Initial Steps in Capacity Management for the Green Datacenter

Stephen Markamer, Aetna Inc.
James Castronovo, Inventa Technologies
Table of Contents

Introduction..................................................................................................... 3
Who Are the Stewards of Green? ................................................................. 4
How Facility Managers Measure Power in the Datacenter ...................... 6
  Power Usage Effectiveness and Datacenter Infrastructure Efficiency .......... 6
  Power versus Energy: Kilowatts versus Kilowatt-Hours ......................... 7
The Capacity Role: Profiling Energy Efficiency with Work Density ......... 8
The Capacity Statement for Green Datacenter Initiatives ....................... 10
Conclusion .................................................................................................. 11
Citations ..................................................................................................... 12
Introduction
This paper provides a vision for Capacity Managers to address system server management within the datacenter before, during and after your company forms a corporate Green initiative. We review how companies start down the path towards the Green Datacenter, and how you, as a Capacity Manager, can make meaningful contributions towards that goal. We discuss various approaches to datacenter greening, our analysis of work density and how we relate it to green programs sponsored by business executives residing outside of IT Operations. By developing the capacity questions and supporting analysis that we describe in this paper we believe that Green Datacenter initiatives give us an opportunity to turn to a new page in the old book of Capacity Management.

This paper is subtitled Beyond Facilities and Power, because as Capacity Managers we do not focus on Facilities or Power Management. Yet Capacity Management still should have a large role in greening the datacenter. What we discuss here is a method that any Capacity Manager using historical system data can apply to the green datacenter efforts at their company. Our discussion will focus on:

- Who Are the Stewards of Green?
- How Power is Measured In the Datacenter
- Our technique: Density Analysis
- How to tie our Density Analysis in with existing initiatives

We will introduce our topic by first looking at the traditional owners of the green initiatives. We will look at how they measure power for Datacenters. However, this paper is not meant to be an in-depth tutorial on measuring power. It will only give a brief overview to help set the stage for the server perspective. There are many papers found throughout the internet, but especially on Green Grid website, [http://www.thegreengrid.org/](http://www.thegreengrid.org/), where Facilities and Power Management perspectives are very well developed. Instead we will explain our capacity-planning technique, Density Analysis and will wrap up considering how our work will add to the greening of the datacenter and support the existing Green Datacenter initiatives.
Who Are the Stewards of Green?

In 2008 Gartner published the results of a survey of IT executives to determine the top motivators for going green with datacenters. The top four motivators for companies to ‘go green’ in their datacenters were:

<table>
<thead>
<tr>
<th>%</th>
<th>Motivator</th>
</tr>
</thead>
<tbody>
<tr>
<td>65%</td>
<td>Reduce energy operating expenses</td>
</tr>
<tr>
<td>32%</td>
<td>Reduce IT operating expenses</td>
</tr>
<tr>
<td>33%</td>
<td>Align IT with a corporate green initiative</td>
</tr>
<tr>
<td>41%</td>
<td>Do the right thing for the environment</td>
</tr>
</tbody>
</table>

They found that two thirds of the executives –that’s the 65% shown in the top line—responded that reducing energy operating expenses related to the data center facility was the main motivator, but that reducing operating expenses for IT provided only half as much motivation. The report goes on to support the idea that green datacenter initiatives are facilities-owned, and not IT owned. That is the underlying reason for such a difference between those two motivators.

In fact, the main motivator of the project with which we were associated is avoiding construction of new datacenters. This is a facilities-driven story. If, however, we examine example projects documented on the internet, we find they are also facilities-centric. In our research, we have found two categories of green initiatives that are broadly IT-based:

- Reducing energy usage of existing hardware
- Replacing hardware with more energy-efficient hardware

Also we have determined that green initiatives concerning computer hardware are companywide and not necessarily exclusive to the datacenter. The result is the following matrix that encompasses most green IT initiatives.

*Figure 1: Matrix of tactics for carrying out Green IT initiatives*
We have placed one example of a successful Green IT initiative within each quadrant of the above matrix.

**Quadrant 1 example initiative: Energy Reduction – Companywide IT**
Quadrant 1’s example is illustrated by Dell Inc’s implementation of software that power-down and power-up 50,000 desktops at close and open of business. The result is an ongoing annual savings of $1.8 million in energy costs which translates into a 40% reduction².

**Quadrant 2 example initiative: Energy Reduction– Datacenter**
Barkley’s Bank fine-tuned the management of their datacenter cooling and reduced energy use by 13% annually³.

**Quadrant 3 example initiative: Infrastructure Replacement – Companywide IT**
Enterprise Rent-A-Car implements a rental transaction system using thin client terminals and replacing desktop computers. This saves 6.5 million pounds of annual carbon dioxide emissions⁴.

