

Part 1

Fog Computing

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

There's a deep fog rolling in across the data center industry as distributed computing further muddies the waters of the old central computing model. Traditional data centers have been defined by a structure with four walls and a roof over it. They can be brick and mortar, modular, containerized, or combinations of these. This has been the long-standing industry location and model for data storage and computing, which can be thought of as the “grounded” or “earth-bound” version of centralized data centers.

In recent years, computing has migrated further away from a centralized model to a distributed model through “edge” data centers and various forms of colocation data centers. The distributed computing model attempts to put more of the data storage and computing power closer to the end use in what we now refer to as “the cloud.”

Taken a step further, the Internet of Things (IoT) continues to push more compute and storage capabilities into the end devices themselves. These devices include traditional known items like smartphones and personal computers, but more recently, less-known devices like autonomous vehicles, home appliances, smart-lighting, meters, personal robotic assistants, and digital displays. This is where things get a bit “foggy” because the lines between the traditional “grounded” model and the “cloud” model have been blurred.

“Fog computing,” as it's become known, is the wider distribution of end-user clients or near-user edge devices that carry out a substantial portion of storage, compute, and communication, rather than

these functions being performed at a centralized location.

Fog computing has made the predicting and planning of centralized and distributed computing, storage, and networking much more challenging, hence the fogging conditions. This column is Part 1 of a series that introduces the concept of fog computing and the evolution of the data center leading up to this condition.

Grounded (Centralized) Data Centers

Brick and mortar data centers are the staple of the data center industry and what most MEP engineers think of when the term “data center” is mentioned.

Original design of these structures involved largely known boundary conditions:

- Known code requirements for building construction;
- Defined spatial parameters;
- Defined power density;

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.

- Defined cadence for the ITE (aisle/equipment spacing);
- Defined air management approach; and
- Defined PUE target.

Further, industry boundary conditions have been published in ASHRAE's *Datacom Equipment Power Trends and Cooling Applications*, 2nd ed.¹

These boundary conditions led to well-defined HVAC requirements and predictable performance for a given geographical location.

Brick and mortar data centers have remained popular for many of the industry giants and corporate enterprise alike because it translates into direct control. Some benefits to having the ability to control include:

- Physical assets, such as:
 - The quality, reliability, optimization, and maintenance of the infrastructure, and
 - The stored data itself; its security, analysis, integrity, etc.;
- Scaling ability; and
- Logical, physical, and cyber-security.

Today's data center and telecommunication models are "centered" on the theme of centralization, known as centralized computing. Enterprise, governmental, telecom, and internet platforms all have centralized data centers to some scale. The consolidation of hardware, software, points of presence, etc., simplified the cost models, regulatory compliance, staffing, cybersecurity, data analysis, and so on.

What is also characteristic of these data centers is their long lead times to budget, plan, procurement, design, approvals, construct, commission, and outfit with ITE for production and operation. As more traditional procurement models have dominated and commoditization of everything has occurred, it has taken even more time for these traditional data center construction approaches to go through these steps. This is particularly true for the budgeting, planning, procurement, and approval phases.

Further, many publicly traded companies prefer to reduce their capital expenditures as they mature and take the hits on the operating side of the balance sheet.

To work around some of these challenges and others we've talked about in prior columns, the industry emerged with new models of off-site premise hosting or "cloud" data centers.

Cloud Computing

The data center terms, such as cloud computing, are derived from terms used to describe the internet and other private network architectures. All forms of cloud are essentially "remote" hosting, computing, and/or storage of data from the owned and occupied premises of a given owner, whether public or private forms of it.

Cloud computing involves a networked aggregation of hardware and software that is abstract or "clouded" from the end-user.

