EXECUTIVE GUIDE SERIES – PART 2

Total Cost of Ownership

by Julius Neudorfer
June 2012

This is the second of a six part series of our Executive Guide whitepapers:
1. Data Center: Build vs Buy
2. Total Cost of Ownership
3. Data Center Energy Efficiency
4. Creating Data Center Strategies with Global Scale
5. Custom Data Centers
6. Data Center Designs

Brought to you by

DIGITAL REALTY
Data Centre Solutions
Introduction

The Total Cost of Ownership, second in the Data Center Executive Guide series

Total Cost of Ownership "TCO" is basic and easy to calculate, just ask any first year economics or business major. Nonetheless, the true TCO of the data center is somewhat more elusive to project accurately. There are many subtleties which can be overlooked or are simply unaccounted for (or perhaps underestimated), over the operational life of a data center. This Executive Guide will examine the many more granular aspects, such as energy costs and operating efficiency, as well as other data center specific issues.

Total Cost of Ownership: An Executive Overview

Like any other TCO analysis, there are the more obvious directly identifiable line items, such as initial capital investment and depreciation; however, the TCO of data center has many more elements, some of which are far more complex. Let's start by defining and understanding the basic factors of TCO calculations.

Whether you decide to build, lease or buy a data center, before you can compare and calculate TCO, it is important to understand and clarify the specific terms of the data center industry:

- Total Space (Building Shell)
- Whitespace (Raised Floor)
- Effective Usable Space (Rack Space)
- Adjacent Total Lot Size (raw lot size)
- Critical Load Power Capacity
- Power Density

Unlike other commercial real estate purchases or leases, the raw square footage of the data center Building Shell is a relatively minor factor in the overall TCO calculation. The more frequently used basis is "whitespace" which is a common industry term for the raised floor area where the computing equipment resides. However, be sure to not limit your TCO space evaluations just on "whitespace" alone. It is the net usable footprint that is available for the computer equipment, the majority of which is housed in racks or cabinets. The effective usable rack space is less than the whitespace and significantly less than the gross building shell space.

The other key factor is the Critical Load Power Capacity. Effectively, and most importantly, this is the net power available for your computing payload, primarily based on the available power capacity of the UPS systems.

Understanding the Critical Load Ratings

Ultimately the purpose of the facility is to support the "Critical Load". This is typically expressed as kVA or kW. With today's power factor corrected computing equipment, expressing the critical load capacity in kVA is no longer as relevant and can skew the real costs. Be sure to base the TCO calculations based on kW capacity, not KVA. This is primarily based on the type and rating of the UPS equipment used in the data center.

It should not be confused with gross utility power to the site, which while it is important to the overall power required to operate the site's other systems such as cooling equipment, is the Critical Power that represents the conditioned and protected power that is available for your computers. (See sidebar: Additional TCO Benefits of Improved Energy Efficiency, on page 4.)

However, beside the Critical Power capacity, the power density will impact how much computing equipment can be placed in each rack. A data center with a lower power density would mean that you may need to use more racks (and whitespace) to house the same amount of computing equipment than at a higher density site. Power density is typically expressed in two ways; watts per square foot or Kilowatts (kW) per rack, or sometimes both. This is primarily based on the design and type of the data center cooling system. Many older data centers cannot effectively or efficiently cool more than 5 kW per rack (some even less), and in some cases their efficiency goes down beyond 3 kW per rack. Even today, not all newer data centers can accommodate medium (5-10kW per rack) or "high-density" racks which require 10 kW or more per rack.
In addition, the overall land area required around the building is a secondary, but necessary consideration. This is used for infrastructure related support systems, such as utility transformers, generators and fuel storage, cooling equipment, as well as security systems such as entry and exit point gates and security kiosks. There is also exterior space required for staff and visitor parking, as well as separate receiving zones for equipment deliveries and cargo security inspection of materials entering and being removed from the data center.

