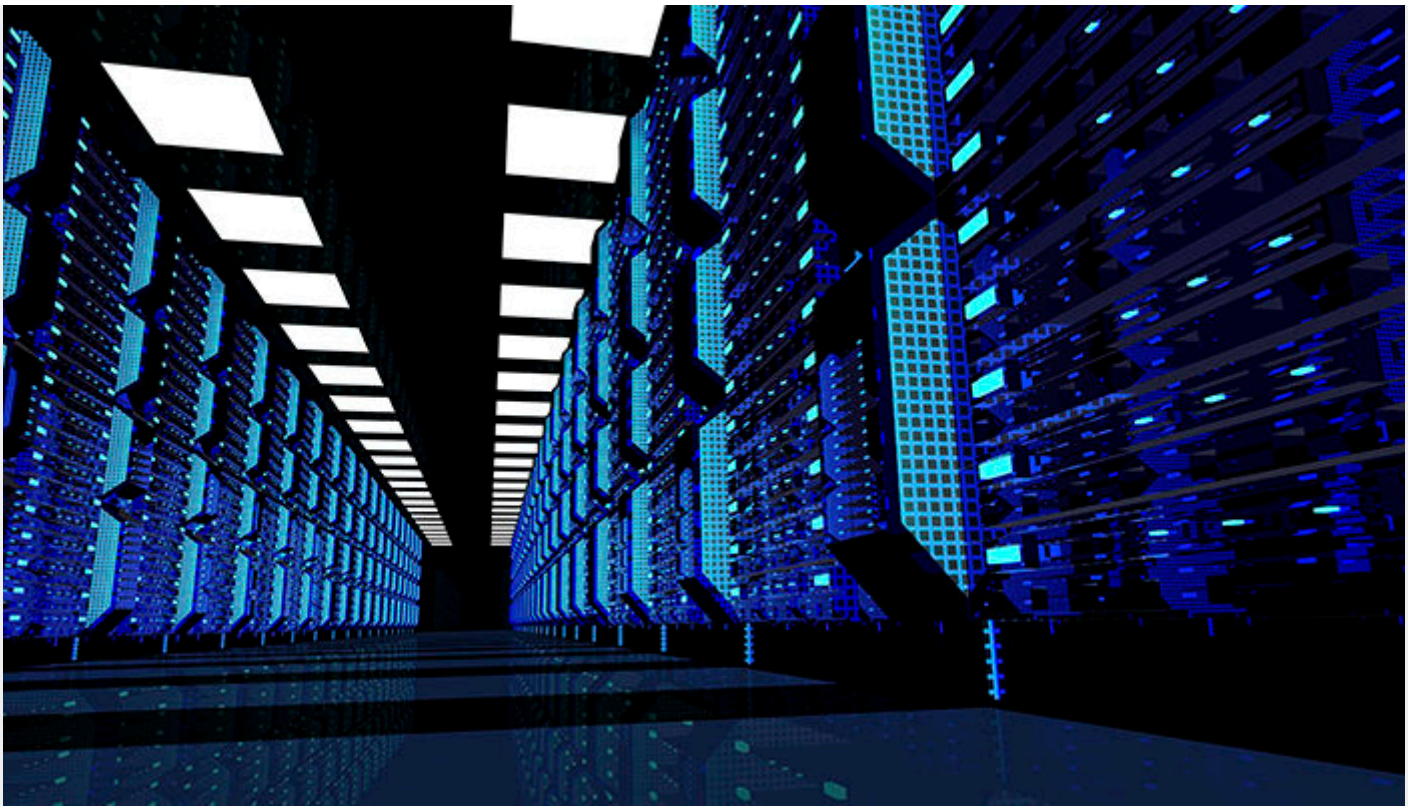
HOW DATA CENTER COOLING WORKS & CAN PROMOTE SUSTAINABILITY



The backbone of every single digital service is a vast network of servers and computing resources that deliver performance and availability necessary for business operations—and hopefully continually improve the customer and end-user experience. These resources are responsible for…