**Quadrant 4 example initiative: Infrastructure Replacement – Datacenter**
Highmark Blue Cross Blue Shield uses Leadership in Energy and Environmental Design (LEED®) certified energy efficient building architecture to construct a datacenter with twice the capacity of their original datacenter yet with the same energy usage⁵.

These are all wonderful successes but they are also very much facilities-driven stories.

Stewardship of green initiatives is mostly located in the realm of Facilities Management. Adding Capacity Management to the green spectrum sometimes takes luck, coincidence, and perhaps a little risk taking. From an organizational standpoint, this entails:

- Extending green initiatives from the facilities realm to include the IT operations
- Specifically take greening into the realm of Capacity Management
- Leverage the fact that Capacity Management owns the management of server utilization

So where do we as managers of capacity fit in? In those companies where the green IT initiatives are owned by an IT executive, not a facilities executive, IT is more likely to have better representation and Capacity Management should be a part of that. Reporting of overall progress for power involved both facilities and IT system services. The IT champion of the project reports to our director of Technology Management Services. That director reports to the CIO. The IT champion needs to understand how the server inventory changes in respect to the power deployments. This means that our Enterprise Management Services organization, responsible for reporting inventory, is on board with the initiative. In Enterprise Management Services, we initiated a project to begin reporting on the full inventory of servers in our datacenters. The reporting included utilization of that inventory. This was not a simple endeavor. It required an understanding
of a very complex and changing environment that included standalone Windows servers, rack Windows servers, ESX servers. It also includes standalone and rack Linux servers; SUN Solaris partitioned frames, and finally AIX on IBM Power series frames. The Capacity Management organization was asked to participate. We became key players in the development of reporting of the server floor. Capacity Management owns the management of utilization.

How Facility Managers Measure Power in the Datacenter

Before we get to Capacity Planning, let us review some facilities-related metrics:

Power Usage Effectiveness and Datacenter Infrastructure Efficiency

We have 2 important metrics measuring power in the datacenter: Power Usage Effectiveness (PUE) and Datacenter Infrastructure Efficiency (DCiE) which are mirror images of the same metric^5 and they deal with the same measures.

\[
\text{Power Usage Effectiveness} = \frac{\text{Total Facility Power}}{\text{IT Equipment Power}}
\]

PUE represents all power needs for the facility divided by the power needs for the IT assets alone. This gives a ratio of the energy that the building needs to support IT assets versus the energy needed just for the IT assets. Accurate mechanisms discount office space and unrelated power requirements.

\[
\text{Datacenter} = \frac{\text{IT Equipment Power}}{\text{Total Facility Power}}
\]

The DCiE metric is the mathematical reciprocal of the PUE metric.

Both PUE and DCiE provide Facilities Managers ways to determine:

- Opportunities to improve a datacenter’s operational efficiency
- Comparison with competitive datacenters
- Opportunities to repurpose energy for additional IT equipment
- Design and process improvements

How are PUE and DCiE used? PUE is a multiplier to determine the total energy a facility needs to support its IT assets. For example a PUE of 3 represents a 3:1 ratio of required energy brought into the building to power the IT equipment as well as additional facilities infrastructure to support that equipment. This metric provides an understanding of how much power the facility needs to support one new asset. For example: you have a
new storage device with a 10 kilowatt (kW) power rating. Using the 3:1 ratio, to support the new storage device, your total power needs for the building are 30 kW.

**Power versus Energy: Kilowatts versus Kilowatt-Hours**

A more basic concept that we need to clarify is the difference between power and energy or the Watt versus the watt-hour. Watts or kilowatts are a unit of Power assigned to an appliance or a consumer of energy and a kilowatt-hour is a measure of Energy consumption.

Here is a clear example: a 40 watt light bulb. The 40 watts describes the power of the light bulb. In one hour, the 40 watt light bulb draws 40 watt hours of energy. Often times though we talk in terms of kilowatts and kilowatt-hours. Figure 2 provides an example of various light bulbs and their energy use.

<table>
<thead>
<tr>
<th>Watts</th>
<th>Power (kW)</th>
<th>Energy Use (1 hour)</th>
<th>Energy Use (24 hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>0.04</td>
<td>0.04 kW hours</td>
<td>0.96 kW hours</td>
</tr>
<tr>
<td>60</td>
<td>0.06</td>
<td>0.06 kW hours</td>
<td>1.44 kW hours</td>
</tr>
<tr>
<td>100</td>
<td>0.10</td>
<td>0.10 kW hours</td>
<td>2.40 kW hours</td>
</tr>
<tr>
<td>200</td>
<td>0.20</td>
<td>0.20 kW hours</td>
<td>4.80 kW hours</td>
</tr>
</tbody>
</table>

*Figure 2: Example of light bulb power versus energy consumption*

To apply that concept to servers in a datacenter, we have an example of a set of 6 HP servers in the figure below along with their power and energy consumption in one day, one month and one year. One note about the server type: HP figures prominently in our inventory and therefore in our study illustrations. The actual hardware model does not matter. The use of these hardware types is just a reference to our real-world example.

