Microsoft

PROJECTING ANNUAL NEW DATACENTER CONSTRUCTION MARKET SIZE

By Christian L. Belady, P.E.

GLOBAL FOUNDATION SERVICES | MARCH 2011

A DAY IN THE MICROSOFT CLOUD

Author's Note: this work is intended to be an exercise to estimate the size of the datacenter construction market over the next decade and is intended to seed the industry for further discussion on the topic through good public debate.

Introduction

Jevons' Paradox [1] has been one of the reasons given for the incredible growth we are seeing with online services. The explosion of services now available on the internet has fueled one of the fastest growing industries today: Datacenter construction. However, there has been little publically available information that quantifies the actual year-over-year monetary opportunity for the industry. Clearly, the opportunity is huge, and some businesses have been trying to estimate the size of market that their technology can capitalize. In addition, the past year has brought us significant changes in the cost per megawatt (MW) of capacity as a result of broader adoption of airside economization and advances in software resiliency. Today, some estimates claim that datacenter costs are on average about $15 million per MW. This falls right in line with The Uptime Institute's estimates of datacenter costs ranging in the $10-22 million per MW depending on redundancy.[2] However, new datacenter designs are emerging that are as low as $6 million per MW. How does this shift affect the overall market? This paper will attempt to "size the global market for new datacenter construction" using recognized industry data, as well as some general observations and sub trends that may actually decrease expected growth rates in the near term.

In early 2007, Jonathon Koomey's research showed that datacenter growth from an electrical power footprint perspective was 14 percent year-over-year in the United States (US) and 16 percent globally.[3] Later that year, this work was heavily leveraged by the US Environmental Protection Agency's (EPA) report to congress on datacenter power use in the US and has been heavily referenced since its publication.[4] Figure 1 shows the power footprint for datacenters in the US with a 14 percent growth rate (or a doubling every five years).