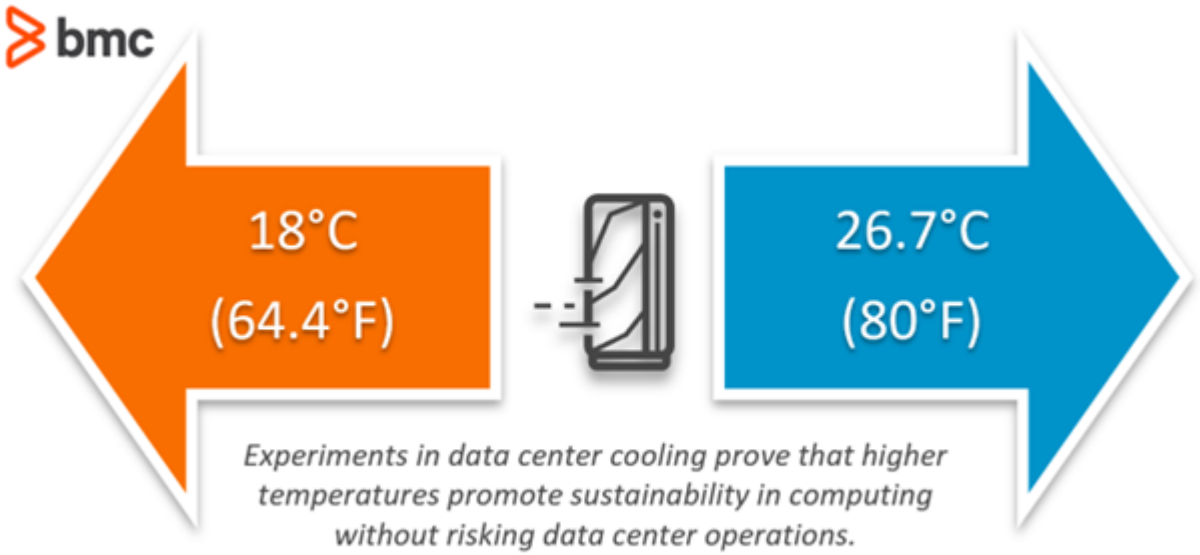
- Performing search queries
- Transferring data
- Delivering computing services

…for millions of users at any given moment, all around the world.

Now take all those servers and consider the power and heat they generate. Anyone who plays games on a laptop, desktop, or gaming platform knows how hot the equipment gets. At the height of the business world, with server rooms stacked with aisles and aisles of computing machines, that problem is significant.

Keeping servers cool is important, especially for enterprise-size businesses—but it also has a significant negative impact on Planet Earth. So, in this article, let's take a look at data center cooling and how companies can harness simple practices and modern technology to make data center cooling a more sustainable practice.

(This article is part of our Sustainable IT Guide. Use the right-hand menu to explore topics related to sustainable technology efforts.)



What is data center cooling?

Data center cooling is exactly what it sounds like: controlling the temperature inside data centers to reduce heat. Failing to manage the heat and airflow within a data center can have disastrous effects on a business. Not only is energy efficiency seriously diminished—with lots of resources spent on keeping the temperature down—but the risk of servers overheating rises rapidly.

The cooling system in a modern data center regulates several parameters in guiding the flow of heat and cooling to achieve maximum efficiency. These parameters include but aren't limited to:

- Temperatures
- Cooling performance
- Energy consumption
- Cooling fluid flow characteristics

All of the data center cooling systems components are interconnected and impact the overall efficiency of the cooling system. No matter how you set up your data center or server room, cooling is necessary to achieve a data center that works and is available to run your business.

(Explore best practices for data center migrations & infrastructure management.)

Is data center cooling necessary?

Yes—but maybe to a lesser extent than we have long believed. The general rule of thumb has been to ensure an entire room's ambient temperate stabilizes at 18 degrees Celsius. But this is unnecessarily low, as Sentry Software explains:

> *It has been proven that computer systems can operate without problems with an ambient temperature significantly higher...This is the fastest and cheapest method to reduce the energy consumed by a datacenter and improve its P.U.E.*

The oft-cited example is Google, who successfully raised the temperature of their datacenters to 26.7˚C (80˚ Fahrenheit). We'll talk more about the famous Google case study later on.

