

Unpacking Data Center Loads

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

At its core, data centers are governed by the interdependent relationship between facilities, hardware, and software (the Data Centers Trinity). A given data center's ability to scale, react, change, and adopt new practices and a new business model is largely constrained by its own inner dynamics between the Data Center Trinity. But, most designers invest little time understanding the hardware and software, which can lead to data centers underperforming or failing. This column provides both data center novices and experts some sense for the impact hardware and software has on data center design. And, it highlights the importance of software's current and future full load profile in a data center.

The Data Centers Trinity is filled with challenges, beginning with the disparate gaps in terminology. Since the three parts of the industry do not speak the same language, there is inefficient communication, which slows the industry's ability for agility and change.¹

Another challenge is the gross mismatch of life cycles between facilities, hardware and software. Infrastructure (facilities) life cycles are measured in decades, while hardware is measured in years. Yet, software change is measured in months, weeks, even days.²

Further, data center business models are extremely diverse. Methods of monetization, regulatory policies, industries supported, contractual arrangements, and the software itself shape the underlying business models of each portion of the data center industry.

Both the challenge and the opportunity is that the data center has to be affordably designed for not just today's requirements, but for the range of possible future requirements (e.g., in five years).

Over the past decade, data centers have also become the poster child for energy conservation, with many opinions on the best methods to save energy.³

By only considering facilities, hardware, and software domains as silos and ignoring facilities and hardware, the risk of suboptimal results, if not actual failure, is

high. To be successful, today's data center designers must invest heavily in understanding and remaining current with the IT hardware and software domains.

To this point, software is the real load in a data center, and it's currently undergoing many new rounds of disruptive change that directly impact the loads data center designers have to manage to. Critical to any power and cooling system design is the load. It is no longer adequate to simply consider maximum load, but rather a full load profile including both transient and steady-state conditions.

Consequently, ASHRAE TC 9.9 pushes hard to publish current important content and expand the coverage of hardware and software (*Figure 1*) because is critical for today's data center designer to be proficient in their knowledge and use of TC 9.9 content, in addition to investing heavily in gathering content from other IT hardware and software sources.

The full load profile (current and future) of a data center's primary occupant, software, is a very complicated and important topic since it directly impacts construction cost, energy use, and operating cost. One should

Donald L. Beaty, P.E., is president and David Quirk, P.E., is vice president of DLB Associates Consulting Engineers, in Eatontown, N.J. Beaty is publications chair and Quirk is the chair of ASHRAE TC 9.9.

either personally obtain this knowledge or use a subject matter expert (SME).

Background

The dominant direct load in a data center is the hardware and software (not climate). That load consists predominantly of IT racks with servers, routers, etc. Therefore, it is important to understand the maximum load at maximum configuration and lowest load at idle (including the various levels of idle or no load condition). How the hardware responds and what efficiencies it operates at based on load, from min to max, is also important.

The actual work performed by hardware is essentially a resource service to software. The software is not physical and essentially represents virtual load on the hardware.

The number of people in companies creating and changing software is somewhat beyond belief. Essentially, anyone with a computer has the potential for creating a small app or piece of software. Since the Internet is so open, people have the unlimited opportunity to search for software. Also, end users see things that they like and do not like, which spurs additional ideas and changes. These changes can happen daily, even hourly.

The data center industry has not only come up with a way of virtualizing hardware (e.g., one physical server becoming multiple virtual servers) it has now virtualized software workload and termed it software workload containerization. The use of the word container is valuable and impactful. There is a direct parallel to shipping containers in that there is no need to know what is inside the container to ship or receive it. The same is true with a software workload container.

All this combined creates a real challenge for data center designers. They are being pressured to deliver services and product at a lower and lower cost, forcing a general trend toward standardization and modularity.

While there is benefit in modular products and standardization, the challenge is that it is also important to invest sufficient time to understand the workload. Without this investment, the design is at risk of underperforming.

Unfortunately, the fees and margin do not justify the amount of effort required to understand the loads to this level of detail, since it also requires obtaining and maintaining a good working knowledge of general hardware and software trends.

