GIGAOM RESEARCH

The value of green HPC

Dave Ohara

October 30, 2013

This report was underwritten by Verne Global.

Cloud

TABLE OF CONTENTS

Executive summary

Forward-thinking CIOs are anticipating increased regulation of carbon emissions and want lower and more-predictable energy costs over the long term. As part of that process, they are looking at ways to go green. They know that data centers are under scrutiny for how sustainable they are, and they know that demand for data center services is growing while the costs of fossil fuels are already high, getting higher, and becoming difficult to predict.

Green data centers present one solution because they use renewable energy sources, have efficient data center facilities, and use efficient IT equipment. The savings these data centers offer can be transformed into more processing power, which gives new opportunities for increased business revenue. Many of these data centers are located where they can take advantage of an area's natural resources (cool climates, for example) and sources of power such as wind, geothermal, and hydroelectric.

However, not all applications are suitable for offloading to a data center, whether it's green or not. Deciding which applications can be placed in a green data center while still satisfying business and performance specifications is critical to success. Among the candidates to consider are high-performance computing (HPC) applications. HPC was once limited to scientific research, but many businesses now use it to analyze large amounts of data and to create simulations and models. HPC applications are compute-intensive and, when applied at scale, require large amounts of energy. However, because users of these applications don't require real-time responses, you have flexibility in where you place these applications. This means that you can take advantage of the lower energy costs a green data center offers, no matter where it's located. This report analyzes these topics as well as the following areas:

- Three factors to consider in choosing a green data center for HPC are the source of the data center's power, the efficiency of its IT equipment, and the data center's efficiency.
- Today's CIOs have the options of building a new data center, refurbishing an existing data center, using co-location, and using the cloud. Each option needs to be balanced against the following criteria: the requirements of increased data center traffic, government regulations, volatile energy costs, and sustainable practices.
- Latency is the single most important criterion for choosing the appropriate applications for cloud or co-location. Following latency, other considerations are whether the application must peer with another company, the business requirements, the application architecture, current and predicted application workload, and the application's resource consumption rate.

Why green data centers?

CIOs, CTOs, and IT facilities managers, who determine their organization's business strategies and IT policies, must handle the conflicting demands for increased use of data center services and pressures to reduce data center budgets. At the same time they must cope with:

- Current and upcoming environmental regulations
- Increasing energy costs
- Customer sensitivity to sustainable practices

The emerging regulatory environment

An increasing number of jurisdictions have regulations that mandate the reduction of greenhouse gases (GHGs), with a particular emphasis on carbon dioxide, which is released when fossil fuels burn. In the United States, at the federal level, there is the EPA mandatory reporting of GHGs. States have regulations of their own, such as California's AB 32, which requires California to lower greenhouse gas emissions to 1990 levels by 2020. This is the equivalent of taking approximately 15 million cars off the nation's roads. Other carbon legislation includes the Regional Greenhouse Gas Initiative (RGGI) and the Western Climate Initiative.

The EU has made a unilateral commitment to reduce overall greenhouse gas emissions from its 28 member states by 20 percent compared to 1990 levels by 2020. It has offered to increase this reduction to 30 percent if other major economies agree to undertake their fair share of a global emissions reduction effort. The 20 percent reduction commitment is included in the climate and energy package of binding legislation. It is also one of the headline targets of the Europe 2020 strategy for smart, sustainable, and inclusive growth.

These regulations use a cap-and-trade model, in which there is a limit to the total amount of carbon emissions allowed and each data center would be assigned a carbon allocation. Exceeding the allocation would result in fines and other penalties. The concept of carbon credits and the ability to trade carbon credits is also a part of these laws.

Fossil fuel risk

The volatility of fossil fuel prices, as well as the overall trend of their becoming more and more expensive, is a prime concern for data centers. The incentive to look at sources of energy that offer more price stability is strong. In a March 2013 article, *Forbes* quotes from a speech that Rick Needham, Google's director of energy and sustainability, gave at a Cleantech forum. Needham says, "While fossil-based prices are on a cost curve that goes up, renewable prices are on a march downward."

