WHITE PAPER

How Artificial Intelligence Is **Cooling Data Center Operations**

Dynamically match cooling to IT load in real time

Executive overview

Recent swift changes in societal habits have increased demand for data usage at an unprecedented rate. While this may mean more revenue for commercial data centers, the surge in usage is also increasing risks of downtime – creating more demand on staff, equipment and energy consumption. These changes give rise to a bigger challenge: how do you scale to meet current demand and plan for future capacity in an age of hyperconnectivity?

Artificial Intelligence (AI) offers data centers promising solutions to improve operations over the long term. In fact, Gartner analysts predict that the early adoption of AI will be a key factor separating data centers of the future from those destined to be dinosaurs of the past. While incorporating AI into an organization's systems can be a long, challenging process, there is good news. Data centers can easily and successfully implement AI in their operations with new thermal cooling solutions.

This paper explores AI and its impact on data centers, using white space cooling optimization as an example of how AI can be implemented today. Starting with a look at the changing data center landscape, this paper provides a glimpse into what the future holds and examines the critical aspects of thermal cooling, specifically thermal optimization. It focuses on how a data center can easily begin integrating AI into its processes through whitespace cooling optimization (WSCO) and reviews how a global financial firm is using WSCO as part of its thermal optimization plan, with promising results.

SIEMENS

Table of Contents

Challenges facing data centers today

Out of necessity, data centers have continually evolved to keep up with customer demands and high-tech innovations. The challenges of staying in front of new trends and changing user requirements are significant. The constant question is: What's next?

In early 2020, Forbes Insights, the strategic research and thought leadership practice of Forbes Media, attempted to answer that question. In doing so it identified three major challenges facing the data center of the future:

- **Lack of readiness to upgrade:** The nature of data centers is changing. To be able to tap into a variety of new cost-cutting opportunities, data centers need to decentralize. They need to move data, processing and resources away from the organization's local data hub to the edge (user devices) – moving aspects of processing to the devices used by end users.
- **Infrastructure challenges:** As 5G enables hyperconnectivity, more users and devices will enter the network, increasing traffic volume and bandwidth usage. To address this new demand, data center leaders need to incorporate smartinfrastructure into their data centers plans, integrating automated systems supported by devices with built-in artificial intelligence capabilities that provide real-time maintenance, configuration and resolution.
- **Staffing challenges:** Forbes Insights cites recent research that reports that one-third of the IT infrastructure workforce will be retiring over the next decade. In response, more than half of the executives surveyed stated that they expect staffing will be taken over by external cloud and service providers.

