

How Capacity Drives Your Data Center Strategy: Build Your Own, Move to a Co-lo Facility, Outsource it All? Part 3 in a 4-Part Series on Data Centers

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

"We are rationalizing, virtualizing, standardizing and consolidating, but we're still out of capacity in our existing data center, and we need a plan for growth."

Not too many IT leaders are in a position to make the above statement. Most have additional capacity that can be found in their existing facilities (with, of course, some input of time and effort to reconfigure systems and to invest in the tools to manage virtualized systems, which offsets some of the savings). They just don't have the processes in place (to apply the rigor needed) to squeeze the last ounce of juice out of their existing facilities.

But when it is time to procure more data center capacity, the following questions invariably arise:

- Should we build our own?
- Should we move to a co-location data center?
- Should we outsource our data center to a managed services provider?

Unfortunately, there are no simple answers, as there is a long list of dependencies. There are serious cost, flexibility and long-term capacity-related constraints, no matter which option you choose.

If you plan to buy or build a primary corporate data center in a location of low real estate cost and low energy cost, you might well conclude owning is worth the investment. The recent drop in real estate prices in the U.S. notwithstanding, at least part of the interest in owning your own data center is to keep the underlying asset costs low and predictable, and possibly even see some capital appreciation over the long-term. However, owning a data center comes with the responsibility to manage the facility, including staff, maintenance and operating expenses, but also the need to have a capacity plan for the center. Knowing how much capacity to buy or build requires extensive planning. Do you move in on day one at full capacity, or do you reserve a percentage of the facility for growth? If you reserve room for growth, how much? How much unused or underutilized capacity are you willing to carry in exchange for the certainty that you will have it when you need it?

Some IT leaders pursue a hybrid strategy of owning some of their primary data centers (for the reasons already cited) and leasing additional co-location or third party data center space (ping, power and pipe only) as overflow or "safety valve" space.

But what is a co-location data center? What do you mean by "ping, power and pipe only"? A co-location center typically is a third party data center, which is multi-tenant and shared by many companies but also has data communications connections to major network providers' networks in the same building (or complex). (See the [Colocation Center Wiki](#).) The advantage of locating high-availability, customer-facing applications in a co-location center is they are typically Tier 3 or Tier 4 in terms of uptime and reliability, and by having high-bandwidth Internet connections already on-site, allow for lower network costs (compared to having to extend high-bandwidth circuits to your own facility or a remote 3rd party location). However, the additional uptime comes at a cost, and the companies that offer co-location services have to make huge capital investments and take the risk that their facilities won't be fully utilized. Their pricing, therefore, has to compensate for both the expected return to investors or shareholders and the long-term risk of underperforming on capacity utilization.

Still, co-location is an extremely popular option these days for customer-facing Internet applications, especially those that require high network performance. And because of their popularity, prices have been going up dramatically year-over-year, due to pressures on demand (but also due to increases in energy prices).

Before we get into energy price issues, what is "ping, power and pipe?" Just because you decide to lease a data center space from a co-lo or third party data center services provider, doesn't mean that you have to turn over management of

the servers and data network equipment residing in the space to them to manage. Yes, there are companies that will buy the servers and equipment, provide the data center space, and manage the systems for you, but this is an entirely different business model, and the pricing is considerably different (and much higher) for this type of service. “Ping, power and pipe” means that the 3rd party is only providing you with a secured (caged) area in a shared data center where you can then place your own servers (ping), network equipment and connections to network carrier circuits (pipe) and provide the *power* to plug it all in.

There are four very important considerations when moving into a co-lo or third party data center arrangement that often do not get the attention they deserve:

1) You don't get away from having to forecast your capacity needs.

