



The Data Center: An Ecosystem Under Pressure

The Problem of Thermal Management

If you're not having trouble with the thermal management of your data center yet, you soon will - It's virtually inevitable. This short paper outlines why heat will continue to grow as a major contributor to downtime, and suggests steps you can take to maintain a thermal environment that protects the most valuable asset of any modern enterprise. Further, this document will also illustrate how thermal considerations impact other components of the data center ecosystem, such as the need for space, asset management, and effective organization.

As any data center manager can attest, the modern enterprise has a voracious appetite for information. Terabytes of transactional and supply chain data allow business performance to be measured in near real-time. Compliance mandates for publicly traded companies such as Sarbanes-Oxley require visibility into every aspect of the business. New initiatives such as RFID allow the enterprise to track and streamline the flow of every part and every finished product.

In fact, data volumes continue to double on average **every nine months**, resulting in an increased need to query, report, and otherwise analyze data to enhance business performance in a hyper-competitive economy.

This is a double dose of bad news for the thermal integrity of the data center facility. First, with the increase in the processing power of a single CPU suggested by Moore's law comes an increase in electrical power consumption, and a subsequent increase in heat discharge by that CPU. Second, because the increase in processing power is not keeping pace with the demand of the modern enterprise for data, more servers are required in less space, putting an even greater strain on the ability of the data center to provide adequate cooling to each server.

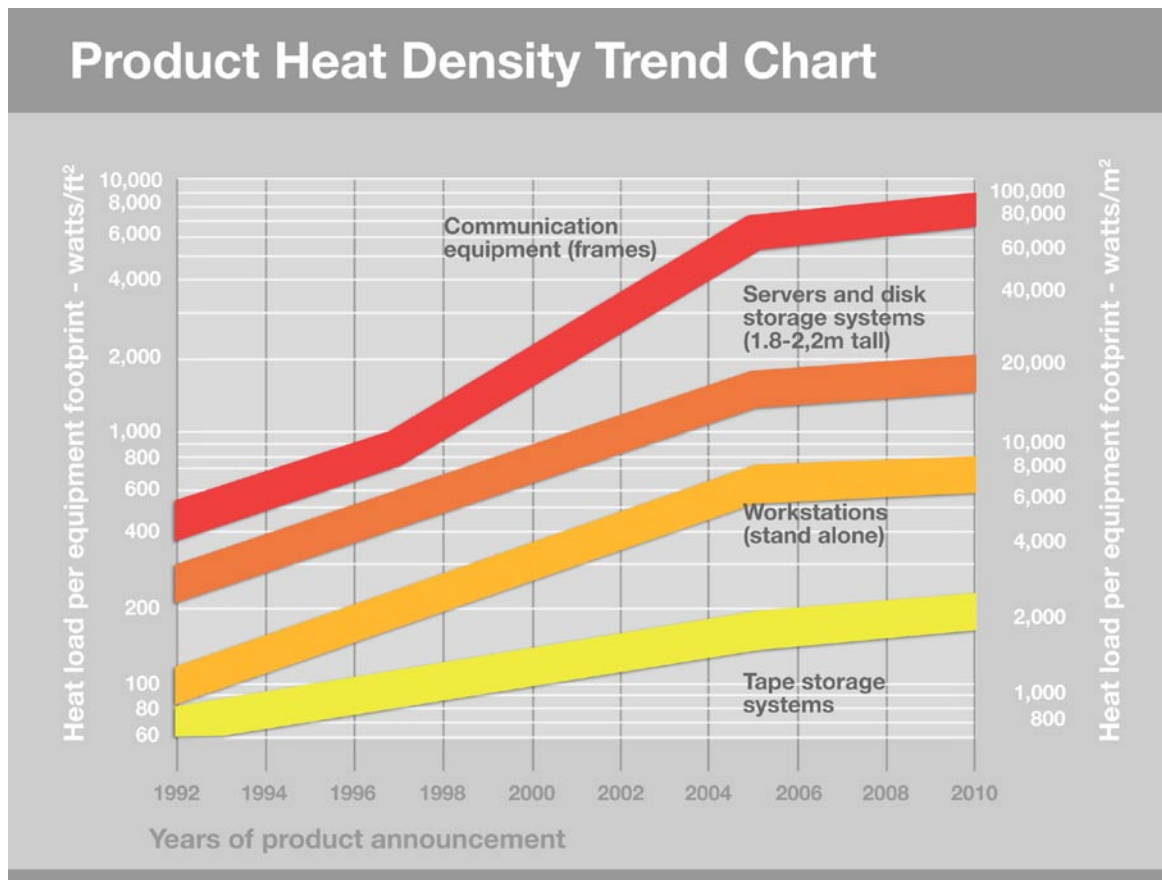
Moore's law states that the density of transistors, and thus processing power, in an integrated circuit will double every eighteen months. Based on the growth rate of data above, we can deduce that data is doubling roughly twice as fast as the ability to process the data with a fixed number of processors!