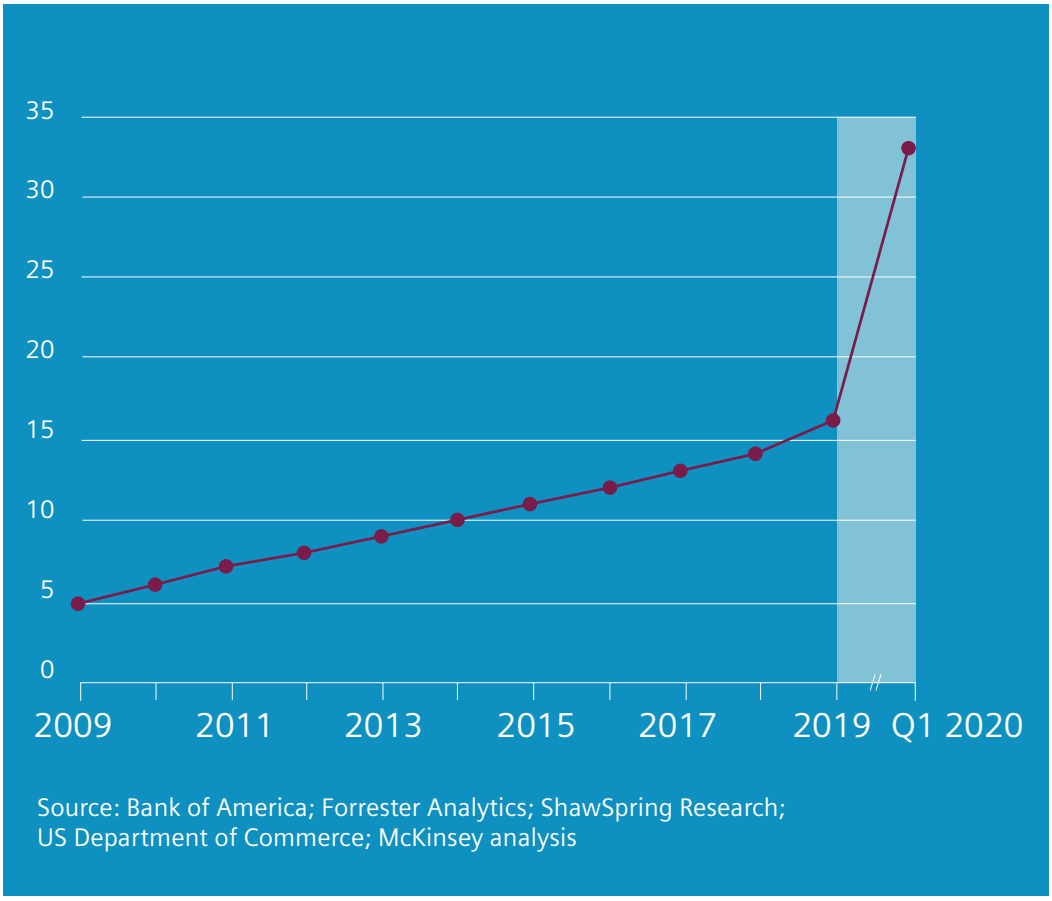
Surprisingly, very few organizations surveyed were ready to rethink their data center strategies in preparation for the next five years. Just 11% of C-suite executives and 1% of engineers said their data centers are ahead of the curve and primed for higher data volumes, according to Forbes Insights.

Surprisingly, very few organizations surveyed were ready to rethink their data center strategies in preparation for the next five years.

Since the survey was taken, the world has changed. While readiness, infrastructure and staffing remain areas of significant concern for the future, the main challenge facing data centers today is exponential growth. Mobile networks propelled by 5G technologies are providing a new level of hyperconnectivity. In addition, more people are working remotely, transforming American workplaces and adding a significant spike in demands on data center use.

At the same time, online shopping and ecommerce have surged, further fueling the demand for data and data storage. In fact, ecommerce experienced the equivalent of 10 years of growth in the space of 90 days due to COVID-19, according to an analysis conducted by McKinsey & Company. In the 10-year period between 2009 and 2019, U.S. ecommerce penetration grew from 5% to roughly 16%. In the first three months of 2020, ecommerce penetration doubled in the United States, growing from 16% to almost 35%.

US e-commerce penetration, %



Source: Bank of America; Forrester Analytics; ShawSpring Research; US Department of Commerce; McKinsey analysis

CEOs and CIOs are now focusing on how to digitally transform their operations so they can keep up with rapid increases in demand.

With this exponential growth, CEOs and CIOs are now focusing on how to digitally transform their operations so they can keep up with rapid increases in demand. Digital transformation is a fundamental rethinking of an organization's use of technology, according to George Westerman, senior lecturer at the MIT Sloan School of Management and author of the book Leading Digital: Turning Technology Into Business Transformation. Westerman argues that IT departments must break out of their traditional roles in organizations and quickly become co-creators of "digital transformation" in order to help their companies thrive in today's rapidly changing world.

AI's impact on data centers

Artificial Intelligence (AI) is part of the digital transformation and is poised to have a tremendous impact on data center management, productivity and infrastructure. Several companies already are experimenting with the idea of fully automating data centers for periods of time, allowing a "lights out" operation. The theory is that a number of energy-efficient strategies can be implemented if data centers were built and optimized entirely based on IT considerations and monitored remotely via DCIM software. By using AI and robots to eliminate the need for employees in the physical space, data centers could lower oxygen levels to reduce fire risk, eliminate the need for lighting, create more efficient cooling designs, and increase rack heights.

