Justin Scott

Climate-Conscious Data Center Cooling In an AI Era

BY JUSTIN SCOTT

Growth in power-hungry artificial intelligence workloads, rising costs and sustainability initiatives could drive new approaches in data center cooling. Sustainable, high-efficiency components, such as fans, pumps and motors could play a bigger role for HVAC equipment that must work harder to keep data centers cooler.

The massive growth in ChatGPT and other generative artificial intelligence (AI) technologies could place more energy demand on data center cooling systems with the shift toward a new generation of power-hungry chips built for AI.

AI computing workloads have been doubling every three to four months since 2012, and they are likely to accelerate further, spurring demand for AI computing hardware over the next several years.[1]

AI requires high-performance processors that generate more thermal power than traditional chips. Many of the chips also require lower operating temperatures, such as ASHRAE Class H1 as defined in ASHRAE's *Thermal Guidelines for Data Processing Environments*, which will likely push data center operators to approach cooling differently.

Other pressures, such as rising energy and material costs, as well as space limitations, have operators maximizing their investments in data centers and working to reduce energy demand. Limited supply and strong demand drove up some rental rates in the second half of 2022 by 14.5% year-over-year.[2]

Sustainability Initiatives

At the same time, momentum is building with government officials for more guidance related to data center efficiency and sustainability. Regulators in many countries are working toward requirements to report data centers' operating information and environmental performance metrics. The European Commission (EC) Energy Efficiency Directive (EED) was updated in September 2023 and is a fundamental principle of EU energy policy. It mandates three levels of information reporting, the application and publication of energy performance improvement and efficiency metrics and conformity with certain energy efficiency requirements.[3]

In late 2022, the U.S. Department of Energy (DOE) announced $42 million in funding to help develop high-performance, energy-efficient cooling solutions for data centers. An industry report also found 63% of data center operators think authorities in their region will require data centers to publicly report environmental data in the next five years.[4]

Justin Scott is applications engineer for Infinitum.

Motors used to power data center cooling equipment.

Investors are also demanding that companies report sustainability-related efforts and actions because sustainability is seen as a competitive advantage. Google, Microsoft, Facebook, Amazon and Apple are all investing in efforts to reduce greenhouse gas emissions to become carbon neutral or negative and to at least match electricity consumption with renewable energy.[5]

Cooler Applications for AI Growth

With a perfect storm of generative AI growth, rising costs and pressure to report on carbon metrics, HVAC equipment manufacturers are prioritizing sustainable, power-dense equipment that can handle higher levels of heat, while keeping data centers operating at lower temperatures.

Data center cooling applications that include liquid cooling, air-handling units that can leverage free cooling and economizers, and precision cooling are effective ways to increase the energy efficiency of data centers with AI workloads. Each approach—liquid cooling, free cooling and precision cooling—has its own benefits and, depending on the data center's needs, may be applied in combination to optimize cooling operations.

Liquid Cooling

Compared to traditional air cooling, liquid cooling is more efficient at transferring heat away from electrical components. In fact, liquid's heat-carrying capacity can be up to 3,500 times greater.[6] Water-cooled data centers use about 10% less energy and thus emit roughly 10% fewer carbon emissions than many air-cooled data centers. In 2021, water cooling helped a major internet provider reduce the energy-related carbon footprint of their data center portfolio by roughly 300,000 tons of CO_2.[7]

A few ways exist to apply liquid cooling; each has challenges. With direct-to-chip cooling, coolant travels through pipes to a cold plate incorporated into the motherboard processors, which disperses heat. The extracted heat is fed into a chilled water loop and transported to a chiller plant at the facility. This method is among the most efficient forms of cooling a server room. While this is great for the processor, it doesn't cool other components. A second approach is bringing liquid cooling to the rack where air supplements the cooling. A third is immersion cooling of server racks in a tank where the fluid surrounds server boards and removes heat. The challenge is moving parts in and out of fluid and potentially contaminating/damaging other components. With liquid cooling, data center operators still require traditional air coolers, chillers and evaporative cooling.

Free Cooling

ASHRAE Standard 90.1 allows data centers in Northern climates to draw cool air directly into the data center, reducing energy demand. Airside economizers bring outdoor air into a building and distribute it to servers, and instead of being recirculated and cooled, the exhaust air from the servers is directed outside. If the outdoor air is particularly cold, the economizer may mix it with the exhaust air so its temperature and humidity fall within the desired range for the equipment.

Other cooling applications can turn off compressors to save energy. Water-cooled chillers with cooling towers provide the capability to turn the chiller compressor off and circulate water/glycol from the cooling tower directly to the air-handling unit (AHU). Air-cooled chillers allow turning off the compressor and circulating water/glycol from the chiller to the AHU. Rooftop units can turn off the compressors, open ducts and bring outdoor air directly into the data center. A computer room air handler (CRAH) uses modulating fans to circulate hot return air from the data center over the chilled water coil to reject the heat and provide cold air back to the data center. When used in areas with colder annual temperatures, it can use free cooling and reduce the need to operate compressors.

Precision Cooling

Close-coupled data center air-conditioning units bring heat transfer closer to the equipment rack and precisely deliver inlet air and more immediately capture exhaust

air. In-row air conditioners with external condensers are a type of open-loop configuration for close-coupled cooling. They are installed inside the rack rows, allowing airflow to follow a short, linear path. This reduces the power needed to start the fans, helping increase energy efficiency.

While racks of servers require the most energy to operate a data center, server cooling equipment is close behind. Given an understanding of effective ways to cool data centers, efficient components—such as fans, pumps and motors—can have an impact on the sustainability of HVAC equipment, and they are important for reducing energy demand as AI workloads increase.

