

Part 3: Interactions Between IT Systems and Data Centers

Changing Landscape Of Data Centers

BY DONALD L. BEATY, P.E., FELLOW ASHRAE; DAVID QUIRK, P.E., MEMBER ASHRAE; JEFF JAWORSKI

From the moment a data center's design is finalized, it may no longer be optimal for the information technology (IT) equipment it will someday contain. Not only does the IT equipment in a data center rarely remain unchanged, but also the loads the IT equipment can experience may vary significantly over time or even within the same day. These changes can easily affect cooling, power, airflow, acoustics, and other considerations, creating an ever-changing landscape the data center must accommodate.

An inherent mismatch of time scales exists between IT equipment and the data center in which it operates. IT equipment is responsive, adapting immediately to its workload and environment. Data center facility equipment, on the other hand, reacts when changes within the data center reach critical mass where larger responses are necessary. For example, if servers consume more power by handling a higher compute workload, internal fans can speed up instantly to remove the additional dissipated heat. In contrast, the facility cooling may wait for a measurable increase in return air temperature before increasing cooling capacity.

Further, many interactions exist between the IT equipment and the data center facility systems, not just those related to power consumption and heat dissipation. The differential air temperature (ΔT)

between supply and return air is inversely related to how much airflow is required to remove some amount of heat. The ability of a server to affect airflow is affected by the pressure differential between the intake and exhaust of the IT equipment. Even acoustics within the data center are directly related to fan speeds.

Combining the multitude of interactions and dependencies with the mismatch of time scales creates a potentially challenging environment in which all the equipment is in a constant state of reacting to the reactions of other equipment. This highlights the importance of understanding and designing for these interactions in ways that yield predictable and/or addressable results.

Donald L. Beaty, P.E., is president, David Quirk, P.E., is vice president, and Jeff Jaworski is an engineer at DLB Associates Consulting Engineers, in Eatontown, N.J.

Differential Air Temperature

For a long time, data center operators were mainly concerned with IT power consumption. This was a clear indicator of the amount of power that would need to be supplied to the equipment and of the heat that would be dissipated by the equipment and need to be removed. But in the quest to improve data center efficiency, other important metrics have emerged— ΔT being one of them.

Before entering the spotlight, ΔT was typically just a consequence of the supply air temperature and the dissipated power of the equipment. Whatever heat could be picked up by the supply air defined the ΔT . However, to increase server efficiency, additional sensors and more complex thermal management techniques have allowed servers to operate at higher average temperatures while managing the risk that any portion of the server could overheat. This has allowed both a reduction in fan speeds within the server (to not provide more cooling than needed) and more efficient heat rejection (less airflow is needed to remove the same amount of heat).

As IT manufacturers began to design equipment for higher ΔT , the consequence has been that exhaust air temperatures have increased since the typical supply air temperatures have certainly not been decreased (colder supply air temperatures would be less efficient for the facility to provide). This was compounded by a general increase in supply air temperatures in data centers as IT manufacturers agreed that warmer air at IT intakes would not impact reliability.

Between 2008 and 2010 a rapid increase occurred in typical ΔT for new IT equipment, climbing over 12°F (7°C) in just one or two generations of equipment. At that point, ΔT ranged from just under 30°F (16°C) to over 50°F (28°C), based on the type, size, and utilization of equipment. Since then, the lower range has steadily increased.¹

But at 50°F (28°C) ΔT , the upper range of this trend has been pushing the limit of what constitutes acceptable exhaust air temperatures, with exhaust temperatures reaching 140°F (60°C)! The concerns at this temperature include touch safety, certification of ancillary equipment such as PDUs and switches, and general human comfort. As such, the upper limit of ΔT is not expected to increase any further.

As mentioned, increasing ΔT corresponds to decreasing airflow requirements per unit of heat. Until 2010, the increase in ΔT meant less airflow was needed for the same load, which allowed server manufacturers to actually increase the amount of heat that could be dissipated by equipment without necessitating major changes to existing facilities. However, as equipment nears 50°F (28°C) ΔT , the amount of airflow levels off to a minimum of ~60 cfm/kW (~100 m³/kWh). Beyond this point, simply put, more airflow will be required to remove additional heat.

Airflow

If IT equipment power consumption and density continue to increase while near or at the maximum practical ΔT , then airflow must continue to increase (or another cooling method must be used, such as liquid cooling). Depending on the flexibility and future proofing of existing data centers, this may or may not necessitate major changes during future IT refreshes.

