

This article was published in ASHRAE Journal, September 2017. Copyright 2017 ASHRAE. Posted at www.ashrae.org. This article may not be copied and/or distributed electronically or in paper form without permission of ASHRAE. For more information about ASHRAE Journal, visit www.ashrae.org.

Part 4: Future Disruptive ITE and Paradigms

Changing Landscape Of Data Centers

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

Data centers are about to undergo yet another major disruption in the evolution of the ITE. This next disruption may be more about the monumental shifts in software and data demands, but there is also the potential for another step function in the computer processors.

This may sound familiar to those who lived through the bipolar to CMOS transition, or to those who have recently experienced the step functions in power density within existing data centers, which was more a by-product of the lagging technology refresh rates occurring at a steeper part of the changes in the ASHRAE power trend curves shown in *Figure 1*.¹

This next disruption is likely to be much more apparent. The “occupant,” or software, is currently getting a significant boost from artificial intelligence and its application to everything. Until recently, the data generated in data centers was largely the by-product of humans. However, now software exists that writes itself and is capable of generating enormous amounts of data; far more significant than humans could generate alone.

To keep pace with this extreme growth rate in software, the hardware must yet again evolve. In both cases, this means the HVAC systems must also evolve. By understanding the IT hardware and software, its changes, and how it interacts with the data center HVAC, data center designers will be more equipped to handle this changing landscape.

This final article in this series will recap the previous ITE articles, extrapolate to the future ITE, and summarize the significance for data center HVAC designers.

Changing Landscape of Data Centers Recap

Part 1 in this series explored the current state of IT equipment in the data center, covering the different

equipment types, their roles, and their general arrangement within the data center. In generic terms the ITE is divided into these basic functions:

- Servers: data computing and processing;
- Networking: data transmission; and
- Storage: data storing.

Today’s ITE consists of several very distinct types of hardware, each with their own unique designs and functions for supporting different types of data and services. Each of these types of ITE will correspondingly interact with the data center infrastructure in different ways.

Not all data centers will have the same composition of these pieces, but that composition does matter for HVAC design. By understanding the fundamental roles of IT hardware (servers, networking, and storage), its changes, and how it interacts with the data center HVAC, data center designers are more equipped to handle the changing landscape.

Part 2 in this series explored the ITE thermal design, the major heat generating components, and how those components work collaboratively amongst one another and interactively with the HVAC infrastructure.

From the smallest component, to the largest ITE chassis, thermal design plays an incredibly important role and influences design throughout the data center industry. It is not a straightforward problem of keeping

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.

equipment cool, but rather an optimization of keeping equipment within acceptable temperatures while enabling the right blend of performance and efficiency.

For the HVAC design engineer, the most important focus is providing the right “entering” thermal conditions (air or liquid) to the ITE. For the ITE manufacturers it’s an optimization problem involving a lot of different pieces of equipment and trade-offs.

Part 3 in this series focused on the interactions between IT equipment and facility (addressing changes to ITE design and use, and how they related to the DC).

There is an inherent mismatch of time scales of thermal responses 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 to where larger responses are necessary.

Further, there are many interactions between the IT equipment and the data center facility systems, not just 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 of the equipment is in a constant state of reacting to the reactions of other equipment.

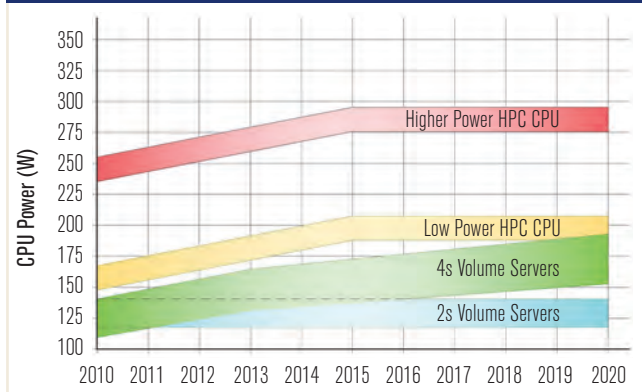
This highlights the importance of understanding and designing for these interactions in ways that yield predictable and/or addressable results.

Data Center HVAC Design Recap

In the beginning of data centers, ITE loads were treated as block-loads in the space with uniformity on that load. This meant that the HVAC cooled and managed that load much like a lighting load.

As industry awareness increased through publications, it became more obvious that cooling that space was not the solution, but rather the intakes on the ITE. As we’ve learned from the previous section, the thermal management inside the ITE is addressed by the ITE

FIGURE 1 Increasing CPU power consumption ranges by application & year of production.



manufacturer design.