Customer demand

Data center customers are becoming more aware of sustainable practices. Rackspace's April 2012 survey "Green Survey: Key Findings" asked customers how they weighed sustainability against cost. They found that:

- When "Two Choices Were Equal," 54 percent of customers felt that "Greener was Better."
- When "Two Choices Were NOT Equal," 20 percent of customers would choose the greener option, signaling that there is tangible value to the reduced risk, higher performance, and higher efficiency among service providers that embrace sustainability.
- Only 26 percent of respondents said that cost outweighed a greener option.

The report goes on to say that "sustainability gives a service provider an edge over another regardless if two choices are equal or not, according to 74 percent of respondents."

What is a green data center?

A green data center is designed to maximize energy efficiency and to minimize environmental impact. How do you quantify these characteristics? There are three factors to consider.

- **Source of data center power.** What is the carbon footprint of the electricity that powers the data center? Are renewable sources of energy used?
- **Efficiency of the IT equipment.** How efficient are the processors and memory of the physical servers and routers in the data center? In other words, how much computation can they perform per kilowatt-hour consumed?
- **Efficiency of the data center facility.** What percentage of the total energy consumed by your data center goes to computing, as opposed to noncomputing operations such as cooling, humidity control, and power delivery?

Source of power

The basic premise of a green data center is that the power it uses has low, or preferably no, carbon emissions. The most efficient data center isn't green if it uses electricity from a high-carbon source such as coal.

The geographic location of a data center can be important. A data center located in an area with access to hydro, geothermal, or wind power, for example, would have a lower carbon footprint than a data center located in an area that depends on coal, oil, or natural gas. For example, the Data Centre Risk Index reports that in the search for renewable energy and carbon neutrality, "The colder climate allows improved free cooling or at worst reduced mechanical cooling and with access to almost limitless supplies of hydro power or alternative renewable energy at comparatively inexpensive rates, the Nordics have become an increasingly attractive prospect."

IT efficiency

The efficiency of the IT equipment is a key component of a green data center. Jonathan Koomey of Stanford University says in his June 27, 2013, blog post, which discusses a paper on the characteristics of low carbon data centers:

> "The critical lesson from the analysis is that IT efficiency (which includes higher utilization and performance improvements as well as purchasing efficient hardware) is the most important issue on which to focus. Most recent efforts in the industry have been on improving infrastructure efficiency, which has many beneficial effects, but is not as important a lever as is the IT efficiency."

Some of the best-known ways to increase IT efficiency are virtualization, good capacity planning, and using energy-efficient equipment. The Green Grid's paper "Using Virtualization to Improve Data Center Efficiency" discusses the advantages of virtualization and some implementation strategies. Organizations such as EPEAT can help you find energy-efficient equipment.

Another issue is that IT equipment purchasers are often insulated from energy costs, especially secondary costs such as cooling, which may come out of another budget. ZDNet's summary of the Uptime Institute's 2012 data industry survey, "The results are in," reports that in only 20 percent of companies did the IT organization pay the bill.

Data center efficiency

A data center's efficiency is often represented by its power usage efficiency (PUE), which The Green Grid developed in 2007 and is widely used in the industry. This metric is expressed as total facility power divided by IT equipment power. A data center that uses 100 percent of its power for IT equipment would have a PUE of 1.0. In practice, all data centers have a PUE greater than that because of infrastructure requirements such as cooling and power distribution.