Data center managers should understand how failing to implement effective cooling technology in a server room can quickly cause overheating.

Poorly configured cooling could deliver the wrong type of server cooling to your center (for example, bottom-up tile cooling technology being used with back-front configured modern servers). This, again, would lead to serious overheating, a risk that no business should accept willingly.

Benefits of data center cooling

The business benefits of data center cooling are abundantly clear.

Ensured server uptime

Proper data center cooling technologies allow servers to stay online for longer. Overheating can be disastrous in a professional environment that requires over 99.99% uptime, so any failure at the server level will have knock-on effects for your business and your customers.

Greater efficiency in the data center

Data doesn't travel faster in cooler server rooms, but it travels a lot faster than if it was trying to travel over a crashed server!

Because data centers can quickly develop hot spots (regardless of whether the data center manager has intended a cold aisle set up or a hot aisle on), creating new solutions to cooling needs to be efficient and easily done on the fly.

This means only using liquid cooling technologies that are easily adaptable or air-cooling systems that can easily change the way cold air is used. Overall, this allows for greater efficiency when scaling up a data center.

Longer lifespan of your technology

Computers that constantly overheat at going to fail before they reach their expected end of life (EOL). For that reason, the expense of cooling systems in a data center quickly starts to pay for itself.

Introducing data center cooling technologies allows for a piece of hardware to survive longer and for a business to spend less on replacing infrastructure. Companies should be moving towards greener IT solutions, not actively creating industrial waste.

(Learn about modernizing the mainframe & your software.)

Drawbacks of data center cooling

Though data center cooling is critical for business success, it has a significant impact on the bottom line—and for Planet Earth.

Costs can be prohibitive

Data center cooling is expensive: so much energy is needed simply to run the servers and data centers. You'll spend additional energy on reducing the heat these systems generate.

For small-sized operations, state-of-the-art cooling systems are simply not possible. Expensive HVAC systems and intricate water-cooling systems that are specifically designed for data center temperature control can cost well over what SMBs can afford.

But this does not mean that SMB data centers must fail. Simple solutions like blanking panels can be used to encourage the easy flow of cold air throughout the server room. Similarly, organized cables (i.e. not a cable nightmare for an IT technician) can allow for better airflow too.

Severe impact on the planet

As much as 50% of all power used in a data center is spent on cooling technologies. Major enterprises are all moving towards reducing their carbon footprint, which means cooling technologies either have to change—or need to go.

According to the Global e-Sustainability Initiative (GESI), in their Smarter 2030 report, the digital world today, at this very moment, encompasses:

- 34 billion pieces of equipment
- More than 4 billion users

With the network infrastructures and data centers associated with these billions and billions, our digital world is responsible for 2.3% of global greenhouse gas (GHG) emissions. Data centers themselves account for 1% of the world's electricity consumption and 0.5% of CO2 emissions. And science recognizes that significantly reducing GHG emissions is a mandate in order to slow and reverse the effects of climate change.

As a global collective, tech companies must work together to reduce energy consumption. That's why we're seeing more and more companies of all ilk announce their plans to cut GHG emissions—and many of them are turning to data center cooling with a fresh approach.

How to cool data centers

Although setting up a full data center cooling system might seem daunting task, it is a necessary step. All server rooms must have the adequate amount of cooling systems that the technology demands.

In the high-performance world we live in, failing to install a necessary water loop can cause serious issues; a lack of constantly cooler air throughout data centers leads to failures—so get everything installed at the start!

Many specialist organizations work with data centers to properly identify the necessary cooling solutions, install the technology, and manage the data center's equipment from installation to EOL. Finding one in your local area is the best decision for an IT team that does not have experience in configuring cooling systems in data centers.