Efficient Components for Data Center Cooling Applications

Below, we'll discuss design considerations to help ensure higher levels of efficiency and sustainability in fans, pumps and motors and other components.

Fans

Data center facilities need well-designed fan systems to maintain airflow between aisles, to cool servers and to sustain appropriate humidity levels, but fans often use a lot of energy. Even a modest investment in a more efficient unit can have a big impact. Typically, centrifugal fans have a higher peak efficiency than axial fans, but the actual efficiency at the operating point of the fan depends on the flow and pressure conditions of the application. Generally speaking, axial fans are preferred for high volume flow rates (cfm), low static pressure and non-ducted applications. Centrifugal fans are used for higher static pressure systems.

When designing fans, design teams choose a fan model that can deliver on flow rates, static pressure and air density (altitude and temperature) requirements amongst others to obtain the fan brake horsepower. Then they select a motor to drive the fan at the customer's operating parameters. Today many of the fan systems used to cool data centers use motors that have more horsepower than the application requires. Rightsizing and customizing the motor to match the application's horsepower, speed and

torque requirements can reduce the overall electrical infrastructure required for a data center.

When selecting a custom motor, designers should look for high efficiency across a wide range of load conditions. Matching a motor's peak efficiency with a fan system's peak efficiency provides the optimal efficiency for the system overall. Customizing motors also allows designers to get close to the exact motor configuration needed, and it reduces input current requirements, which can add up to significant energy and electrical equipment savings for the whole system. A savings of 5.6 amps for a motor that fits system requirements amounts to 22.4 amps for a single four-fan array module. For 100 fan arrays, that adds up to a savings of 2,224 amps, which could reduce electrical infrastructure requirements.

Pumps

In data centers, pumps move coolants and liquid for air-handling units, cooling coil boosters and liquid cooling applications. Pumps that vary speed can reduce their energy consumption by reducing pump speed to match load requirements. Integrated variable speed controls can also eliminate the need for throttling downstream valves to match demand, saving both energy and infrastructure wear and tear. Recently, the U.S. Department of Energy (DOE) issued a report[8] on pump efficiency that found that improvements to motor efficiency and demand-based variable speed controls can yield greater energy savings than those from improved hydraulic efficiency—saving upwards of 65% of energy use depending on the application. Motors with advanced control capabilities enable the remote monitoring of vibration, temperature, speed and efficiency, which feeds back to pump controllers and allows motors to adjust and self-protect on demand to reduce maintenance issues and energy consumption.

Pumps that integrate compact motor systems can free up valuable facilities' space by taking a typical packaged pump skid solution and reducing its footprint by 50%. Selecting motors with axial-flux designs and thin PCB

stators (versus heavier iron-core motors) can result in 50% smaller and lighter form factors that can reduce overall space required in data center air-handling, chillers, coolers and liquid cooling applications. This can allow for closer proximity to data center equipment and for more effective, efficient cooling and removal of heat.

Motors

Currently, 40% of the energy in data centers is consumed by motors that power cooling equipment, such as fans, pumps and compressors to reduce heat generated by servers. A typical data center averages 500 electric motors in their HVAC systems alone, so prioritizing efficiency should be a top design consideration.

In a data center, where loads can vary widely depending on the time of day, it's important to identify a motor that has a flat efficiency curve across a wide range of load conditions to optimize overall HVAC system efficiency and operations. Doing this will ensure the system will benefit from optimum efficiency under a variety of operating conditions. Advanced air core motors can run at variable speeds with partial load efficiency, saving energy at off peak times, while maintaining high levels of efficiency across a wide range of loads and speeds.

Data centers have hundreds to thousands of motors with valuable materials, such as aluminum, steel and magnets that can benefit from refurbishment and reuse. Today, the majority of motors end up in landfills after 10 to 20 years, but advances in circular motor design are allowing for extended component life and reuse. Motors that take advantage of modular design can make the motor easy to maintain and allow for components to be reused multiple times.

Other Components

Other HVAC components, such as alternative refrigerants with low CFC levels, equipment housing and coils made using fewer materials and resources, and higher-efficiency air filters, can contribute to the overall sustainability of cooling systems and data centers.

Conclusion

As AI workloads accelerate, costs remain a challenge, and pressures mount to report on sustainability metrics, HVAC system manufacturers designing data centers must prioritize higher-efficiency components. Doing so can reduce the overall energy demand and carbon footprint of data centers, while limiting the impact on the environment and the next generation.

References

1. J.P. Morgan. 2023. "Is Generative AI a Game Changer?" J.P. Morgan. https://tinyurl.com/bdzaxaa8

2. CBRE. 2023. "North America Data Center Trends H2 2022." CBRE. https://tinyurl.com/3tab5vee

3. Dietrich, J. 2023. "First Signs of Federal Data Center Reporting Mandates Appear in U.S." Uptime Institute. https://tinyurl.com/4vmu7v9c

4. Davis, J., D. Bizo, A. Lawrence, O. Rogers, M. Smolaks. 2022. "Uptime Institute Global Data Center Survey 2022." Uptime Institute. https://tinyurl.com/hrfv4efz

5. Ziser, K. 2023. "AWS, Microsoft Lead the Charge in Data Center Sustainability." LightReading. https://tinyurl.com/2mmzkx7n

6. ASHRAE. 2013. *Liquid Cooling Guidelines for Datacom Equipment Centers, 2nd Ed*. Atlanta: ASHRAE.

7. Holzle, U. 2022. "Our Commitment to Climate-Conscious Data Center Cooling." Google. https://tinyurl.com/ead9mmmt

8. Federal Register. 2022. "Energy conservation program: energy conservation standards for circulator pumps." Federal Register 87(233) 10 CFR 431. https://tinyurl.com/rx8xe24n ∎