But there are challenges. While leading organizations expect to double the number of AI projects over the next two years, according to the Gartner 2020 CIO Agenda Survey, most are struggling to implement them. Stakeholders face challenges that include transparency and "interpretability" around algorithmic decision-making, data quality and bias, safety and security, accountability, social and economic impact assessments, and governance. One of the biggest obstacles is scaling AI pilots into enterprise-wide production. AI can address many issues in a controlled environment; however, replicating the experience across an organization is fraught with difficulty. Due to these challenges, most companies question how quickly AI can provide a return on investment.

What is AI?

Artificial Intelligence (AI) is the creation of human-like intelligence using computer-based algorithms. According to the Oxford Dictionary, AI is the "theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages."

AI systems are based on three types of computer-based machine learning: supervised learning, unsupervised learning, and reinforcement learning. White Space Cooling Optimization (WSCO) uses supervised learning, where you train the machine to use data that is "labeled." In other words, the computer is fed data that is tagged with correct answers to potential questions, or scenarios. Its algorithm learns from this labeled training data how to map answers between input examples (questions) and the target variable (best answers). The algorithms teach the computer memory to how to predict outcomes from similar, unforeseen data and respond to future events.

Where can AI provide immediate results?

One area where AI can immediately deliver real benefits is data center cooling and control. As demand for data grows, so does the need to better manage cooling conditions in data centers.

Among the factors that make control challenging is the mix of legacy equipment from a variety of manufacturers found in many data centers, each with its own recommended standards that often call for more cooling than is actually needed. Managing airflow is also complex, especially with the dynamic nature of data centers and the unpredictable demands of IT equipment. Added to this is the fact that on average, 20% to 30% of servers in large data centers are unused or obsolete but still consume electricity. These ghost or zombie servers not only add costs to the electrical bill, they also create excess heat and drive higher demand for cooling throughout the data center. As a result of all these factors, most data centers are overcooled, wasting valuable resources.

On average, 20% to 30% of servers in large data centers are unused or obsolete but still consume electricity.

The answer: thermal optimization!

Thermal optimization is a comprehensive Siemens solution that eliminates the need to manually maintain the optimal, cool and consistent temperatures required to house data center equipment safely. Its benefits go beyond temperature control to include the optimizing of energy use and personnel time to better manage resources.

Thermal optimization features a package of facility improvement measures based on data collection and analysis. The performance of everything from the cooling tower to the rack is measured and approved. The solution also applies intelligence about the density of IT workloads to regulate the cooling within a data center. The goal is to optimize primary and secondary cooling cycles as well as the temperatures within the “white space” – the unusable square footage of a data center.

In this paper, we define thermal optimization specifically as the combination of Demand Flow® chiller optimization and White Space Cooling Optimization (WSCO).

Demand Flow: To reduce energy costs and ensure ongoing operational efficiency, Demand Flow takes a holistic approach that employs thermodynamic principles. It focuses on optimizing every subsystem of the chiller plant, including chillers, chilled water pumps, condenser water pumps and cooling towers. A powerful solution that provides immediate results, Demand Flow collects and analyzes data to deliver the proper amount of chilled water to meet the current cooling load.

White Space Cooling Optimization (WSCO): Using a network of sensors, WSCO collects temperature and air supply data. Its AI Engine applies the data to algorithms and calculates the required adjustments in airflow to maintain the correct temperature for each aisle of racks. It goes beyond alerting human operators to temperature fluctuations and automatically makes adjustments itself. By automating the control of cooling fans, it reduces the risk of a thermal outage and maintains consistent air temperatures among server racks and white space. It also reduces wasted energy by dynamically matching cooling in the IT load in real time, automatically responding to temperature changes andeliminating overcooling. At the same time, it gives the management team critical data that supports future operational business decisions.

Chiller optimization with Demand Flow

Demand Flow operates on the simple premise that we need to think differently about chilled water systems. It challenges the industry norm of pumping a constant volume of water through the chilled water system all year long with little regard for the varying load on the system.