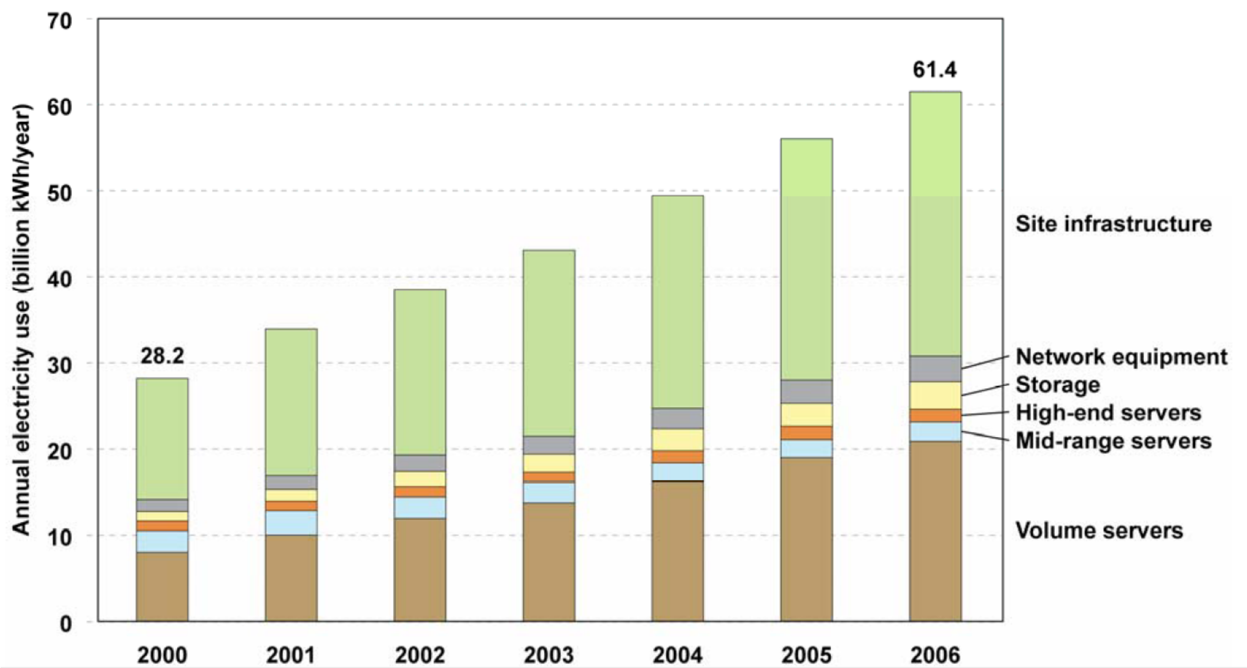

Figure 1. Year-over-year Datacenter Power Footprint[4]

Unfortunately, the EPA does not provide a similar figure for global electrical power footprint. In this case we need to leverage Koomey's work. However, when reconciling Koomey's work, there appears to be a discrepancy between the two reports. For example, in 2000, Koomey reports that the footprint was 23 billion kilowatt hour (kWh), while the EPA reports 28.2 billion kWh. Upon further inspection, this discrepancy is reconciled with the addition of network and storage power which was not included in Koomey's original body of work. Why is this important? This 20 percent addition for storage and network is used to adjust the Koomey's global estimate as shown in Figure 2:

Region	Year	Koomey (TWhr)	EPA (TWhr)	Adjusted (TWhr)	Comment
US	2000	23	28.2	NA	EPA numbers include 20% for storage and network
	2005	45	54	NA	
Global	2000	58	NA	69.6	Koomey numbers adjusted by 20% to include storage and network
	2005	123	NA	147.6	

Figure 2. Comparison of Koomey's and EPA's reports on Datacenter Power Footprint

Assuming these trends continue, with all other factors remaining relatively constant, this data can be used to generate a projection to 2020 as shown by Figure 3. For both US and global trends, the curves represent an approximate doubling every five years. Clearly, these are projections based on past performance and there are many factors that can change this such as energy costs, efficiency and cost gains, technological shifts, and new business models. However, for this discussion, it is assumed that Figure 3 represents a plausible future.