Still, in more recent times, industry awareness has increased again to recognize that those loads are not uniform in the space. This prompted a close review of airflow management and the introduction of various airflow metrics. Those metrics revealed the necessity for various forms of aisle containment and that sometimes the total HVAC required in a space would be driven by airflow, not cooling capacity.²

To further complicate the HVAC design, the configuration of the hardware, as deployed by the end users, can vary significantly from rack-to-rack and row-to-row. This makes the proportion of distributed load across the floor extremely variable from 0% to 100%.

Today, there are ITE applications where the amount of compute power varies largely as a product of the actual compute workload. The servers at idle can have a very different load profile than those same servers at 100% CPU utilization.

Due to the lag in the technology refresh rates and a combination of the items noted previously, many data centers failed to see the rapidly increasing power density rates, in parallel with the industry forecasts in the ASHRAE’s *Datacom Equipment Power Trends and Cooling Applications, Second Edition*. Yet, ASHRAE TC 9.9 recently confirmed that those industry peak power trend forecasts have proven highly accurate.

Without a complete picture of all of these items there’s potential for a false sense of security regarding the trajectories of the future utilizations in many data centers. The next section will highlight how some of the current trends may significantly disrupt the current utilizations and hence the HVAC design in many data centers.

New Disruptive Technologies and Their Impact on IT Hardware and Software

Disruptive software is where it all begins. After all, software is the true occupant of the data center, despite that most of it is very transient, coming and leaving the data center at the speed of light. Software is evolving more rapidly than ever before and most recently it's receiving a significant boost from itself (without human intervention).

As we've seen with historical trends, data generation and consumption continues at an exponential rate globally. Nowhere has this been more apparent than in the social media space and the rate of increase of data usage in "An Internet Minute."³

Beginning with new forms of software developed by humans, emerging platforms such as cryptocurrencies are fundamentally driving significant changes in our approach to the electronic exchanges of currency, goods, and services.

The most prominent of them, Bitcoin, has skyrocketed in value across the last few years, creating a computing

"gold rush" to mine the currency—an effort that is estimated to collectively consume around 40 GWh per day.⁴ Importantly, these efforts have rapidly increased from individual machines to large-scale clusters of highly specialized GPU platforms, powered and cooled in traditional data center spaces.

Perhaps the most disruptive software of all, artificial intelligence (AI), has emerged on the global stage and is sweeping across every industry. Unlike other forms of software, various forms of AI are self-replicating, self-developing, self-healing, self-producing, and self-extrapolating. As a result, AI is capable of generating quantities of data that were unheard of with prior forms of software that required constant human management and interface.

AI is taking on numerous applications and spreading across all industries at an accelerating rate. As an example, Microsoft Tay, an AI chatterbot released on Twitter in March of 2016, was capable of producing over 95,000 Tweets, 35,000 direct messages, and a lot of internet traffic in under 24 hours without any human

Advertisement formerly in this space.

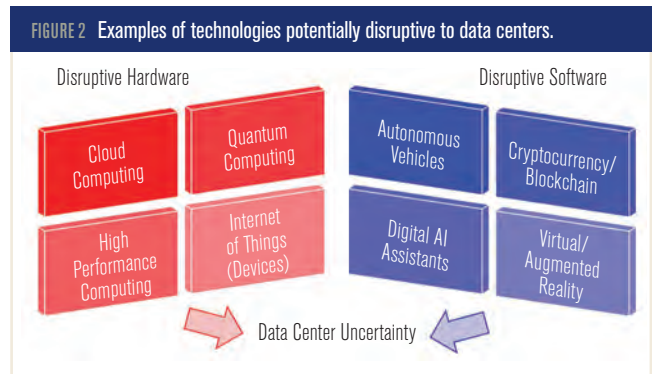
Advertisement formerly in this space.

intervention on the software side. Tay (later succeeded by Zo) was similar to or based on Xiaoice, a similar Microsoft project in China. It's reported that since late 2014 Xiaoice had had "more than 40 million conversations apparently without major incident."⁵

Another example of "AI of everything" is autonomous vehicles. Tesla Motors has been leading the charge, debuting their Autopilot system in 2014, which was capable of auto-steering, traffic-aware speed control, and driver-directed lane changes. Autopilot uses deep-learning techniques, and Tesla boasted that the system had been in operation for more than 300 million miles prior to 2017.