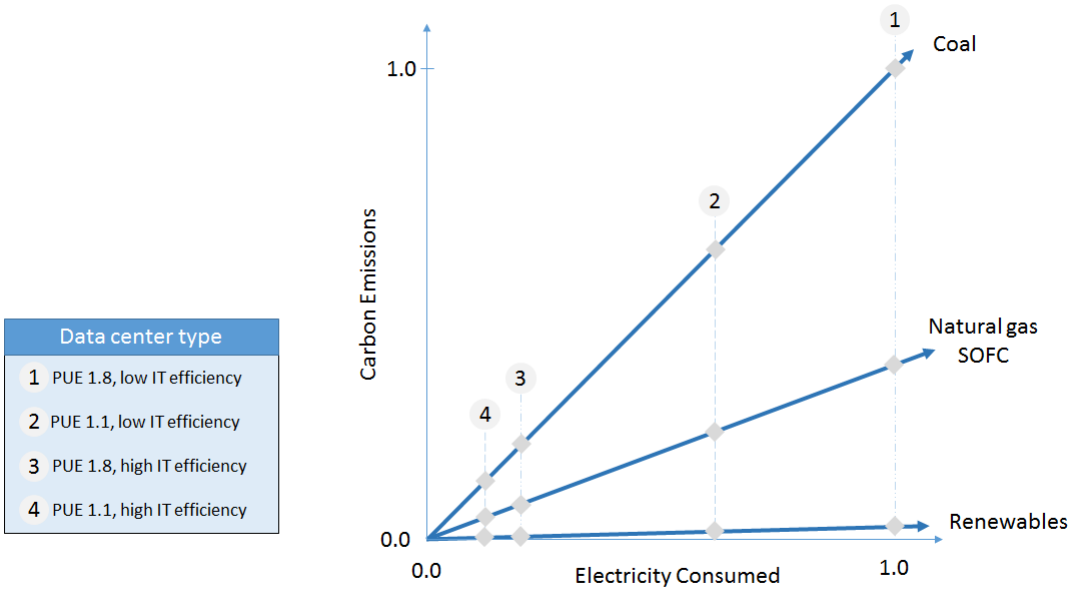
PUE is a good metric for the facilities side of the data center. It allows engineers to measure the impact of changes they make to the infrastructure, such as upgrading to a higher-efficiency cooling system or increasing the voltage to the rack. It's also a good yardstick to use when measuring improvements you make to an existing data center. For example, moving your data center from a PUE of 1.7 to 1.6 represents

an increase in efficiency. However, PUE says nothing about IT efficiency. You should be careful not to use PUE as the single global measure of data center efficiency.

Putting it all together

Low-carbon power, IT efficiency, and data center efficiency are all components of the green data center. Each of the four data centers in the figure below performs the same amount of computing but with different carbon emissions. The highest emission scenario is scaled to 1.0. The vertical axis represents the amount of carbon emitted in each scenario. You can see the dramatic reduction in carbon emissions when greener sources of energy are chosen. The horizontal axis is the amount of electricity consumed by a particular data center. You can see how carbon emissions can be reduced by improved IT efficiency and also by improved data center efficiency. For more information on the study that produced this data, see "Characteristics of low-carbon data centres."

Comparison of four types of data center under various power-source scenarios



Source: GreenM3, Gigaom Research

Options for going green

The Cisco Global Cloud Index predicts nearly a fourfold increase in traffic by 2016. It forecasts that:

- Annual global data center IP traffic will reach 6.6 zettabytes (10^{21}) by the end of 2016. By 2016 global data center IP traffic will reach 554 exabytes (10^{18}) per month (up from 146 exabytes per month in 2011).
- Global data center IP traffic will nearly quadruple over the next five years. Overall, data center IP traffic will grow at a compound annual growth rate (CAGR) of 31 percent from 2011 to 2016.

Forward-thinking CTOs are trying to balance this growth in data center traffic and its associated energy requirements with ways to comply with government regulations, protect themselves from volatile energy costs, and promote sustainable practices. Most importantly, they know that increased IT efficiency and data center efficiency coupled with stable (and reasonable) energy prices mean lower costs. In turn, lower costs can increase revenue through savings and by having more money to invest in increased compute power.

What are some options for achieving these goals? Two choices are to build a new data center or refurbish an existing one. Two other possibilities are co-location and the cloud.

Building or refurbishing a data center

If your company has the organizational and engineering skills on hand as well as the financial wherewithal, you can, of course, build your own data center or refurbish one so that it meets your environmental criteria.

If you are looking for best practices and guidance, companies such as Microsoft and Google have been generous in sharing what they've learned by building their own green data centers. One short paper, "Microsoft's Top 10 Business Practices for Environmentally Sustainable Data Centers," discusses specific ways to reduce energy consumption, waste, and costs while increasing efficiency and return on investment (ROI).

Google has published many articles on its green philosophy. For example, there are "Efficiency: How we do it," which discusses its approach to PUE, and "Google Green: The Big Picture,"

which gives an overview of its approach to incorporating sustainable practices into all aspects of its business.

Another good resource is the Lawrence Berkeley National Laboratory's website for building energy-efficient data centers. It has an extensive set of scientifically vetted recommendations and best practices. Many of these apply to new facilities only, but others make sense for existing facilities.