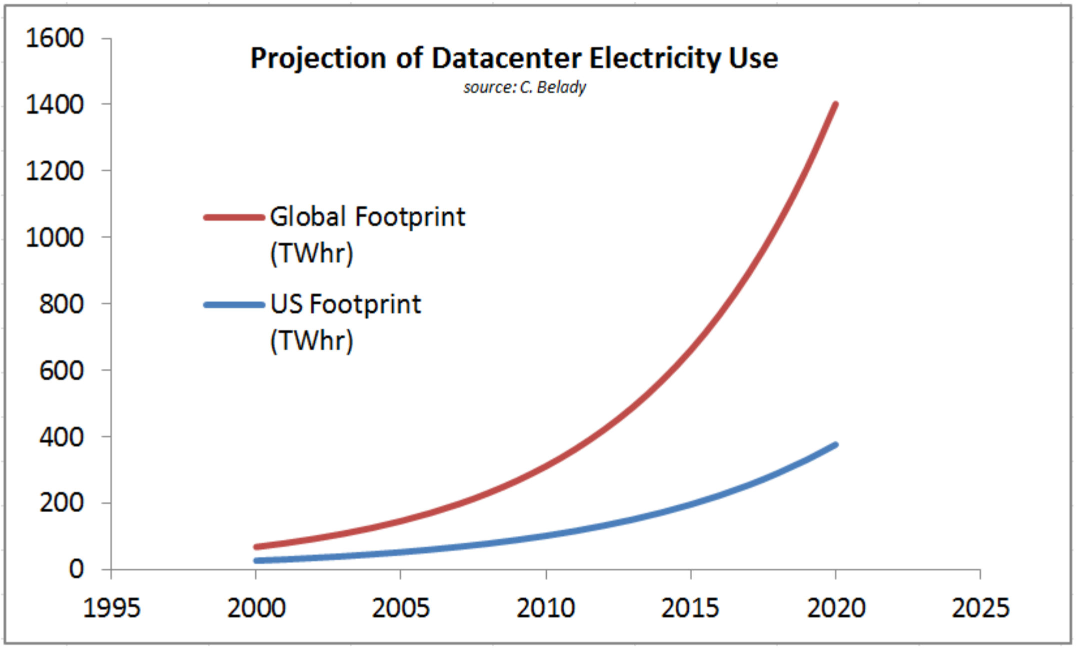

Figure 3. Projection of total electricity use by datacenters in the US and the world based on Koomey's[3] and EPA's[4] data

Projecting the Market

Figure 3 projects the EPA data out to 2020, but is based on a Power Usage Effectiveness (PUE) of 2.0. Clearly PUEs have been changing, particularly for cloud companies such as Microsoft, who are now building datacenters that have PUEs below 1.2. These new hyper-efficient datacenters are built at costs around $6 million per MW because they are taking advantage of innovations such as airside economization and others. It is important to note that these costs (as is true with the Uptime Institute's values) are all based on critical load capacity and not facility utility load. This means that we need to use the critical load as the basis for the calculating datacenter costs. As a result, since the PUE is 2.0 in Figure 3, the critical load for both the US and the globe would be 50 percent of that shown in the curves. So the methodology for calculating the annual construction dollars both in the US and globally is as follows:

1) De-rate both the US and global curves in Figure 3 by 50 percent to set the basis for critical load instead of utility load.
2) Calculate the additional capacity each year by subtracting the previous year's critical load from the current year's load. Take this result and divide by 8,760 hours to get the power capacity growth for that year.
3) Using the basis from (#2), now multiply those numbers by the appropriate cost per MW to get a market size for the given year.

Figures 4 and 5 represent the size of the datacenter construction market each year as a function of the type of datacenter being constructed each year for both the US and global markets, respectively. For example the Tier 4 datacenter represents a market size that assumes every datacenter built is Tier 4. Assuming the "average" mix of datacenter redundancy remains constant, for the next decade, the likely market size by 2020 will be $50 billion and $218 billion, respectively.