Obviously, with an increase in processing power comes an increase in the electrical consumption, and therefore heat emission. Even assuming that processing ability per CPU could keep pace with the growth of data, the data center would still have an increased demand for cooling over time.

As we know, this optimistic view of the data center is not even close to being true. In reality, to keep up with the growing demand to process data, the data center will have to accommodate an exponential increase in the sheer number of CPUs. To maximize the amount of processing power that can be accommodated, data center managers will have to squeeze smaller servers into a limited amount of space.

In some cases, footprint pressures will mean fitting more dual-processor 1U servers into one universal cabinet. In other cases, “Blade” servers will allow the data center manager to fit 8U worth of server hardware into 3U of space. Although these new, high-density servers process more data using less physical space than their predecessors, their space advantage – estimated at 14 to 17 percent per year of added free space – can easily be taken up by the additional cooling equipment they require.

Given that each CPU is generating more heat by itself, it’s easy to understand how serious the combination of an increase in server density and an increase in processor speed can be. The chart below summarizes the projected thermal implications of these two trends:



<http://www.upsite.com/TUI/pages/whitepapers/tuiheat1.0.html>

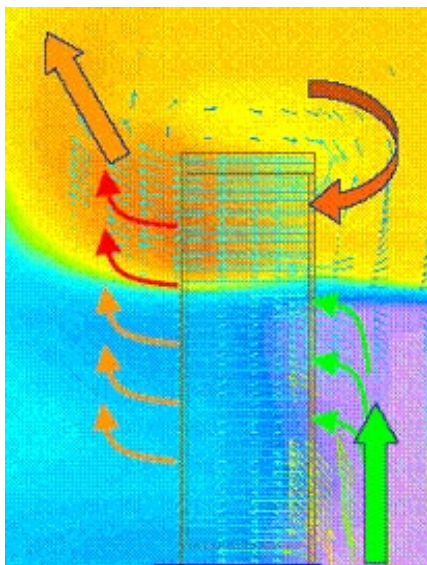
This chart shows how the heat management requirements of data centers will be impacted by the new, denser hardware products that are coming on the market. But older types of hardware will get hotter over time as well. Today, the typical server runs at less than 50% of capacity, mostly due to the fact that data center personnel are rightly unwilling to take a risk on system uptime by running systems harder. Going forward, data center managers will be unable to afford this luxury, and will require more precise cooling management to ensure thermal stability. Unfortunately, new software also complicates the ability of system administrators to run servers at near-idle. For example, server virtualization allows multiple applications – and even multiple operating systems – to run on the same server, thereby increasing processing load.

In summary, the potential is real – As the need to process data increases, existing data centers will be severely challenged in their ability to deliver an adequate thermal infrastructure to ensure uptime.

So what can be done? Of course, if data centers had an unlimited expanse of raised floor space, density would no longer be an issue. But they don't.

The key to dealing with thermal problems is delivering adequate cold air to every server's air intake. The increasingly common practice in raised floor data centers of creating hot and cold aisles has increased efficiency and helped with the density problem, but this practice is not adequate in many situations.

The higher the server is located in the cabinet, the more difficult the cooling problem, as cool air is typically supplied from the floor of the data center. Imagine a cabinet that has 600 cfm of cool supply air available, and is populated with 10 servers, each of which draws 100 cfm at its air intake. The lowest server draws 100 cfm of the rising air, leaving 500 to cool the remaining 9 servers. The next server up in the enclosure draws another 100 cfm, leaving 400 cfm to cool the remaining 8 servers, and so on. Obviously, there will be no cool air left for the seventh server at the top of the enclosure. Typically, the remaining servers simply draw air from whatever source possible, often times the hot exhaust from and around the ceiling! This phenomenon is illustrated by the diagram below.



If the majority of the cabinets in the data center have hot spots as shown above, there is no recourse short of a major redesign to increase total thermal capacity. This can be very expensive, because increasing thermal capacity is not simply a matter of adding another air conditioner. The airflow within the room is equally important, and in most raised floor data centers that airflow is determined by the area of the room itself, along with the design of the floor tiles.

Fortunately, a second, much less expensive approach exists: redistributing servers so that the hotter ones are interspersed with others that have less demanding cooling requirements. Obviously, this is not a process to be undertaken casually. Success is best achieved with the help of experts who specialize in the complex thermal dynamics of data centers, and can model proposed

D.S. De Lorenzo, "Air-Cooling in Servers and IT Facilities", Apr. 11, 2002, Intel Labs

solutions before they are implemented.

For those data centers where there are only isolated instances of overheating, a significant interim solution exists that will allow the data center to significantly increase the potential to deploy servers more densely: spot cooling. Devices that focus increased airflow within the data center to specific locations have been shown to be very effective. For example, an air curtain installed at the bottom of a problem cabinet can increase the cold air available to the servers at the higher elevations within the rack by 25%.

The need for better thermal management is one of many interdependent factors that must be balanced in order to maximize a data center's ability to meet its SLAs. Others include spatial consolidation, power delivery and cabling, security and site management, and of course, cost control. Papers on all these topics will soon be available at www.sharkrack.com.