Now on the second generation of hardware, each vehicle's eight cameras, 12 ultrasonic sensors, and radar produce a large amount of data that is processed in real-time by NVidia's Drive PX 2 platform. This represents a truly massive amount of data being generated each day, although it is disaggregated in this case.

At the same time that these new software platforms are proliferating across multiple industries, creating,



processing and storing almost inconceivable amounts of data, the potential for disruptive hardware advancements loom on the horizon. Advancements in high performance computing (HPC) and the possibilities of quantum computing are game changers that many businesses may not see coming.

The increasing performance and availability of high performance computing is driving much advancement that historically may have required expensive and highly specialized supercomputers, which very few companies

Advertisement formerly in this space.

Advertisement formerly in this space.

had the means to use. Nowadays, there are cloud-based platforms with intense computing capabilities that can be accessed by any individual with an email address and credit card.

Not only is HPC more available and the physical hardware more powerful than ever before, but new system architectures may also lead to rapid performance improvements. One such HPC platform is putting shared memory at the center of the system architecture, as opposed to CPU, enabled by optical interconnects. Fundamental changes to system architectures could have major implications for the data center's layout and infrastructure.

On the other hand, Quantum computing offers the potential to be revolutionary not because of performance improvements, but because of what types of computation it might enable that would be otherwise out of reach. Instead of a binary system made up of bits of data that are either a 1 or 0 at any given time, quantum systems use quantum bits that can be in multiple states at once.

Theoretically, these systems would be able to solve classes of problems very quickly that are very computationally intensive on classical computers. The most common example is integer factorization, which would allow for easy decryption of some cryptographic systems commonly used today.

Even though quantum computing is in its absolute infancy, the implications for HVAC requirements are known to be interesting since most quantum effects require superconducting components to be cooled to within fractions of degrees above absolute zero. Such an environment is so highly specialized that the Quantum Data Center may look nothing like any traditional data center today.

Figure 2 shows just a few of the disruptive software and hardware developments that are ongoing or on the horizon that may together dramatically affect the data center, both in its design and operation.

Closing Comments

Data centers and their ITE have been changing radically since its inception. Those changes are on an increasing exponential scale. Historical disruptions have included step-functions in the CPU type, exponential growth in CPU power consumption and continued optimization of the ITE to thermodynamic extremes.

Data center HVAC designers have to consider:

- Hardware
 - Type of hardware, which generations are deployed, disaggregation of the types used, hardware refresh rates, utilization and occupancy of racks, etc.
 - Thermal considerations such as; power trends, differential temperature (or ΔT), acceptable entering conditions (air or water), thermal limitations of various elements, disaggregation of loads, etc.
- Software
 - Types of software used, utilization trends and its impact on hardware utilizations, growth trends, etc.
- Interdependencies
 - Airflow management between ITE and HVAC
 - Thermal responses, sensitivities of systems, and time scale mismatches.

It is reasonable to believe that the IT industry is currently undergoing more rapid change than almost anything it has experienced. Not only is data within existing segments being created, stored, and processed more quickly than ever before, but entirely new IT segments are emerging, such as blockchain, virtual reality (VR), and self-driving vehicles. These changes are placing increased demands on data centers and the IT equipment in them.

IT software and hardware are refreshing at an accelerated rate, while MEP infrastructure and design has largely remained the same and is much slower. This mismatch leads to incompatibilities with matching loads to capacities and even more importantly responses to those loads.

The trends are clear though, future technologies are resulting in increased unpredictable advances in the software growth and hardware evolutions to keep pace with it. The dynamic responses and rate of changes with ITE will demand an increased importance on future-proofing the HVAC designs in data centers.

References

1. *Datacom Equipment Power Trends and Cooling Applications*, Second Edition, 2015.
2. Beaty, D., D. Quirk, J. Jaworski. "Changing landscape of data centers part 3: interactions between IT systems and data centers." *ASHRAE Journal* 59(7).
3. Quirk, D., B. Durham. 2015. *Social Media Data Centers, Telco's, and Net Neutrality Battles*. *The DataCenter Journal*.
4. Digiconomist. 2017. "Bitcoin Energy Consumption Index." <http://digiconomist.net/bitcoin-energy-consumption>, accessed 7/17/2017.
5. Wikipedia. 2016. Tay(bot). [https://en.wikipedia.org/wiki/Tay_\(bot\)](https://en.wikipedia.org/wiki/Tay_(bot)), accessed 7/17/2017. ■