Even though the data center provider has thousands of square feet of raised floor, their incentive is to sell out the space to full capacity (as we've discussed before, whether the capacity limit is power, cooling or floor space is dependent upon the situation). Once a facility is “sold out,” and you suddenly realize you need more capacity:

- a. You may have been forward thinking and reserved some buffer capacity within your cage
- b. Your cage may be at maximum capacity, but the 3rd party provider has other cages within cable distance that you can lease and extend your network to
- c. The provider may have capacity in another building in the complex within cable distance
- d. The provider may be completely at capacity limits within the complex, requiring that you:
 - i. Move your entire footprint (extremely costly)
 - ii. Build out another “overflow” center elsewhere, with an attendant increase in network costs

Unfortunately, experience has shown that most companies are not willing to pay upfront for ‘buffer space’ (for future growth) or seriously underestimate their needs for growth space. In either case, the data center provider sells out all of the cages around yours, and your cage becomes ‘landlocked’. This is a difficult issue for many companies to understand, grapple with and resolve, as it becomes a “pay me now, or pay me later” issue. If the company doesn't want to pay for the underutilized capacity to allow for sufficient growth, the cost to move a mid-to-large-sized data center later can run into the millions of dollars (which would have paid for a lot of months of underutilization), and can be very disruptive to the business.

Understandably, businesses would like to avoid the entire discussion, and would rather talk about “pay by the drink” computing (utility model or transaction-based pricing model), grid computing, or software as a service (SaaS) models. Of course, there is no free lunch, and all of these models have their own sets of issues and trade-offs. (Too many complications for this discussion, but these topics will be kept in mind for future articles).

2) Energy costs have to be factored into the data center providers' pricing algorithm. Many 3rd party providers are still charging by “rack and run,” i.e., by the number of racks (floor space) you occupy and a charge for each electrical cable supplying power to your space (which varies based upon whether you are running 110v, 20 amp circuits, or 220v, 30 amp circuits, etc.). The problem with “rack and run” pricing methods is that the provider typically will set the price of an electrical run for the term of your contract, but may have price increases from its utility during the term. These costs have to be factored into the price per month rate to you. Even if you don't use the full capacity of an electrical run of cable, the provider has to charge you as if you did, because you *could* install equipment, at any time, which could push the circuits to maximum capacity. In this case, you are essentially paying for what you reserved, not what you actually used. Because of the essential “unfairness” of this approach, some providers are converting to power metering and usage-based charges for power instead of a fixed price per power cable run.

3) Regardless of whether the pricing for power is circuit-based or usage-based, many providers are forced to set a maximum power usage capacity limit (power cap) on each cage. The server and data equipment in use today can consume so much electrical power per unit of floor space, if every tenant crammed their racks completely full of high power-consuming equipment, the overall power draw to the facility would overwhelm the supply. Power caps must be kept in the forefront of the designer's plans for your data center power, rack and equipment layout, as the cap typically

becomes the initial most important factor in how densely the racks can (or should) be filled. (Refer back to Part I of this series for a discussion of the power-density problem in present day data centers.)

4) At the end of the contract period, the data center services provider typically does not have to renew your contract, and certainly not at the same rates. We have seen more than one situation where a company's lease has expired and the provider is being inflexible on price increases or wants to force the tenant to move out (for the provider's reasons). This can result in a 'panic' to find an alternate location and result in a near-emergency move of the data center to a new location under extreme time pressure, putting the business at risk. Contract terms around move out considerations have to be carefully constructed to prevent adverse impacts to the business.

Moving data center responsibilities to a managed services provider *does* make sense for many companies, especially for common corporate (non-customer-facing) applications that the third party is well equipped and staffed to support. Yes, it's true; there will be no direct comparison in costs between third party managed services and third party co-lo, as the staffing, tooling and on-going management cost models will be completely different. Constructing a cost model to compare internal company costs to managed services requires advanced skills and experience in information technology costs, operations and financial modeling.

Generally speaking, companies that are successful in sourcing their data center support to a managed services provider have:

- 1) Good documentation and processes in place around their existing internal IT support services
- 2) Good understanding of their existing cost structure, including cost/unit of service
- 3) Good contract review and negotiation skills (legal and procurement)
- 4) Experienced supplier relationship managers in place

Falling short of a 'good' rating in any of these areas often leads to a rocky relationship (at best) or disastrous results for your company, in terms of poor service and/or higher costs to the business.

So, after all of this, you plan to build your own data center, or source it to a co-lo, "ping-power-pipe" or managed services provider? Where on the planet should we place the equipment? What are the considerations in selecting a location for your company's data center? In our next installment, we discuss the issues and trade-offs around the need for global applications performance vs. network latency and where to locate a data center, considering the effect of natural disasters (seismic and weather-related events) and energy costs.

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