Instead, Demand Flow takes a holistic approach that uses thermodynamic principles. It focuses on optimizing every subsystem of the chiller plant, including chillers, chilled water pumps, condenser water pumps, and cooling towers. Through the use of high-accuracy precision instrumentation and variable pressure curve logic, Demand Flow is able to dynamically adapt to load changes while maintaining optimal total system performance.

A Demand Flow solution typically includes installing variable speed drives on all chilled water pumps, condenser water pumps, and cooling tower fans as well as employing the high-accuracy controllers and instrumentation needed for proper control.

With Demand Flow, water flow varies through both the evaporator and condenser tube bundles in near-real time to perfectly optimize total system energy consumption in response to system load fluxuations. The effects of optimized flow through the chillers in near-real time produces significant energy savings at the chiller and can increase the nominal tonnage of a typical plant by as much as 20%.

Benefits of thermal optimization

By optimizing chilled water cooling (Demand Flow) and white space cooling (WSCO) together, thermal optimization can help a data center achieve more savings than using one or the other solution individually. The combination delivers many benefits, including:

1. Collecting indoor environmental data in real time and providing valuable insights for making short- and long-term operational decisions
2. Freeing up capacity and better positioning data center workloads
3. Providing dynamic control and energy savings by managing thermal airflow
4. Enabling systems to adjust in real time to match facility cooling needs as IT loads change
5. Maintaining optimal performance by dynamically controlling pumps, chillers and fans
6. Optimizing central plant operations to help a facility achieve energy savings up to 50%

Benefits of thermal optimization

By optimizing chilled water cooling (Demand Flow) and white space cooling (WSCO) together, thermal optimization can help a data center achieve more savings than using one or the other solution individually.

AI and WSCO

AI is what makes White Space Cooling Optimization (WSCO) so effective in improving temperature control and cool air distribution with less staff support. WSCO's AI engine utilizes a type of machine learning known as supervised learning that enables the WSCO software to continuously get smarter based on the calculations of a series of algorithms. The software learns by continuously analyzing the sensor data for environmental changes. Past and current records of temperature control data are used as a basis for the algorithms and to establish baselines. The software quickly learns how to best respond to temperature variations.

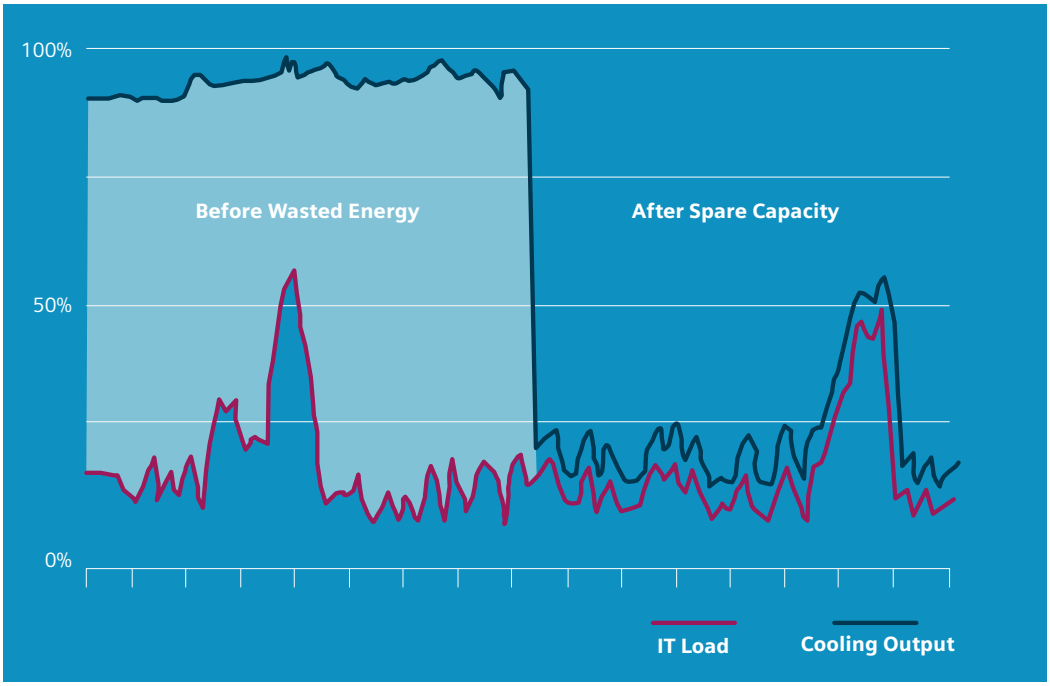
The backbone of WSCO consists of a wireless mesh system with a dense array of sensors and controllers that fuel the powerful analytics and machine learning process. Rack-level sensors capture both top and bottom rack temperatures while cooling unit controllers provide data such as supply and return temperatures, status, speed, and cooling unit power. This data helps identify cooling influences on the racks so that WSCO's learning software can dynamically determine how much cooling is needed for ideal conditions.