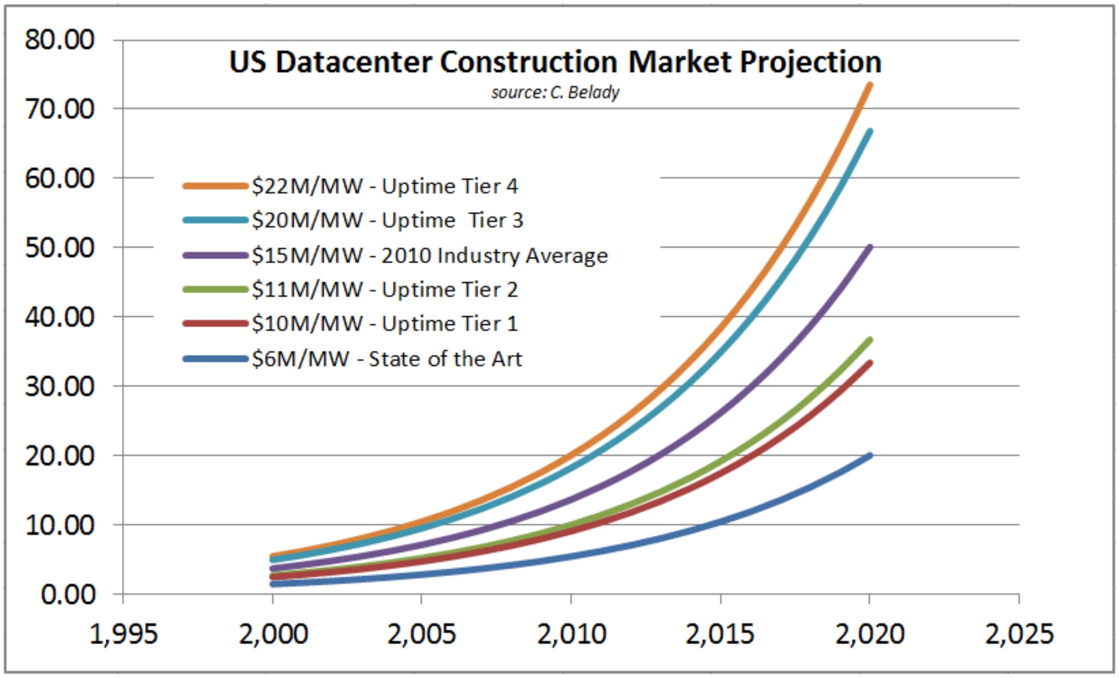

Figure 4. US datacenter construction market size as a function of cost per megawatt

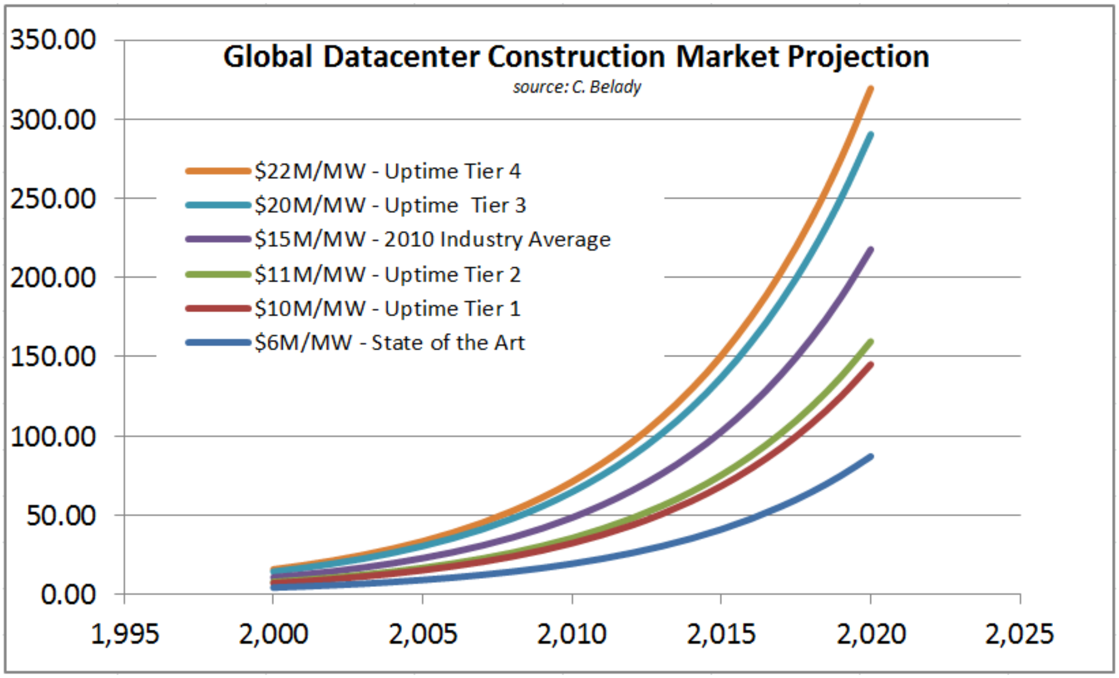

Figure 5. Global datacenter construction market size as a function of cost per megawatt

As impressive as these numbers look, they likely only represent an upper bound for the datacenter construction market size and will be subject to other forces that will limit this growth. Some of these could be quite unpredictable such as:

1) **Global Economy:** To date, the IT industry has weathered the economic downturn relatively well and the question is whether this will be sustained over the next decade.
2) **Energy availability and Cost:** Clearly, for IT operations, higher energy costs could slow down the growth of the industry.
3) **Construction and Labor Costs:** Marked increases in the cost of construction and/or labor could alter datacenter economics and retard growth
4) **Carbon Regulation:** Related to (#2) is the possibility of carbon taxation or limits for emissions on IT operations. These would all impede growth.
5) **Innovation:** Leading to breakthroughs in storage, networking and semiconductor technologies.