Of course, even if you can build or refurbish a data center, that may not be your best option. For example, you may be located somewhere with high and volatile energy costs or where you can't expand quickly enough to accommodate your anticipated growth. In these situations, you may still want to consider either a green co-location site or a green cloud service.

Choosing a green cloud

The Open Data Center Alliance (ODCA) develops open, interoperable standards for the cloud. In 2011 it joined with The Green Grid to use its work to provide customer-focused cloud usage models. The ODCA's report "Carbon Footprint and Energy Efficiency Rev. 2.0" (PDF) reflects this partnership. The report describes what cloud subscribers should expect from their providers. Here are some questions to ask:

- Does the provider supply energy-use and carbon-emission figures? If not, subscribers should demand them. Cloud subscribers should accumulate and aggregate the data for carbon reporting. (The ODCA notes that using and reporting these figures will likely become a necessary part of doing business in the near future.)
- Cloud subscribers only use a portion of a cloud provider's data center and may also use multiple providers. Does the cloud provider have methods for allocating carbon usage to specific cloud subscribers? In turn, cloud subscribers should have methods for aggregating the amount of carbon used from various cloud providers and from in-house production.
- Can the provider describe any approximations that it uses in its calculations?
- What is the PUE of the data center?
- What kind of energy monitoring and data center infrastructure monitoring (DCIM) does the provider use to ensure IT efficiency? Good monitoring is vital to containing energy costs. One

published source writes that "energy savings from well-managed data centers can reduce operating costs by as much as 20%."

- The ODCA report quotes research by Pike Research from 2010 that reports the adoption of cloud computing could lead to a 38 percent reduction in worldwide data center energy expenditures by 2020.

Choosing a green co-location center

Because of scale and utilization rates, cloud computing may not be for you. Co-location might be the most effective solution. If you choose co-location, you should ask a number of questions to ensure your goals toward sustainable IT are met. Some of them are the same questions you would ask a cloud provider.

In many situations, companies provide their own equipment if they are using co-location. If this is the case, you can make sure that your IT equipment is used efficiently. Remember that improving IT efficiency has a multiplicative benefit: Better IT efficiency means less equipment to cool. Improvements to IT efficiency can play a big part in a carbon-reduction scheme.

Here are some questions you can ask:

- Does the co-location center use prefabricated modules in its building design? These cannot solely make a data center more efficient, but your provider will be able to accommodate you if you need to expand quickly.
- Does the co-location center use hot-aisle containment to ensure that even cooling is delivered to the IT equipment with minimal wasted energy?
- Where is the provider located? Is it taking advantage of any environmental advantages the location offers?
- How is the temperature of the site controlled? Does the provider perform an annual assessment of computational fluid dynamics to test for hot spots and general floor-cooling effectiveness?
- Is the service located in a business-friendly location, with a well-educated workforce that speaks your native language?

- Do you work well with the service provider? Does your IT staff get along with its staff so that problems can be quickly resolved?
- Does the co-location provider have any awards or certifications that show its commitment to sustainable practices? For example, does it have a Green Globes award from the Green Building Initiative? Does it have LEED certification? If so, what level?

Choosing the right application

If you've decided to use green computing resources from a cloud provider or from a co-location service, you still have an important issue to consider: What applications are best suited for green environments? You need to have accurate profiles of your applications in order to make the right choices.

An important issue is latency, which is the amount of time delay an application experiences. Interactive applications, for example, require low latency because users expect quick responses. Stock-trading programs are a good example of applications that demand near-zero latency. Low-latency applications need to be located in data centers that are physically close to their markets.

Here are other questions to ask yourself, whether you are opting for the cloud or for co-location.