<table>
<thead>
<tr>
<th>HP Server</th>
<th>Power</th>
<th>Energy 1 day</th>
<th>Energy 1 month</th>
<th>Energy 1 year</th>
</tr>
</thead>
<tbody>
<tr>
<td>HP BL 460c</td>
<td>0.2 kW</td>
<td>5 kW hours</td>
<td>144 kW hours</td>
<td>1,752 kW hours</td>
</tr>
<tr>
<td>HP DL 380</td>
<td>0.3 kW</td>
<td>7 kW hours</td>
<td>216 kW hours</td>
<td>2,628 kW hours</td>
</tr>
<tr>
<td>HP DL 580</td>
<td>0.4 kW</td>
<td>10 kW hours</td>
<td>288 kW hours</td>
<td>3,504 kW hours</td>
</tr>
<tr>
<td>HP DL 585</td>
<td>0.5 kW</td>
<td>12 kW hours</td>
<td>360 kW hours</td>
<td>4,380 kW hours</td>
</tr>
<tr>
<td>HP 1850R</td>
<td>0.8 kW</td>
<td>19 kW hours</td>
<td>576 kW hours</td>
<td>7,008 kW hours</td>
</tr>
<tr>
<td>HP RX8640</td>
<td>3.9 kW</td>
<td>94 kW hours</td>
<td>2,808 kW hours</td>
<td>34,164 kW hours</td>
</tr>
<tr>
<td>Total</td>
<td>6.1 kW</td>
<td>146 kW hours</td>
<td>146.4 kW hours</td>
<td>53,436 kW hours</td>
</tr>
</tbody>
</table>

*Figure 3: Example of server power versus energy consumption*
The HP BL 460c distinguishes itself as the lowest power server in this table. The HP Integrity RX8640 server stands apart from the other five servers as an outlier which you can appreciate with the table of energy consumption below.

![Figure 4: Distribution of the kW hours by server type](image1.png)

**The Capacity Role: Profiling Energy Efficiency with Work Density**

You would think from the graph above that the prudent green decision is to migrate datacenter capacity to HP DL servers, HP BL blade servers and other servers with low power. In fact that is a green direction at many companies. The technique entails moving over server work, one to one, on to Blades. Let us take a look at how that happens. Starting with an old server with a Pentium III processor in it, we measure peak utilization at 90%. Note that it has a power of 500 watts.

<table>
<thead>
<tr>
<th>Server</th>
<th>Core</th>
<th>Core Count</th>
<th>Utilization</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old Server</td>
<td>Pentium III</td>
<td>1</td>
<td>90%</td>
<td>0.50 kW</td>
</tr>
<tr>
<td>New Blade</td>
<td>Xeon</td>
<td>8</td>
<td>5%</td>
<td>0.23 kW</td>
</tr>
<tr>
<td>Savings</td>
<td></td>
<td></td>
<td></td>
<td>0.27 kW</td>
</tr>
</tbody>
</table>

![Figure 5: Work Migration Scenario 1](image2.png)
The work moves on to a new blade that has 4 dual-core Xeon processors, which does the same work using just 5% of its CPU resources. The new blade has a power of just 230 watts. Now is this green? Most observers will say yes it is for two main reasons:

- Because you are going from 500 watts to 230 watts
- There is a migration of work off of old technology to new which requires less maintenance concerns

Both of these points are good. However, we are also going from 90% utilization to 5% utilization. This is counterproductive and we feel like something is missing here. What we are missing is the opportunity to take advantage of the increased processing power. For example, let’s say we can take the work of 7 of the Pentium III servers and put it on that same blade.