It has been commonplace for a long time to adopt hot-aisle, cold-aisle layouts within the data center. This practice concentrates the exhaust air for more efficient removal while simultaneously reducing inefficient mixing of supply and return air. To supply the air, many data centers use raised floor plenums to distribute the supply air across the data center floor, with perforated tiles in the cold-aisles.

However, this standard layout typically allows for, at most, a single perforated tile per rack, creating a possible constraint on the maximum amount of airflow that can be supplied. Based on standard raised floor pressures of 0.03 in. w.c. to 0.05 in. w.c. (7.5 Pa to 12.5 Pa), perforated tiles (~25% free area) could supply up to 500 cfm (850 m³/h) while very open grates (~60% free area) could supply up to 1,800 cfm (3060 m³/h). This range already exceeds what may be required for a full 42U rack.

Attempting to simply raise the underfloor pressure to increase airflow may not be effective, due to increased leakage and unexpected pressure effects where the velocity of supply air under the floor is high. For example, perforated tiles near computer room air handler (CRAH) units may actually experience reverse airflow due to low underfloor static pressure according to the Bernoulli principle. High

pressures also increase the typical air velocity exiting the raised floor, and can cause air to overshoot the servers in the bottom of the rack and increase mixing with the return air.

A few potential solutions exist to overcome this restriction. The cold aisle could be widened to allow two perforated tiles per rack, but this would require significantly more floor space, leading to a reduction in the number of racks that could fit into the data center. Alternatively, active tiles could be used that contain fans and can increase airflow up to 3,000 cfm (5100 m³/h) or more. It turns out, though, that these are best used sparingly for spot cooling as opposed to increasing airflow throughout the data center. Unless enough supply air is provided, widespread use of these tiles could lower the pressure under the raised floor causing more harm than good.

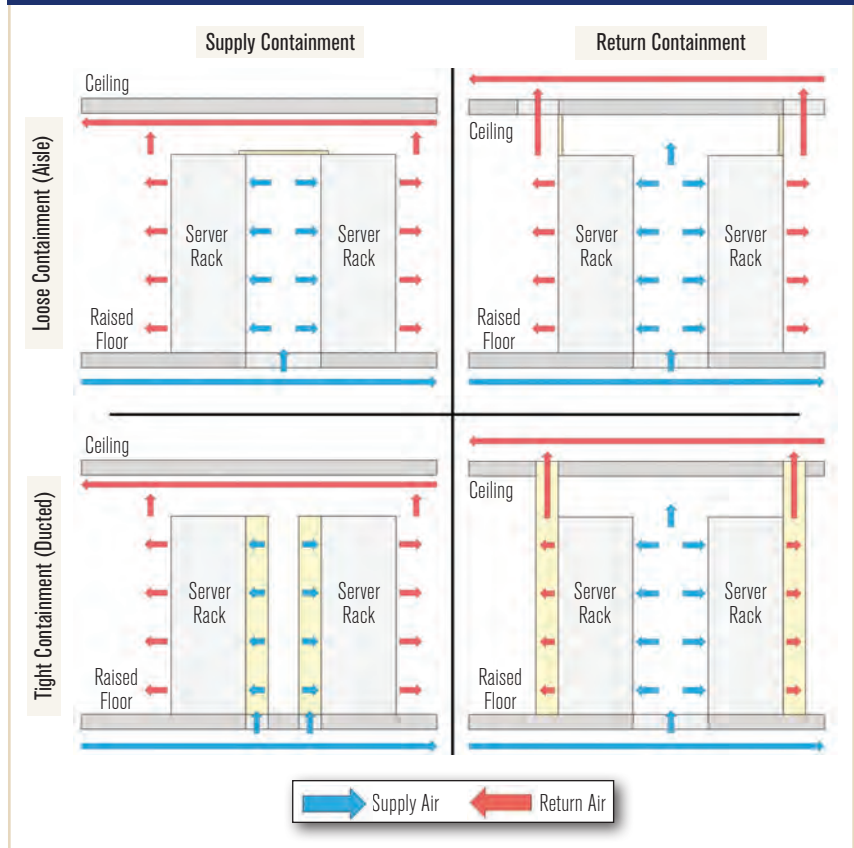
Directly Coupled Air Paths

If the IT equipment requirements are in conflict with the amount of air that can be reasonably supplied in a typical manner, it may be advisable to consider directly coupled cooling solutions in which supply or return air is entirely contained. By containing one of the airstreams, pressures can be increased without leakage or mixing, making it easier to supply additional air.

