

The Untold Story: Hidden Data Center Costs & Limitations

By [Bruce Whetstone](#), Executive IT Management Consultant with Taos' Office of the CIO Practice

"How could we be out of space in our data center?" asked the IT client services manager, incredulously. "I was just by the data center a few weeks ago, and we have plenty of rack space open."

(This brought to mind the old joke, "How could I be out of money? I still have checks in my checkbook.")

The reply he received was, "We're not out of rack space, we're out of cooling capacity."

"But I only need two more servers! Surely the cooling system can handle just two more servers?" the IT manager asked.

To which he was told, "Sorry, we're already over the cooling capacity limits for the room, and there is a long list of other requests for more servers ahead of yours in the queue."

IT professionals, though experts in their specific domains, often know very little about data center infrastructure and consider the data center to be an infinite resource.

In this series of articles, I hope to point out some of the most critical issues facing businesses and their data center managers by relating real-life examples.

The primary components necessary to build a data center include rack space (real estate), electrical power and cooling capacity. At any given time, one of these components is likely to be a primary capacity limitation, and over the past few years, the most likely suspect is power. The obvious requirement is power for the servers and network equipment, but sometimes less obvious is the power required to run the air handling and cooling systems. Unless you have built your data center right next to a power station, and have a very long contract in place for guaranteed supply of power at a nice low rate per kilowatt-hour, you are likely seeing your data center costs rise dramatically as the cost of electricity increases. Many IT professionals are finding themselves in the awkward position of begging their data center hosting providers for increased electrical capacity, or they are facing the unpalatable option of relocating their data center facilities.

While the industry is very aware of the issue of equipment power consumption, and the major manufacturers are already designing in power reduction and conservation features to their latest products, we are far from turning the corner on the demand for more data center electrical power. The demand for computing capacity continues to rise, which implies a demand for "more, better and faster" systems. This results directly in higher demands for rack space, power and cooling in your data center, and related increased costs.

The increasing costs are receiving a lot of executive management attention, especially given current economic conditions. Server rationalization is the buzz word of the day. Do we need to buy a new server, or can we re-use one we already have? Are we using the servers we have to their full capability? A typical data center has very high storage utilization (as no one voluntarily throws old data away), and server utilization is low (as different functions within a company don't want to share). Server virtualization has become one of the hottest topics in IT, as it has become a means of ensuring higher utilization of servers, lowering data center costs and providing flexibility for shuffling systems' workload. But there are limits to what can be virtualized, and of course, the overall utilization of a server pool has an upper bound, as well.

But there are other components needed to build a data center, including: network equipment and communication circuits, power distribution equipment (e.g., distribution panels, cable and outlets), power backup equipment (including generator(s) and uninterruptible power supplies (UPS)), cable trays (for both network and power cables), and fire suppression systems.

And, yes, believe it or not, sometimes these other components can become the constraining factor in data center capacity. "Cable trays can be a limiting factor?" you ask. Yes, just two years ago, we ran into a situation where an older data center couldn't add capacity to a specific cage because the weight of the cable already in the tray was at design load limits. We couldn't risk the tray ripping out of the ceiling by laying in more cable, and we couldn't shut the network and power down long enough to pull out all of the old cables and put new ones back in without excessive downtime to the business. A very expensive migration to a new cage in the data center became the only feasible option.

Though it may seem hard to believe, there are many IT professionals who have never seen the inside of a data center. Over the years, having a glass-walled computer room as a showcase in your corporate headquarters became problematic for a number of reasons including, security, high-priced real estate for servers that could be better accommodated elsewhere, as well as loss of the glass wall space for other purposes (e.g., communications punch-down blocks, electrical distribution panels). Besides, the number of blinking lights in the data center has steadily decreased over the years. And you don't see white lab-coated workers pushing buttons and handling tapes. So, what's there to see? (Not much, especially in a "lights out" facility.)

But the interesting part is: when corporate executives and managers **do** occasionally visit a data center facility, they still expect to see nice clean rows of equipment, full racks of blinking lights and servers happily computing away. Instead, we now see large amounts of unused floor space and partially filled racks. As servers have become more powerful per cubic inch of space occupied, power and cooling capacity have become increasingly scarce, not the rack space for the equipment. The decreasing server footprint relative to the higher energy per cubic inch requirement is often referred to as a "power-density" problem.

You should ask your data center manager for a tour. Standing behind a full rack of 40+ servers consuming 200 watts (or more) of power each is an amazing experience, likened to having someone turn six 1200-watt hair dryers directly toward you, running full blast.

While the heat load behind a server rack has its shock effect, the more interesting cognitive dissonance for many executives is seeing the empty racks and floor space. When called upon to explain, my simple example has been this: if the power capacity (power cap) in your cage (or computer room) is limited to 100 kWh, it doesn't matter whether you have 10 full racks that consume 10 kWh each, or 20 partially filled racks that only consume 5 kWh each. If you have only one supercomputer that consumes all 100 kWh sitting in the middle of the room taking up only 20 square feet, and there is 500 square feet of unused space all around it, it may look very odd, but you're still out of power.

It would be wonderful if the data center design would "come out even," with exactly the right amount of full racks, power and cooling to look like the space is being well-utilized, but that is not a common occurrence these days. Even if you can optimize one cage, it's extremely difficult to optimize across the entire data center floor.

Building out a new data center space requires careful planning and engineering of all the components discussed in this article. And even then, changes over time in server utilization, increases (and decreases) in the business' needs, virtualization and server technology all conspire to generate re-work of your carefully thought-out and well-balanced design.

Bruce Whetstone is an Executive IT Management Consultant with Taos' Office of the CIO Practice. Contact him at bwhetstone@taos.com