Leveraging data collected by the sensor network, WSCO's AI engine automatically creates a real-time model of the facility's thermal environment. The AI engine maps influences and determines the precise cooling influence of every Computer Room Air Handling (CRAH) unit, both individually and collectively, at every spot across the data center. In most facilities, the AI engine is able to identify influence patterns and measure cooling influences in less than 24 hours.

The WSCO system then takes dynamic control of the cooling units – turning them on and off, and ramping fan speeds up and down — to meet pre-specified temperature settings in the most efficient manner possible. As the AI software learns the effects of control actions, it manipulates the cooling equipment by itself without staff intervention, automatically managing cooling and balancing airflow in critical areas in the data center hall.

In the event of a system failure or when temperatures exceed a certain threshold, the WSCO system has a "guard mode" that activates cooling. The "guard mode" delivers added protection to the data center until the mechanical system returns to normal operation, all driven by the automation and learning algorithms.

WSCO's impact on air cooling efficiency



The AI engine uses real-time data to produce algorithms that predict the best level of cooling that will deliver the desired temperature at each sensor.

WSCO solutions are infrastructure agnostic, so they are immediately compatible with existing systems with minimal on-site configuration and setup. The WSCO wireless architecture makes installation non-intrusive and flexible. Without the need to run hundreds of cables, even large sites can be up and running in a matter of weeks.

WSCO's key success factor is its ability to dynamically match cooling to the IT load in real time. The AI engine uses real-time data to produce algorithms that predict the best level of cooling that will deliver the desired temperature at each sensor. And because most facilities are over-cooled, WSCO uncovers pre-existing redundancy and redirects airflow. It typically level-sets cooling capacity while delivering significant reductions in energy use.

Increasing staff efficiency

As a truly autonomous solution, WSCO can be capable of increasing staff efficiency in numerous ways. It minimizes the need for staff supervision and personnel on site, allowing valuable staff to be assigned to other critical tasks and limiting overall site access. While matching cooling requirements to IT load in real-time, WSCO can self-correct and quickly scale to meet demand, eliminating the need for human intervention. At the same time, it provides visibility into real-time system performance so operators can make necessary adjustments with the user interface. As a result, data centers don't need employees or third-party vendors to physically address the tactical, complex and time-consuming task of adjusting cooling systems or to manually recommission the server room. When third-party support is needed, the WSCO system can quantify the effectiveness of the vendor's work by using historical data to compare performance before and after maintenance actions and ensure that repairs are fully resolved.

WSCO increases staff efficiency by minimizing the need for staff supervision and personnel on site, allowing valuable staff to be assigned to other critical tasks and limiting overall site access.

Improved planning and budgeting

Since the WSCO interface can integrate with any building management system (BMS) via BACnet/IP, it gives managers immediate access to real-time, trend-based insights and new levels of control at their fingertips. For example, WSCO may help improve capacity planning. Once overcooling is addressed, managers can use the system to identify how much actual cooling is still available and then determine when and where to deploy new IT equipment. Likewise, managers may be able to defer or reduce capital expenditures since fewer units are needed to meet day-to-day cooling requirements and duty cycles on existing equipment can be reduced, which may lead to extending equipment life. Managers have the visibility and reporting needed to identify exactly which cooling units need to be maintained, which are running optimally and which are underperforming. Easy identification of poorly performing equipment helps managers plan when equipment needs to be retired and which low-efficiency units should be the first to go. In addition, operators can use the interface to identify any area in the data center that faces thermal risk so that high-value IT assets are not deployed there until the problem is resolved.