<table>
<thead>
<tr>
<th>Server</th>
<th>Core Type</th>
<th>Core Count</th>
<th>Utilization</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 Old Servers</td>
<td>Pentium III</td>
<td>1</td>
<td>90%</td>
<td>3.50 kW</td>
</tr>
<tr>
<td>New Blade</td>
<td>Xeon</td>
<td>8</td>
<td>35%</td>
<td>0.23 kW</td>
</tr>
<tr>
<td>Savings</td>
<td></td>
<td></td>
<td></td>
<td>3.27 kW</td>
</tr>
</tbody>
</table>

*Figure 6: Work Migration Scenario 2*

Now the blade runs at 35% and our power savings is 3.27 kW savings; this is well above and much more green than the savings of 0.27 kW in the prior example. Another way of looking at the issue is that after the first work migration scenario 0.23 kW supports or buys you X quantity of work, and after the second work migration scenario the same 0.23 kW supports or buys you 7 times more, or 7X. Now let’s normalize on the power, and not on the work. We do this by dividing the power by the work variable. This is resolved as the following:

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Work Variable</th>
<th>Server Power</th>
<th>Work produced per kW</th>
<th>Work produced over baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>1X</td>
<td>0.50 kW</td>
<td>2.0 X</td>
<td>NA</td>
</tr>
<tr>
<td>1</td>
<td>1X</td>
<td>0.23 kW</td>
<td>4.4 X</td>
<td>2.4 X</td>
</tr>
<tr>
<td>2</td>
<td>7X</td>
<td>0.23 kW</td>
<td>30.0 X</td>
<td>28.0 X</td>
</tr>
</tbody>
</table>

*Figure 7: Work Migration Scenario 2*

So now we can make the statement that as Capacity Planners we have two target scenarios to propose involving the movement of work onto a more energy-efficient blade:
one proposal where one kilowatt of power will produce 4.4 quantities of work and a second scenario where one kilowatt of power will produce 30.0 quantities of work. This difference in work brings to light that the density of work distinguishes the degree to which you may consider a server “green”.

A word about Density: Many administrators of virtualization technologies such as VMware understand density as the number of virtual servers that one frame hosts. We abstract the concept of density further to indicate the quantity of work that a server, and even a datacenter, performs. We will flush out a meaning for “quantity of work” in a discussion of our Density Analysis technique in our presentation and publication entitled, The Green Datacenter: Beyond Facilities and Power. The publication will be available in the proceedings of the 36th annual International Conference for the Computer Measurement Group (CMG ’10). For now, we would like to highlight that by normalizing IT transactions or work in terms of power we enable Capacity Managers to speak a language that the owners of green datacenter initiatives understand.

Also, before moving on we would like you to take one more glance at the pie graph figured above. By looking at the chart you may think that the HP BL 460c server rated at 0.2 kW is more efficient than the HP Integrity RX8640 rated at 3.9 kW. Reason: the DB 460c requires less than one tenth the power. However, consider this: the BL 460c is indeed the less efficient server if the Integrity server supports more than 10 times the work. This returns to the idea that in addition to the power assignment, density of work plays an important role in determining the degree to which you may consider a server to be energy efficient. We are now investigating next steps that will leverage benchmark standards to better formulate a relative Work to Power Efficiency metric.

The Capacity Statement for Green Datacenter Initiatives

To rephrase our objectives, we believe Capacity Planning can best contribute to green datacenter initiatives by measuring the density of server work and by coaxing a greater density of work onto more energy efficient frames. From here, we can formulate a capacity statement related to green datacenter initiatives:

Capacity Management works in support of a green datacenter initiative to promote a greater density of work placed on more energy efficient frames and thereby reducing the total IT Equipment Power.

You may remember IT Equipment Power described earlier in this paper. It is a special term taken from the Facilities Management discipline. We have used it earlier in this paper as the denominator of the PUE metric that Facilities managers use. The significance of that is we have found a way to influence an important Facilities Management metric and thereby play a role in Green Datacenter initiatives.
Conclusion

In this paper we have put forth a variety of ideas. We have explained that corporate Green Datacenter initiatives are traditionally owned by Facilities Management and not IT Infrastructure Management. We developed a matrix to describe most published Green Datacenter initiatives. We provided specific examples of initiatives that emphasize their facilities-focus. Before making an effort to provide a Capacity Management approach, we describe the language of Facilities datacenter initiatives, specifically the PUE and the DCiE. From there we describe two important elements over which Capacity Management has control: Density of work and server power. We then relate those elements back to the Facilities initiatives through the PUE and DCiE. We look forward to presenting a deeper discussion of our techniques in server and datacenter density in the CMG ’10 presentation, *The Green Datacenter: Beyond Facilities and Power.*
Citations

1. Is Green IT Your Emperor With No Clothes? Doug Washburn, Forrester
   http://www.newsroom.barclays.co.uk/content/Detail.asp?ReleaseID=1329&NewsAreaID=2
4. Highmark Inc. was the winner of Computerworld’s “Top 12 Green-IT Users” award. Source: Computerworld Inc., “Top 12 Green-IT Users: No. 1 Highmark Inc.,” February 15, 2008
   http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=310494