For direct coupling, either the supply air or return air can be contained. Further, containment can be as large as the entire aisle or smaller, such as a contained chimney for the hot exhaust, or even smaller still by limiting the containment to the rack itself, as is the case with rear door heat exchangers. Figure 1 shows a graphical matrix representing possible designs that answer these two questions.

The size of the containment will typically be inversely related to the pressure created by flow resistance inside it. For very large containment volumes, there may be a

FIGURE 1 Directly coupled design examples using a raised floor supply.



negligible to minor increase in resistance that must be overcome. For smaller containment volumes such as capturing the hot exhaust and ducting it into a return plenum, the resistance due to pressure may become significant unless there is active assistance increasing the flow. The increased pressure also can achieve higher cooling capacities, thus higher rack densities.

Modern servers can adjust to external pressures by appropriately and continuously speeding up or slowing down fans, assuming there is enough margin to do so. Limitations of the IT equipment must be considered though; if servers are running at their maximum utilization at maximum inlet temperature, the fans may already be running at their design maximum. Additionally, other IT equipment such as legacy servers, storage equipment, and switches may have more basic thermal controls and fail to adequately adapt to external pressures.

Data Center Transients

Within the data center, environmental factors may vary frequently and at different rates. On one

end of the spectrum we have changes to air temperatures, which can change quite significantly and rapidly due to their minimal thermal mass and rapid flow through the data center. If a rack of IT equipment is turned on, then the air temperature at the back of that rack will increase almost immediately.

At the other end, the temperature of the physical data center (walls, floor, etc.) may change the slowest since they have very large thermal masses and do not produce heat themselves. The IT and facility equipment most likely land somewhere in between, as they have large thermal masses but the potential to produce or exchange great quantities of heat.

There are two main things to consider related to the capacity of the data center environment to change. The first is IT equipment specifications that often specify a maximal rate of temperature change per some timeframe. *Thermal Guidelines for Data Processing Environments* states a maximal rate of 9°F (5°C) per hour for tape and 9°F (5°C) per 15 minutes for all other IT equipment.² The thermal mass of IT equipment is often enough to alleviate concerns that the equipment would change temperature too rapidly due to a normal change in environmental conditions.

The second consideration is related to failure modes of the data center and their effect on the equipment. For example, a loss of CRAH fan power could have a rapid impact as return air ceases to be cooled, whereas a loss of pumps to the CRAH would have a slower impact as the return air could continue to transfer heat to the CRAH coils for a short period of time.

All modes of failure should be properly evaluated for their impact to the data center. Based on the design and failure, factors such as the thermal mass along the airflow path, possible mixing of supply and return air, and control schemes may greatly affect how quickly temperatures rise in the data center. This information is critical to creating the right data center response plans in the event of failures.

Some Other Interrelated Considerations

In addition to air temperature-related concerns, many other aspects of data center design and performance need to be considered when making any

changes to the data center. For example, the potential for electrostatic discharge may increase when humidity is very low, but unnecessary humidification wastes power.

Humidity also plays a role regarding air contaminants; corrosive gaseous factors are generally more dangerous when humidity is high.³ Therefore, design humidity may affect what sort of air filtration is required.

Particulate air filtration is another factor that can have large-reaching effects, since fouling of the airstreams on even tiny components within servers can affect the thermal performance of servers throughout the data center. On the other hand, unnecessary air filtration will certainly waste energy.

Even acoustics and weight should be considered. Increasing temperatures throughout the data center could increase server fan speed to the point where acoustical specifications are being approached. Similarly, increasing server density could mean increasing the load on a raised floor beyond its design point, necessitating additional structural supports and creating further airflow obstructions under the floors.

Conclusion

Throughout the life of the data center, changes to both the IT and facilities equipment will almost inevitably occur on a regular basis. However, inherent risk exists in making changes without considering all the effects it might have downstream. One seemingly minor change to one aspect might ultimately cause a massive system failure elsewhere.

Designing for change is a wise strategy, but even observing current trends in IT and technology may not be enough to predict what changes may be required a few years from now. Therefore, potentially the best weapon against future risk is a solid understanding and appreciation for the interrelated nature of systems throughout the data center.

References

1. ASHRAE. 2016. *IT Equipment Design Impact on Data Center Solutions*. Atlanta: ASHRAE.
2. ASHRAE. 2015. *Thermal Guidelines For Data Processing Environments, 4th Edition*. Atlanta: ASHRAE.
3. ASHRAE. 2009. *Particulate and Gaseous Contamination in Datacom Environments*. Atlanta: ASHRAE. ■