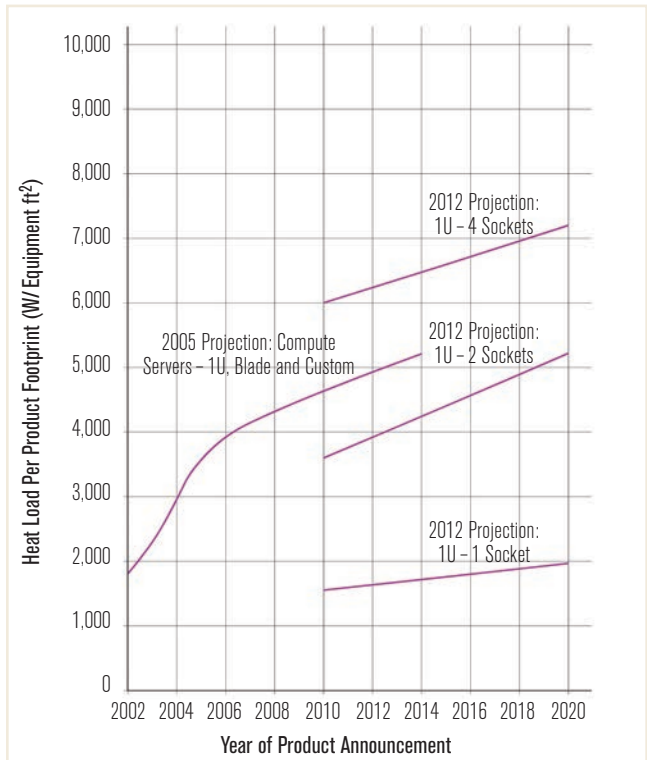


FIGURE 1 Datacom power trends chart from the second edition of *Datacom Equipment Power Trends and Cooling Applications*.

Facilities Domain's Impact

A typical data center designer begins with common owner defined requirements such as power density (W/ft² or kW/rack), redundancy (extra standby systems), resiliency (defining failure domains) and IT equipment intake temperatures.

This terminology is “commonly” understood by data center designers today and the underlying assumptions commonly made with each of these are as follows:

Power Density (Watts per Square Foot or kW/Rack)

- The actual power requirements of the IT equipment will average out to some uniform density across the data center.
- Once the IT equipment is plugged in, the power does not fluctuate much over a given day.
- The power density is based on the nameplate information the owner has provided.

Redundancy (Extra Standby Systems)

- The availability of the data center is wholly dependent upon the infrastructure redundancies.
- There are no interdependencies between the infrastructure, hardware, and software redundancies, resiliencies, and diversities.

- More is always better.

Resiliency (Defining Failure Domains)

- Tradeoff between probability of an outage and its impact granularity.
- Bounding how much capacity or “mission” a specific type of outage will take down (if any).
- How does the data center “flex” when faced with a specific type of outage?

IT Equipment Intake Temperatures

- The equipment can only operate within the Recommended Range as published by TC 9.9.
- Constant intake temperatures are required.
- All IT hardware has the same requirements for intake temperatures.

However, because most designers have invested little time in understanding the IT hardware and software, these assumptions are rarely correct in today’s data centers.

Hardware Domain’s Impact

IT hardware development is driven by software needs and the thermal limitations of silicon. The hardware can be air-cooled or liquid-cooled. In either case, it is important to understand the interface points and syntax of the hardware to provide the proper infrastructure to support it.

Some key hardware domain terminology includes processors or central processor unit (CPU), memory, storage, I/Os and power supplies.

At the motherboard level, the quantity of each component and their relationship to one another has a large impact on the total load and the range of loads that a server can produce. In addition, the thermal and power management software applications running on those servers can further impact that potential range. Each form of optimization can result in a different total power level and range of power levels, depending on the data center application and the software objectives being sought.

For the data center designer, it’s also important to focus on the IT hardware interfaces such as temperature, humidity, air quality and cooling flow (air or liquid).

Details of each of these interface points can be found in the *2015 ASHRAE Handbook—HVAC Applications*, Chapter

19, Data Centers and Telecommunications Facilities, or in the TC 9.9 book *Thermal Guidelines for Data Processing Environments, Third Edition*, which publishes the acceptable environmental guidelines for entering conditions on air-cooled and liquid-cooled equipment.

ASHRAE TC 9.9’s *Datacom Equipment Power Trends & Cooling Applications, Second Edition*, forecasts the past, present and future trends of heat release in watts for various IT equipment, unlocking the potential for implementing affordable provisions for future IT loads (Figure 1). A blind-side for data center designers would be to only consider the single dimension metric of W/ft² or kW/rack.

The items above are perhaps the better known topics within the hardware domain. What is not well understood are the approaches manufacturers use to optimize the hardware performance, longevity,

reliability, energy efficiency, costs, etc. The thermal design of hardware continues to evolve within the boundary conditions defined by these items (Figure 2). Recently this has led to extensive power and thermal management solutions.

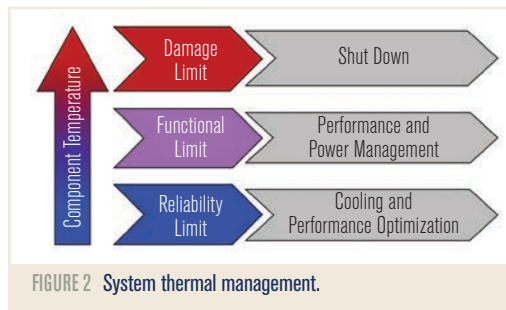
IT Equipment Thermal Management and Controls, a TC 9.9 whitepaper

published in 2012, provides further insight into some of the considerations and strategies used by hardware manufacturers to manage the thermal and performance aspects of a server.

In recent years, hardware manufacturers have made significant headway in reducing the no load losses of their equipment. This amplifies the impact of the load variability or range since it starts with a lower baseline. In other words, the idle load condition of the data center can be orders of magnitude different than the full load condition based on the processor utilizations driven by the software.