To maintain configured limits, operators have the option to continually improve thermal management by directly and automatically managing temperatures as they see fit. They can also operate their sites at higher temperature set points while increasing reliability and maintaining quantifiable SLA compliance.

Building toward the future

In combination with other emerging technologies, AI promises to reshape data center operations in the coming years. However, data centers need to prepare today to be relevant tomorrow. AI-managed thermal cooling solutions like WSCO help data centers lay the groundwork for the future. These solutions help managers address today's challenges – personnel management, operating costs, operational efficiency and energy efficiency – while preparing their operations to adopt other new AI applications. The simplicity of WSCO's agnostic infrastructure design, for example, makes it responsive and scalable to meet future demands. It's an AI application that is ready to implement today and a smart investment for long-term operational excellence.

The simplicity of WSCO's agnostic infrastructure design makes it responsive and scalable to meet future demands.

Use case: WSCO keeps financial data cool and safe

In 2009, a global financial firm invested heavily in the construction of a new data center that was classified as a Tier III facility by the Uptime Institute. Less than five years later, the firm realized it needed to increase the data center's cooling capacity while reducing its energy consumption. The data center was built with redundant mechanical plants for both chilled water and air handling. The chilled water plant used a typical primary\secondary loop chilled water system serving constant volume Computer Room Air Handling (CRAH) units in the data halls. While the data halls were built following typical data center design standards, their rack density underwent numerous changes that created hot spots. The hot spots caused chronic over provisioning of cooling throughout many of the data halls. And like many other data centers, the facility ultimately lost its Tier III rating because it had to turn on what should have been redundant chillers, pumps and CRAH units to mitigate the hot spots. The data center also faced increased energy consumption and a loss of adequate cooling redundancy throughout.

The company hired Siemens Smart Infrastructure to implement a comprehensive thermal optimization strategy for the data center. The first step was to address the chilled water plants' inability to provide cooling to the farthest end of the secondary loop without running its redundant chiller systems. To help resolve this issue, Siemens implemented its Demand Flow solution for chilled water. The solution helped the data center accomplish two initial goals: it was able to reduce the total number of chillers needed to support the load and it was able to utilize the existing plate and frame heat exchanger longer. As a result, Demand Flow helped the data center achieved an immediate 37% reduction in annual energy usage. It also achieved a $200,000 utility rebate and more than $200,000 in annual savings, which resulted in a simple payback in 2.75 years.

Second, after optimizing the chilled water plant, Siemens addressed the data center's airflow issue: it had to run all 72 of its CRAH units in order to satisfy the data halls' cooling needs. Each 30-ton CRAH unit was originally designed as constant air volume (CAV) units. Siemens converted them to variable air volume (VAV) and used ModBus to integrate them to the data center's Building Automation system for greater visibility and control.

The third step was to apply White Space Cooling Optimization (WSCO), which has built-in Artificial Intelligence (AI) for real-time learning and control. The WSCO solution employs wireless technologies to monitor rack inlet temperatures. Then, using AI-enabled predictive control algorithms, WSCO automatically manages cooling and balances airflow in areas of influence in the data center hall. The results were significant: WSCO reduced the need for CRAH units from 72 to 35, restoring the data center's Tier III rating. The solution also achieved a 71% savings in kilowatt hours, from 241.7 kW to 69.1 kW. This generated $241,817 in annual savings, a utility rebate of $150,000 and a simple payback in less than two years.

To top it off, an additional 5% in energy savings was realized at the chiller plant due to a reduction in required pumping after WSCO was completed.

The data center achieved an immediate 37% reduction in annual energy usage. It also achieved a $200,000 utility rebate and more than $200,000 in annual savings, which resulted in a simple payback in 2.75 years.

Legal Manufacturer

Siemens Industry, Inc.
Smart Infrastructure
1000 Deerfield Parkway
Buffalo Grove, IL 60089
Tel: (847) 215-1000

(05/2023, Part # 153-SBT-481)