
2. https://www.energy.gov/eere/buildings/data-centers-and-servers
3. How many hyperscale data centres does the world need? Hundreds more, it seems | TelecomTV
4. https://datacentremagazine.com/data-centres/number-of-hyperscale-data-centres-to-reach-1-200-by-2026
5. https://www.eesi.org/topics/fossil-fuels/description#:~:text=Fossil%20fuels%E2%80%94including%20coal%2C%20oil,percent%20of%20the%20world's%20energy
6. https://www.supermicro.com/CaseStudies/Success_Story_GleSYS.pdf
7. https://www.epa.gov/climateleadership/scope-3-inventory-guidance
8. https://www.eia.gov/tools/faqs/faq.php?id=97&t=3

LEARN MORE AT WWW.SUPERMICRO.COM/GREENCOMPUTING

Super Micro Computer, Inc.
980 Rock Avenue
San Jose, CA 95131 USA
www.supermicro.com