Cost of Energy
Like fuel cost to an airline, the cost of energy represents one of the most significant TCO components in the data center. Over the long term it is one of the most variable costs, and based on current and foreseeable trends, it will continue to rise over time. The cost of energy is typically expressed as cents per kilowatt hour (KWH) and varies widely by region, as well as the underlying source of power generation. It can also vary based on the time of day, seasons and changes in the cost of fuels sources, such as natural gas or coal. In the United States the basic energy rate can vary by a factor of as much as 10:1 ratio (2.5 – 25 cents per kwh), which is primarily based on geographic location.

One of the issues that can also impact the cost of the energy is if it is purchased directly, vs purchasing via an aggregation contract. When doing an energy cost projection analysis, be sure to compare the directly purchased cost from the local utility, which would be based on only your own energy buying requirements, compared to the cost of an energy aggregation contract. Taking advantage of aggregation can be done independently, or in conjunction with some of the larger facility operators, some of whom may offer this lower cost energy as part of their lease. In effect, they may buy 10-20 times more energy for their customers across their multiple facilities and sites, which can lower the energy rate. In addition, energy aggregators are focused on power costs and may have fixed price or capped contracts to prevent price shocks. However, be careful to see if the facility operator is offering to pass-through those savings or if you are forced to purchase power only from the facility at a price that may actually be higher than the regular cost you would have paid directly from the utility.

In addition, there may be some capital costs involved to bring in new utility service if you are building your own site in a newer underdeveloped area, while land may be cheaper, its lower cost may be offset by the high cost of new utility service. In some cases, the facility operator may have already installed a large sub-station on the data center campus, as part of the overall site development. This can not only minimize some capital installation expenses, it can also mitigate the risk of extensive delays in getting utility approvals for your own new multi-megawatt service facilities and potential local right-of-way or permitting issues.

When making an initial decision as to location of the site, energy rates should be given close scrutiny and serious weight in the decision, but should not be the only or overriding factor.
Total Cost of Ownership

Site Selection

Site Selection Factors

• Geographic Location – Stability
• Distance and Availability of Utilities and Communications Providers
• Cost of Energy (Local Rates and Green Energy Incentives)
• Distance and Availability of Communications Providers
• Availability of Skilled Labor and Costs

Of course picking a site location that is physically secure and has reliable access to power, water and communications is an important first step. Since energy is the most significant operating cost of a data center, focus your attention on the cost of power and its long term impact. Energy costs are highly location dependent and are based on local or purchased power generation costs (related to fuel types or sustainable sources such as, hydro, wind or solar), as well as any state and local taxes (or tax incentives). In the United States it can range from 2.5 – 25 cents per KWH and also can vary by the scale and peak energy demands of your site. The US Department of Energy provides a state-by-state guide of average energy prices; however, it is important to check local rates and look for utility and state energy incentives which can provide lower costs, energy efficiency rebates or tax benefits. Another factor is location and long term overall market demand for constrained resources such as power and water, which can ultimately limit the data center capacity and impact the initial capital cost and final TCO.

In addition, having access to redundant power from separate sub-stations is an important consideration, and if possible; two separate power grids (this is not often feasible but highly desirable from a reliability viewpoint). If this is a location with or near a high level of industrial or commercial development, this may reduce the costs of bringing new utility service from a nearby sub-station, however, be sure to investigate overall power capacity and reserves to ensure that power will not be constrained in the future. If the site is relatively remote and needs to be newly developed, be sure to factor in the cost of bringing in new high voltage utility services, which can be expensive and require long lead times to have approved and installed.

Carrier provisioning (fiber network connectivity), which again depending on site location and distance from major carrier(s) Point-of-Presence, can have a significant impact on the bandwidth, redundancy and latency in relation to a target market; such as the real-time financial transactions.

Site Selection can also impact the facility’s energy efficiency (primarily cooling system efficiency), since it is related to the ambient temperature conditions and is location dependent.