- Does the application need to peer with another company, such as Facebook or Google? If so, this, along with latency, can shape your decision about where you can locate it.
- What are the business requirements? What do you expect in terms of performance, availability, reliability, number of users, and response times?
- What is the application architecture? Multitier applications that have a great deal of communication among layers may not do well in the cloud because of latency issues. This is also true of chatty applications, where the application makes multiple calls to the database server for each user request.
- What is the current and predicted application workload? What are the current and predicted capacity requirements of the workload?
- What is the application's resource consumption rate? Resources include memory utilization, CPU utilization, data transfer and bandwidth, and storage requirements. In particular, make sure you understand your input/output (I/O) requirements for both network and storage. This metric is often forgotten. Today's servers, which have many cores and large amounts of memory, can be constrained by the design of the network.
- If you are looking at existing applications, try to collect metrics about what it costs to run the application now, so that you can compare it to prices that providers are offering. Look at a variety of factors such as the costs of running any supporting applications such as the operating system,

the cost of support personnel, the cost of cooling and power, and the cost of server and storage hardware.

Going green with high-performance computing

High-performance computing (HPC) was once used exclusively by the scientific community. This is no longer true. More and more businesses require sophisticated analytics, modeling, and simulations. This increasing demand for HPC applications can make life difficult for data center managers because of the amount of energy these compute-intensive applications consume. Of course, as applications grow, the servers they run on need a commensurate amount of cooling, which adds cost.

There is, however, a silver lining. Unlike transaction-based applications, HPC applications use batch processes. They do not have low-latency requirements, which means you can be more flexible in where you locate them. You can look for providers that offer low-energy costs and renewable sources of power.

Green HPC case study: deploying a risk-management platform

RMS is a risk-management solutions company that helps the insurance and financial industries manage risk by providing sophisticated catastrophe-modeling services. Initially sold as a software package, the company found that as it planned to deliver a new generation of modeling and analytics capabilities to the market, the overall infrastructure requirements became too overwhelming for all but its largest clients. So it decided to provide an extensive suite of modern features and capabilities by using a cloud-based delivery model called RMS(one), which is currently scheduled to go live in mid-April 2014.

Ron Stein, the director of Product Marketing for the RMS(one) platform, describes the product as largely being a SaaS offering, with some PaaS characteristics such as the ability for clients and third parties to add their own models, analytics, and applications and to access and invoke them seamlessly. The majority of RMS' customers are on the North American eastern seaboard, Europe, and Caribbean.

Stein said that because of the unique analytic capabilities of RMS(one), it wasn't possible to use an off-the-shelf PaaS vendor. The company decided that building its own infrastructure made sense but that building its own data centers didn't.

Initially RMS will use three data centers that are identical to one another in all but scale. Two are production data centers, which are located in Europe and North America. The third is currently used to showcase beta versions of the product. Once RMS(one) goes live, this site will be transitioned to being a dedicated disaster-recovery data center. It will also be used to perform load and stress tests and to act as the staging environment for deploying to the production data centers. After an extensive search, RMS chose Datapipe as its provider for this site. Datapipe's Stratosphere HPC green cloud platform is based in the Verne Global facility in Iceland.

Stein said that Iceland was attractive for a number of reasons. One was the growing sensitivity of customers, particularly in Europe, to sustainability. Another was the EU directive to reduce carbon emissions. A third reason was, of course, the cost of energy. Iceland won all because the power costs are low and predictable and because the power sources are green. These issues were particularly important because the footprint of this site is the largest, due to it being used for disaster recovery, load and stress testing, and staging.

Latency also figured into RMS' calculations. Iceland is strategically located near the eastern seaboard of North America, Europe, and the Caribbean. Streaming information from the two production sites to the Iceland site happens within a reasonable time frame, and there is excellent bandwidth availability.

Green HPC case study: automotive engineering

The BMW Group is well-known for its commitment to green manufacturing processes. Since 2005 it has been named by the Dow Jones Sustainability Indices as the world's most sustainable automobile manufacturer. The report looks at industry-specific criteria such as having clean production processes, developing fuel-efficient vehicles and vehicles that use alternative energy sources, and using clean recycling practices. BMW has extended its commitment to sustainability to include its IT practices as well.

To design the i3, its new urban electric car, BMW has used HPC cells for crash simulations, fluid dynamic modeling, and computer-aided design and engineering (CAD and CAE). In keeping with its belief in the importance of sustainability, BMW wanted these compute-intensive activities to occur in a data center that offered a carbon-free alternative.