These are all relatively unpredictable and beyond the scope of this paper. However, there is a force that is effecting the annual growth in datacenter construction that is a bit more tangible: the shift from highly redundant and well-controlled datacenters to lower cost, highly-efficient cloud datacenter designs. These designs leverage their scale and application redundancy, as opposed to hardware redundancy to drive down cost. In addition, these designs use aggressive economizations that employ either liquid or air to help drive down cost and substantially improve efficiency. These datacenters today are characterized by costs on the order of $6 million per MW and will continue to go lower in the future. Figures 6 and 7 show the effect that this could have on the market size for both US markets and global markets, respectively, with the $6 million per MW and $15 million per MW curves for reference.

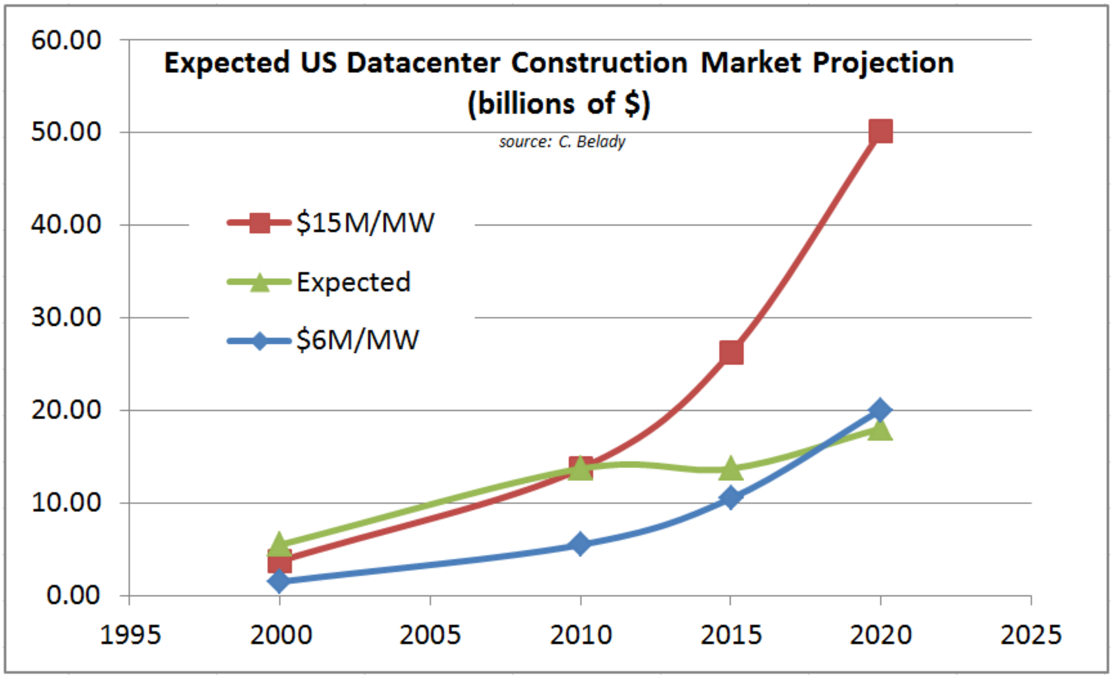

Figure 6. Expected US projection of datacenter construction size each year

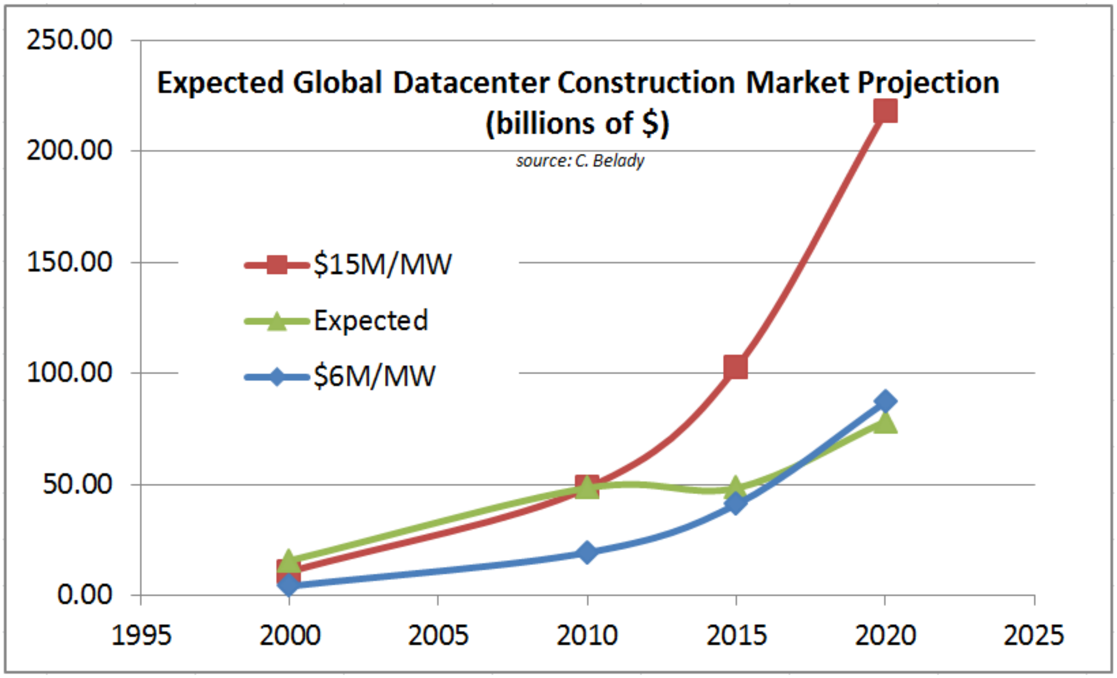

Figure 7. Expected global projection of datacenter construction size each year

What both of these figures show is that as a larger percentage of new datacenter builds are driven by cloud infrastructure expansion, resulting in a shift in the cost average for the next five years to a point where all new datacenters will be built with cloud-like infrastructure costs. In other words, we can expect a flattening until this transition occurs. Clearly the five year conversion time may be suspect, but it is important to note that this shift from typical datacenter infrastructure projects to cloud-based infrastructure projects is likely to occur in a time frame of less than 10 years with five years being a mean expectation. Figure 8 shows a summary for both US markets and global markets.