---

Additional TCO Benefits of Improved Energy Efficiency

Lower Initial Capital Cost and More Power for Computing Equipment

Besides the obvious lowering of energy usage, improving the facility’s operating efficiency (lowering the PUE), there are additional benefits. When designing a data center’s entire electrical system, it must be correctly sized to support the projected Critical Load of the computing equipment, plus all the entire facility’s infrastructure (lighting, power conditioning and cooling, etc) systems. This multiplies upwards when calculating the capacity of the utility system feeds, as well as the size of the back-up generators and related main switchgear.

Typically the utility capacity and generator capacity needs to be approx 2.5-3 times the critical load capacity (at a PUE of 2). By improving the operating efficiency, it also lowers the total power capacity requirements for the facility for the same given Critical Load — so that in a data center with a PUE of 1.3 it may only need to have a utility feed and generator capacity of only 1.8 — 2 times the critical load. This results in lower upfront capital costs for many of the electrical items in the power chain, as well as lowering recurring maintenance costs.

In addition, for any given size of the utility feed, a higher operation efficiency (PUE of 1.2-1.5) would allow a significantly higher Critical Load for virtually the same initial capital cost, when compared to a lower efficiency design (i.e. PUE of 2).
Comparative Maintenance Costs

Unlike a typical commercial building, which normally does not operate on a continuous basis and have downtime maintenance windows (i.e. nights and weekends) data centers do not have that option. While virtually all systems require maintenance, there is additional redundant power and cooling equipment installed to provide “concurrent maintainability”. Maintaining the electrical and mechanical systems in a 7x24 data center requires more complicated equipment, as well as a more sophisticated meticulous approach, since downtime is not an option. This requires a staff trained and experienced with bypassing critical systems, without affecting the computing systems.

Staffing Labor costs

Site staff requires multiple skill sets that must be available 7x24 — both common skill sets such as administrative and physical security (guards), as well as technically qualified electrical and mechanical personnel. In a smaller organization it may be more difficult and expensive to find, train, motivate and retain experienced personnel with the specialized skill sets required for a data center, especially with complex redundant power systems.

The TCO for staffing in a smaller site (or organization) will generally be higher per unit (SF and/or KW). In particular, the resources (skill levels) requirements to operate and maintain a 10,000 SF site or a 50,000 SF data center, are essentially the same.

One of the inherent advantages a larger data center facility provider has is the efficiency of scale. Their support staff can oversee a campus site with multiple data centers covering hundreds of thousands of SF on a 7x24 basis, with almost the same personnel cost as an individual organization would need to properly support their own 10,000 SF site.

Expected and Unexpected Costs

Beside regular maintenance of all the power and cooling equipment, some items such as UPS batteries, require replacement anywhere between 5–10 years, but typically around 7 years. This is a significant cost and should not be overlooked. In addition, there can be unanticipated expenses from unexpected equipment failures, which can be significant. These costs may or may not come into play depending on whether you built your own site or if you leased the site from a facility operator and it may be covered under the terms of the lease.

Occupancy Rate – Design vs Actual

The capital costs are based on the design maximum and the TCO really is based on the effective use (i.e. occupancy rate over time of the actual fill-up of the computing systems). Essentially you need to calculate the TCO based on your initial usage in the first few years, as well as when the site is more fully populated in 7, 10 or more years. Alternately, if your organization computing requirements are such that you intend to fill the space and fully utilize power capacity almost immediately, your TCO numbers will be based on a fairly straight line number.

Reserved Capacity

Realistically, a data center is not operated at 100% of design capacity for a number of reasons, primarily for ensuring equipment reliability and maintaining uptime. Depending on the organization culture, typically systems are operated at no more than 80–85% of design ratings (some may push to 90%) before it is considered “full”.