Mario Mueller, BMW's VP of IT Infrastructure, said the company needed to investigate a number of possibilities. Initially BMW thought it would try to add another HPC cell to the BMW data center, but it couldn't find the room. Another option was to refurbish an existing data center, but it was too expensive. Next BMW tried to find additional data center space elsewhere in Germany, but there were problems: None of the data centers candidates were carbon-free, none of them could promise that BMW would be able to expand rapidly in the near future, and the cost of power couldn't be guaranteed to remain stable.

At that point BMW began to look further afield. Its most important criteria were:

- The data center had to be carbon-free.
- The data center had to be able to accommodate rapid expansion in a short amount of time.
- Energy pricing had to be stable.
- Service-level agreements needed to be clear.
- The company had to value good customer relations and be willing to work with BMW to resolve any problems.

After investigating many candidates, BMW decided to use the Verne Global facility in Keflavik, Iceland, as its co-location provider.

Along with zero carbon emissions, Verne Global satisfied BMW's other requirements, such as stable pricing and the ability to expand quickly. Finally, Mueller stressed the good working relationship BMW has with Verne Global. "Trust isn't given for free. You have to work hard to get it. You need to earn it with real actions, not just PowerPoint presentations." He added that there were certainly issues, but everyone worked together to solve them and that Verne Global's commitment to customer satisfaction was important to BMW.

Currently BMW has five HPC cells at the Verne Global site, which comprise approximately 10,000 cores on about 650 servers. They run 25,000 jobs per month, and a job takes an average of 10 hours to complete. The pipe between Munich and Iceland is large enough so that BMW has no trouble transferring information between the two locations. The servers run 90 percent of the time; there is no idling. In addition, there have been no power outages.

Mueller said that by using Verne Global, BMW saves about 3,570 tons of carbon dioxide annually. This is equivalent to driving around the planet more than 600 times.

BMW has big plans for the future. Along with his other duties, Mueller is the chairman and secretary of the ODCA, and his development plans support the precepts of that organization by also expanding into the cloud.

In addition, Mueller says that BMW plans to add another five HPC cells to the Verne Global facility in 2013 so it can do big data and analytics. This includes BMW's Connected Drive technology, which uses real-time traffic data, cloud-based voice control, and other features connected by an LTE data connection. In an interview on Slashdot, "BMW's Connected Cars Force New Data Center," Mueller said that today just 1 million cars are connected, with data requirements in the hundreds of megabytes. By 2018 some 10 million BMWs will be connected, asking for and receiving more than 1 terabyte of data every day.

When asked about what he's learned from this project, Mueller said, "Everything is possible, don't say no at the beginning. Take your risk. Change the way you work and succeed. Don't talk too much, just do it."

Key takeaways

Forward-thinking CIOs are well aware of the pressures that come from the increased demand for data center services, rising energy costs, and regulations that limit carbon emissions. Here are some key points to remember.

- Government regulations that limit carbon emissions are being enacted globally.
- Customers are becoming more aware of how data centers implement sustainable practices.
- Fossil fuel prices are high, difficult to predict, and produce carbon emissions.
- Using lower-cost, renewable energy and following green practices can reduce costs. Savings can be translated into greater capacity, which generates revenue.
- Green data centers reduce their carbon footprint as much as possible. They try to use renewable energy and efficient IT equipment and maintain a low PUE.
- Possibilities for going green include building a new data center, refurbishing an existing data center, choosing a green cloud provider, or choosing a green co-location provider.
- Evaluate your applications to see which can be located in a green data center. Develop detailed profiles of each application's requirements.
- HPC applications are becoming common in all types of businesses. HPC applications are often less sensitive to location because they don't require low latency. Consider green data centers that offer low and predictable energy costs.

About Dave Ohara

Dave Ohara's corporate career began at HP and continued with Apple and Microsoft. After over 20 years of developing technology, he switched to the data center industry to better understand the challenges in going to market with online services. He now consults with companies in the data center community. He holds a degree in Industrial Engineering and Operations from the University of California, Berkeley.

About Gigaom Research

Gigaom Research gives you insider access to expert industry insights on emerging markets. Focused on delivering highly relevant and timely research to the people who need it most, our analysis, reports, and original research come from the most respected voices in the industry. Whether you're beginning to learn about a new market or are an industry insider, Gigaom Research addresses the need for relevant, illuminating insights into the industry's most dynamic markets.

Visit us at: research.gigaom.com.