The context of cloud computing is the attempt to marry information technology (IT) strategies and overall business strategies. The rapid advancements of the former combined with the infinite variability of the latter make for a myriad of possibilities of how to achieve an aligned outcome. This is the driver behind why there are so many variants for the types and models of cloud computing.²

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

The cooling challenges of a cloud computing facility involve being adaptive and flexible enough to cope with an unguaranteed, limitless, capacity on-demand load-profile while still being energy efficient and resilient to scheduled and unscheduled outages (e.g., equipment maintenance, utility interruptions, etc.).

Cloud computing is a rapidly evolving industry. The cloud's need to be flexible and adaptable is driven by technology advancements and highly variable business needs that are expected to be unlimited and on-demand but with little to no risk, guarantee, or commitment on the part of the customer.

As a result, the data center facilities that house cloud computing infrastructure are subject to highly volatile power and cooling loads in terms of their magnitude and profile. This compounds the challenge for data center designs to not only maintain their 24/7 availability, but to do so for loads that can significantly fluctuate while still attempting to be energy efficient and not capital or operating cost prohibitive.

The diversity in cloud models brings about different challenges for the associated infrastructure. With enterprise data centers or private cloud, the growth model may be known for the company managing the data center infrastructure. However, some other models of

public, hybrid, or community cloud often create unpredictable growth requirements because the services are driven by many different businesses operating in this cloud environment outside the control or visibility of the ones managing the data center.

The use of the services in the cloud can be highly variable, resulting in more variable load swings within the data center. This results in the need for extremely scalable infrastructure.

Despite all the variations of cloud, one thing is consistent; they still largely take the form of “brick and mortar” data centers, including all the variations of modular, container, etc., construction.

The boundary conditions of cloud data centers vary slightly from the centralized “grounded” model in that optimization and scaling are more challenging and unknown since the data center owner has little control over how their customers use the space and power they purchase. Ironically, the “cloud” serves dual meaning here and also represents the nature of the unknowns (clouds constantly change form) for

scaling and optimization of the infrastructure.

Edge Computing

While much of the data center industry is in the midst of “mega” consolidation of their data centers, one could argue they are moving in the wrong direction. Data analytics and knowledge generation at the source of data, otherwise known as edge computing, is the next frontier.

It is true there are strong arguments for centralized computing and data storage, yet distributed edge computing is emerging as a better means of managing the growing number of connections, data development, and data analysis as networks of all types are stretched to their capacities.

As the term implies, edge computing shifts the “computing” functions to the edge of the telecommunications and IT networks. It covers a wide range of technologies including this small sample:

- Mobile technologies such as mobile edge computing, cloud-based radio access, mobile data acquisition, and mobile signature analysis;
- Wireless sensor networks;

Advertisement formerly in this space.

- Fog computing;
- Distributed data storage and retrieval;
- Remote cloud services; and
- Augmented reality.

The edge computing trend is being fueled by many other important trends, including the explosion of edge devices or sensors (Internet of Things), analytics of consumer data as sources of targeted advertising, and limitations on network bandwidths, among other things.

In the past, data centers were used primarily to support internal functions, now they are being used to support end-user applications and content. Companies are shifting to obtain more revenue from those applications and content through advertising, in particular, targeted advertising.

Edge computing essentially moves the “brains” of computing where less data needs to be moved across telecommunication networks or internet service provider (ISP) networks. ISP’s and telecom providers are increasingly under pressure to expand the capacity of their networks to manage the exponential growth in data demands. One way to solve this problem is to architect applications and software solutions with edge computing capabilities to reduce the amount of network traffic.

Perhaps one of the best examples of edge computing is the smartphone in the palm of your hand. The smartphone is one example of the mobile edge computing (MEC) trend.

Edge computing translates into more agility for future data center growth. Increasingly, companies are now beginning to realize the potential competitive advantage obtained through increased use of edge-computing strategies. They are also beginning to realize there are significant cost savings available through a divergence from centralized data management.

For data center consultants, this means looking at future-proofing concepts with more dimensions. Those dimensions need to no longer only consider increased capacity but the higher potential of stranded assets and/or the need for more dynamic exit strategies.