Total Cost of Ownership

Oversize or Undersize Design Capacity Impacts TCO

When deciding on the design capacity of the data center there are many competing factors that influence the decisions. The fear of making it too small and running out of space or power in only a few years, is a very realistic scenario and fear. In recent years the growth in computing power demands and power density has made many data centers that were built less than 10 years ago functionally obsolete, a real risk of then looking for additional space or power that is not readily resolvable with a dedicated single site. Conversely, oversizing will mitigate that risk, but will increase both the Capex and Opex and of course, the TCO.

Modular Design

One method to mitigate the potential of over or undersized data centers is modular design. Capacity Planning and modular capacity designs can help mitigate the risk of capacity or functional obsolescence. In some designs, the total space and utility capacity is designed and built upfront, but only individual sections are fully outfitted with the UPS, generators and cooling equipment. This saves both upfront capital cost and recurring maintenance expenses. Moreover, it also improves energy costs, since the smaller sections are more fully occupied and operate at a higher efficiency. This modular design still allows for growth, but in a pre-planned, stepped fashion.

Understanding Power Usage Effectiveness (PUE) Metric – The Green Grid

Power Usage Effectiveness “PUE”, was initially introduced by The Green Grid in 2008 and has become a basic globally recognized metric of facilities infrastructure energy efficiency. It became an internationally agreed upon metric in 2011, when the US Dept. of Energy, EPA, European Union and Japan agreed to it as a mutually satisfactory metric. Previously data centers were primarily focused on reliability, and not on energy efficiency. The basis of the PUE metric is relatively straightforward; it is the ratio of the Total Energy being used by the facility divided by the Energy used by the Computing Equipment (measured on an annualized basis). The range for PUE measurement is 1.0 (theoretically perfect — 100% efficient) with no upper limit (very inefficient). In reality, the average older data center used twice as much energy, as was delivered to the computing equipment (this represents a PUE of 2.0 or 50 percent operating efficiency). In some cases the PUE of some older sites are even worse with PUE of 2.5 – 3.0. New mainstream data centers are being designed and built that have much better operating efficiencies, with a PUE range of 1.2 – 1.5. Needless to say this has a direct impact on energy usage and therefore TCO.

Beside the improved efficiency of newer designs and newer more efficient power and cooling equipment, the cooling systems of data centers represent the largest use of energy other than the computing load itself. This is also directly related to cooling energy necessary to maintain the very tight operating temperature and humidity for the computing systems. Raising the allowable temperature and widening the humidity ranges has a significant impact on lowering the cooling energy required in the data center (note to reader: the next upcoming edition of the Executive Guide series will focus on Energy Efficiency).
The Bottom Line

It would be prudent to not oversimplify the TCO evaluation of building a data center. Do not be shocked if the costs may seem much higher than expected, this reaction is especially common to those not intimately familiar with build-out costs of the data center industry.

While the evaluation can be done by your staff in-house with the appropriate experience, consider doing it as a team effort in conjunction with an outside trusted and respected consultant (either as a collective effort or as independent opinion). The TCO analysis should result in a projected range over time (at 7, 10 or even 15 years), with allowances for different operational scenarios and varying occupancy rates over time, as well as possible fluctuating energy costs, not just a simple fixed number, which would not adequately reflect the known and unknown variables.

Julius Neudorfer Bio

Julius Neudorfer is the CTO and founder of North American Access Technologies, Inc. (NAAT). Based in Westchester NY, NAAT's clients include Fortune 500 firms and government agencies. NAAT has been designing and implementing Data Center Infrastructure and related technology projects for over 20 years.

Julius is a member of AFCOM, ASHRAE, BICSI, IEEE and The Green Grid, as well as a Certified Data Center Design Professional “CDCDP” designer and instructor. Most recently, he is also an instructor for the US Department of Energy “Data Center Energy Practitioner” “DCEP” program.

Julius has written numerous articles and whitepapers for various IT and Data Center publications and has delivered seminars and webinars on data center power, cooling and efficiency.