How cooling works: drawing heat out

In order to maintain optimal performance of the computing infrastructure, the data center must maintain an optimal room and server hardware temperature. The cooling system essentially draws heat from data center equipment and its surrounding environment. Cool air or fluids replace the heat to reduce the temperature of the hardware.

Data center cooling techniques

Data center cooling is a balancing act that requires the IT technicians responsible for it to consider a number of factors. Among many, some of the most common ways of controlling computer room air are:

- **Liquid cooling** uses using water to cool the servers. Using a Computer Room Air Handler (CRAH) is a popular way to combine liquid cooling and air cooling, but new emerging technologies like Microsoft's "boiling water cooling" are used for cooling data center servers and driving evaporative cooling technology.
- **Air cooling** uses a variety of Computer Room Air Conditioner (CRAC) technology to create east paths for hot air to leave the IT space.
- **Raised floor platforms** create a chilled space below the raised platform where a CRAH or CRAC can send the heat via chilled water coolers and other technologies which create cold aisles underneath the servers.
- **Temperature and humidity controls**, such as an HVAC which controls the cooling infrastructure, and other technologies provide air conditioning functionality.
- **Control through hot and cold aisle containment** allows hot aisles to feed onto cold aisles through the server room. Proper airflow, the use of a raised floor, and other cooling technology such as liquid cooling or HVAC cooling solutions are supported by hot and cold aisles within a data center.

Though necessary for business, these techniques require significant energy spend.

Using AI & neural networks

Significant improvements in cooling system technologies in the last decade have allowed organizations to improve efficiency, but the pace of improvements has slowed more recently. Instead of regularly reinvesting in new cooling technologies to pursue diminishing returns, however, you can now implement artificial intelligence (AI) to efficiently manage the cooling operations of its data center infrastructure.

Traditional engineering approaches struggle to keep pace with rapid business needs. What worked for you in terms of temperature control and energy consumption a decade ago is likely not enough today—and AI can help to accurately model these complex interdependencies.

How AI improves cooling

Google's implementation of AI to address this challenge involves the use of neural networks, a methodology that exhibits cognitive behavior to identify patterns between complex input and output parameters.

For instance, a small change in ambient air temperature may require significant variations of cool airflow between server aisles, but the process may not satisfy safety and efficiency constraints among certain components.

This relationship may be largely unknown, unpredictable, and behave nonlinearly for any manual or human-supervised control system to identify and counteract effectively.

Organizations today can equip data centers data centers with IoT sensors that provide real-time

information on various components, server workloads, power consumption, and ambient conditions. The neural network takes the instantaneous, average, total, or meta-variable values from these sensors to process with the neural network.

Google pioneered this approach in the 2010s, and today, more and more organizations are embracing AI to support necessary IT operations.

(Learn more about AIOps.)

Case study: Google data center

Google's implementation of the neural network reduced the error to 0.004 Power Utilization Effectiveness (PUE) or 0.34-0.37 percent of the PUE value. The error percentage is expected to further reduce as the neural network processes new data sets and validates the results against the actual system behavior. These numbers translate into a 40% energy savings for the data center cooling system.