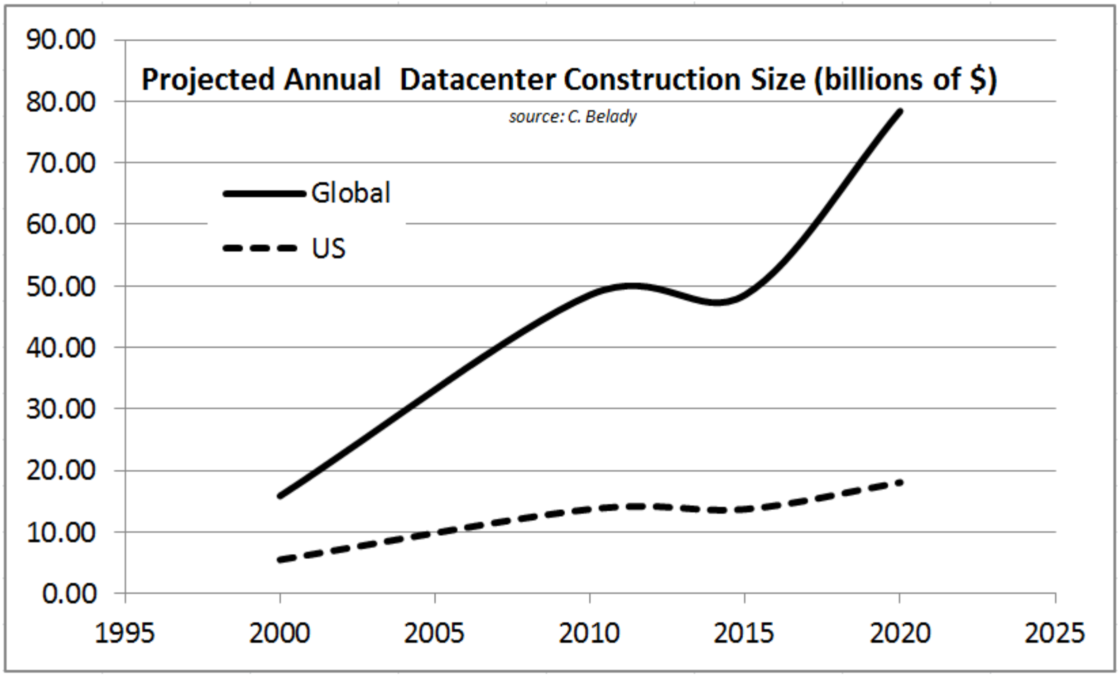

Figure 8. Year-over-year datacenter construction market size

Conclusion

This paper is an attempt to project the size the construction market for the datacenter industry year-over-year. Admittedly, there are many assumptions behind the projections that may prove to be incorrect in the future. However, this study does provide a rudimentary sizing of the market. This work could help some businesses project their true market opportunities, but will also require subjective assessments of the variables that impact these projections.

Some key take-aways from this work are:

1) Infrastructure capital expenditures will likely flatten over the next few years as a result of innovations in new designs that bring down cost. However, once these innovations become prolific and further improvement opportunities diminish, the growth will continue at historic rates.
2) In 2000, the annual construction market size globally was three times that in the US. By 2020, it will likely be four times as large. This should be no surprise that the growth will be higher globally, then in the US.
3) By 2020, we can expect the construction market size for datacenters to be about $18 billion in the US and $78 billion globally.
4) This is merely a projection to seed further dialog (in the absence of a more in-depth study that could provide more guidance). However, the future is unpredictable and precautions need to be taken when using this data.

In conclusion, expect long term growth in capital spending on datacenters. However, in the short term (next five years), expect some flattening due to the emergence of hyper-efficient datacenter designs.

References:

1. Belady, C., "Does Efficiency in the Data Center Give Us What We Need?", Mission Critical Magazine (Spring 2008) http://www.missioncriticalmagazine.com/CDA/Articles/Features/BNP_GUID_9-52006_A_10000000000000340120

2. Turner, P., J. Seader, "Dollars per KW plus Dollars per Square Foot Are Better Data Center Cost Model than Dollars per Square Foot Alone", Uptime Institute 2006 http://uptimeinstitute.org/

3. Koomey, J., "Estimating Total Power Consumption by Servers in the US and the World", AMD 2007 http://sites.amd.com/us/Documents/svrpwrusecompletefinal.pdf

4. US Environmental Protection Agency, "Report to Congress on Server and Data Center Energy Efficiency Public Law 109-431" http://www.energystar.gov/ia/partners/prod_development/downloads/EPA_Datacenter_Report_Congress_Final1.pdf

5. Belady, C., "In the Data Center, Power and Cooling Costs More than the IT Equipment It Supports", Electronics Cooling Magazine, Volume 13, Number 1, February 2007