Questions that arise from a wide range of load and rapid changes of loads may include:

- Can the infrastructure operate across the range of loads?
- Does the infrastructure operate efficiently across the range of loads?
- Can the infrastructure adequately respond to the rapid changes in loads and still meet requirements such



Advertisement formerly in this space.

as maximum permissible entering temperatures on the IT hardware?

Due to the extensive pressure placed on the industry for energy conservation, the hardware is controlled in many different states and at an unprecedented level of optimization. These new states have created a much wider band of operation and use of the various hardware assets on a server.

As a result, the predicted loads, range of those loads, and their rate of change has significantly increased. Data center designers need to be intimately familiar with the hardware domain.

Advertisement formerly in this space.

Software Domain's Impact

As discussed in a previous column,² software represents the actual occupancy of a data center. A solid understanding of software is paramount for a data center designer to properly quantify the project's needs.

Software can be divided into two main categories: systems software and applications software.

Systems software includes the operating system (OS) and all the utilities that enable the computer to function, sometimes referred to as operating systems and firmware.

Applications software includes programs that do real work for users. For example, word processors, spreadsheets, and database management systems fall under the category of applications software. (See Figure 3 for the hardware/software pyramid.)

Applications can be further divided into storage (databases, spreadsheets, etc.), compute (word processors, web browsers,

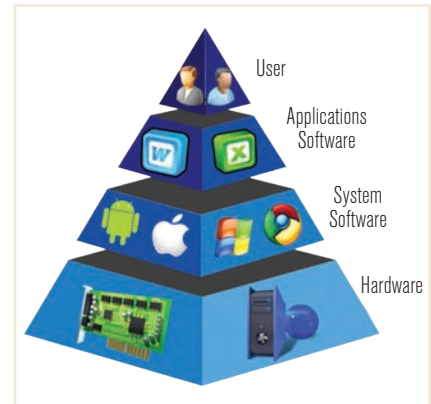


FIGURE 3 Hardware/software pyramid.

computational fluid dynamics [CFD], etc.), and communication (TCP/IP, Modbus, Layers 1–3 of BACnet, etc.).

Many applications can have components of two or three of these fundamental features.

Some data center designers are familiar with the concept of virtualization, which is masking from server users hardware resources, including the number and identity of the individual physical servers, processors, and operating systems.

The server administrator uses a software application to divide one physical server into multiple isolated virtual environments. In other words, a single piece of hardware is made to look like many pieces of hardware, so many operating systems and applications can be installed on a single server. In doing so, the use of the hardware resources can significantly increase.

These software terms are less known by data center designers, yet their impact on the load is potentially significant.

Unified computing system. A converged data center architecture

that integrates computing, networking and storage resources to increase efficiency and enable centralized management.

Hyper-convergence. A type of infrastructure system with a software-centric architecture that tightly integrates compute, storage, networking and virtualization resources and other technologies from scratch in a commodity hardware box supported by a single vendor.

Distributed computing. A model in which components of a software system are shared among multiple computers to improve efficiency and performance.

Hadoop. A free, Java-based programming framework that supports the processing of large data sets in a distributed computing environment. It is part of the Apache project sponsored by the Apache Software Foundation.

Follow-the-Sun. A data center that moves workloads to different geographic locations to more efficiently balance resources and demand.

Geo-Replication. A type of data storage replication in which the same data is stored on servers in multiple distant physical locations.

Software has shown that it can reach beyond the known boundaries of hardware. It is constantly changing the behavior of hardware development. Subsequently, it is going to radically change the data center infrastructure too.

Closing Comments

In the pursuit of optimization at all levels, the interdependent relationship between facilities, hardware, and software (The Data Centers Trinity) will drive the need for much more predictive and agile responses from the hardware and facilities/infrastructure of data centers.

Software is evolving at an ever-increasing rate in the pursuit of business efficiencies and optimization. It is bound only by the hardware it operates on and the data center infrastructure used to support it. However, software works with such agility it's even found ways to break these boundaries. Consequently, data center designers must adapt and adopt to software's ever-changing needs.

The current disparate gaps in the terminology within the Data Centers Trinity need bridging. Through a

better understanding of The Data Centers Trinity and associated terminology, data center designers will be better prepared for the agility being demanded of data centers.

The choice is to acquire this information yourself and actively remain current or access subject matter experts. In either case, ignoring the significance of the changing load profiles of software and hardware within the data center will surely result in its underperforming to clients' future needs.

Software operates in the unknown, unseen, and unheard environment of the data center. However, software has become an ever increasingly dynamic load that must be understood for the successful design of the supporting infrastructure.

References

1. Beaty, D. 2013. "Why datacom?" *ASHRAE Journal* (3).
2. Beaty, D. 2012. "Data Centers and Comfort Cooling." *ASHRAE Journal* (10).
3. Beaty, D., and D. Quirk. 2015. "Driver for energy savings: carrot or the stick." *ASHRAE Journal* (8). ■

Advertisement formerly in this space.