Today’s designs are primarily singularly focused on future increases in loads and space needs. However, as we’ve seen with the introduction of virtualization, some disruptions can result in the reverse trend. Edge computing is another such trend that aims to diversify the data center across more assets in smaller segments.

There is a link between virtualization, IoT, cloud, Software Defined Networks (SDN), and edge computing. While SDN is focused on data centers and the core

network, over time there is potential for the principles of rapid and automated reconfiguration of network resources to be applied more widely and closer to the edge. A distributed computing resource would become part of this paradigm and that resource now has a name: “fog computing.”

Further details on edge computing can be read in the past column, *Exploring Edge Computing*.⁴

Fog Computing

The industry is pushing the edges of the centralized (ground) and cloud models as they seek to gain more and more data on everything the consumer and its associated devices are doing. This race to optimize everything and gain data insights to everything has created a “fog” in between.

Fog computing has been considered a virtualized platform that is typically located between end-user devices (smart objects and sensors) and the cloud data centers.

Fog computing has been driven by a host of factors:

- Increasing compute capabilities in small form-factors;
- Intermittent connectivity with the cloud (interference, remote, etc.);
- Insufficient network bandwidth and increased latency (software applications are outpacing network advances and deployments);
- The need for device-to-device communications for more real-time applications and machine-to-machine (M2M) communications; and
- An explosion of smart-devices and small sensors for everything.

Fog computing has been labeled many things, but perhaps it’s single greatest purpose is to enable “Aware Machine-to-Machine Communication.” A way to think of this is sensors in virtually everything. In other words, everything becomes digital and has a digital signature and data associated with it.

There needs to be a means of network communication amongst devices that is not dependent upon bogging down the same network that humans use because machines will communicate orders of magnitude greater. Machine-to-machine communications generate data at different velocity, veracity, and volumes than humans.

As stated in the last column, *Changing Landscape of Data Centers, Part 4: Future Disruptive ITE and Paradigms*, data centers are about to undergo yet another major disruption. This next disruption may be more about the monumental shifts in software and data demands.⁵

Closing Comments

Just when the industry thought things had cleared to a predictable model in the data center industry, the fog settles in and blurred the lines between the traditional “grounded brick and mortar” data centers and the “cloud” data center models.

The centralized model meant that data center consultants had a single large, predictable target to focus on. Edge computing changes the game entirely and distributes the data center everywhere and anywhere. The definition of the data center will need to be revisited as its location, size, and loads will no longer be as predictable.

Edge computing emerged to solve a host of problems associated with the ground and cloud models, but it's not the final frontier. Solutions like mobile edge computing, wireless mesh networks, remote cloud services, machine-to-machine, and a host of other services aim to further optimize data collection, compute, storage, and transmission.

Fog computing aims to further distribute software

services to the end devices. This includes both mobile and stationary assets like smart appliances, autonomous vehicles, personal AI assistants, virtual and augmented reality wear, drones, M2M sensors, and even robotics. We'll further explore fog computing and the IoT evolution it supports in the next column.

Rapid changes will continue in the data center industry, but all indications are that the new norm will be the hybrid that is a combination of centralized computing (cloud) and distributed computing (edge).

References

1. ASHRAE. 2012. *Datacom Equipment Power Trends and Cooling Applications, Second Edition*.
2. Beaty, D. 2013. “Cloud Computing 101.” *ASHRAE Journal* (10).
3. Quirk, D., B. Durham. 2015. “Social media data centers, telco's, and net neutrality battles.” *Data Center Journal* (2).
4. Beaty, D., D. Quirk. 2015. “Exploring edge computing.” *ASHRAE Journal* (12).
5. Beaty, D., D. Quirk, J. Jaworski. 2017. “Changing landscape of data centers, part 4: future disruptive ITE and paradigms.” *ASHRAE Journal* (9). ■

Advertisement formerly in this space.