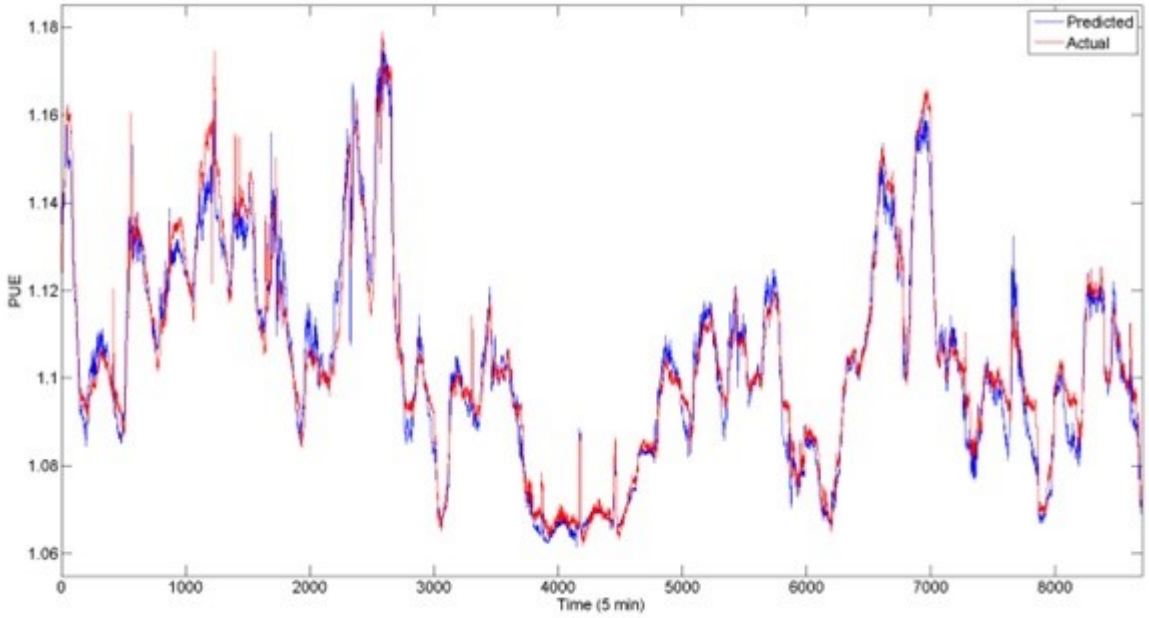


Fig. 3 Predicted vs actual PUE values at a major DC.

The graph below demonstrates how the neural network implementation delivered PUE improvements over the years. Since these results are aggregated from multiple data centers operating under different environmental and technical constraints, the optimal implementation of the machine learning algorithm promises even better improvements in comparison with traditional control system implementations.

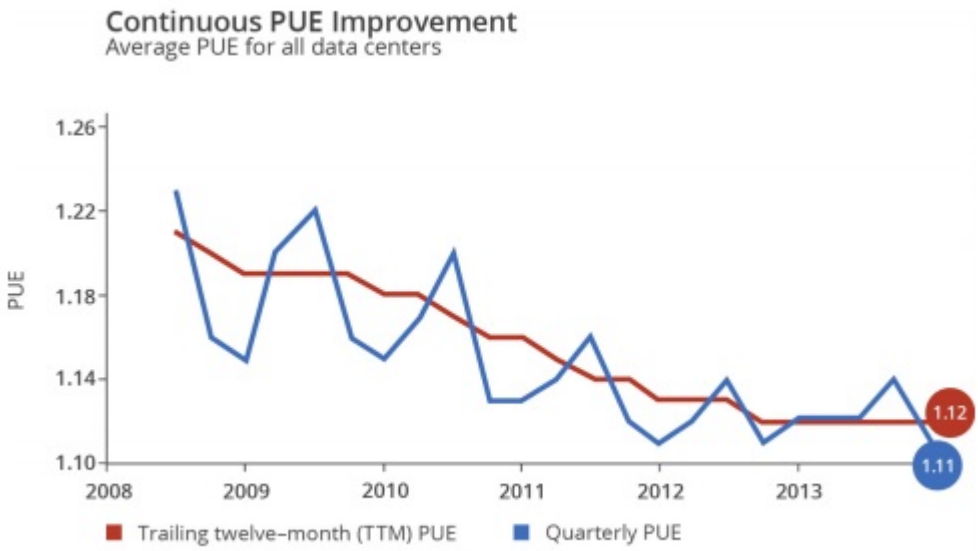


Fig 1. Historical PUE values at Google.

The neural network algorithm is one of many methodologies that Internet companies including Google may have implemented for data center cooling applications.

Related reading

- BMC IT Operations Blog
- BMC AIOps Blog
- Data Center Guide, a multi-part series
- Data Center Infrastructure Management (DCIM) Solutions Explained
- How To Secure a Server Room
- IT Salary Trends Today: The Complete Roundup