

Solar-Powered Cloud Computing Datacenters

Laura Hosman, *Illinois Institute of Technology*

Bruce Baikie, *Green WiFi*

In the developing world, reliable and affordable electricity isn't always available, presenting challenges for the Western “bigger is better” datacenter model. A small-scale, cloud computing datacenter, however, could leverage recent technological breakthroughs to instead rely on solar energy.

Most developing countries face financial, geographical, infrastructural, and power constraints that have largely prohibited local development and adoption of IT-related activities. Consequently, datacenters reside primarily in Western markets, where electric power is reliable and more affordable. Yet emerging economies stand to gain a great deal in terms of socioeconomic growth and development from a robust Internet connection and from telecommunications and related technology advances. Unfortunately, information and communications technology (ICT)-related development can't move forward unless the energy to power the technology is sufficient, affordable, and

reliable, and the ICT equipment is engineered to serve this market.

The picture is bleak for datacenters built in developing regions following the Western model of high-energy consumption. Electric grid outages are common, and datacenters rely on diesel-powered generators. With ever-increasing diesel fuel costs and unregulated pollution levels from such generators, a new approach is urgently needed.

Tinkering at the edges of technological improvement while retaining the “bigger is better” mindset is neither sufficient nor sustainable. Rather than simply rearranging the deck chairs—or racks full of servers and storage—we propose a fundamental reconceptualization of the datacenter—from macro to micro. We also need

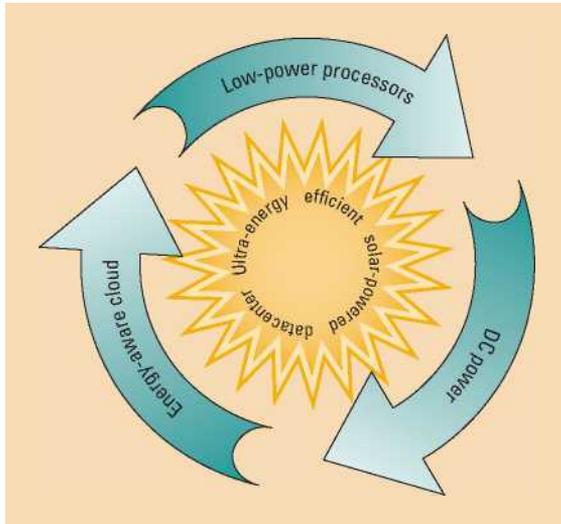


Figure 1. A rethinking of the datacenter combines three innovative design approaches to reduce energy needs. The lower energy usage makes renewable energy options previously not viable—such as solar photovoltaics—realistic, leading to a new generation of green, solar-powered cloud computing datacenters.

to rethink constituent datacenter elements—their processors, storage, energy, design, size, logistics, configuration, networking technology, hardware, and software.

We propose a specialized small-scale datacenter based on cloud computing. By combining several recent technological breakthroughs—namely, low-power computing platforms, energy-aware cloud computing, and DC power distribution—we can use solar energy to power such a datacenter (see Figure 1). This ultra-efficient approach moves computing and storage closer to the user, increases security for local data, and reduces energy costs and the environmental footprint.

The Looming Energy Crisis

In 2007, the Environmental Protection Agency reported that US datacenter energy consumption had doubled from 2001 to 2006 and was expected to double again over the next five years.¹ As companies began to face the painful reality of rising energy bills and the environmental effect of their energy consumption, they launched a wave of green datacenter initiatives. Unfortunately, the fixes proposed were simply band-aids that didn't address the fundamental problems.

Datacenters continue to consume copious amounts of power. Furthermore, they turn the majority of that power into unwanted heat

and still generally maintain the big-box, large-footprint datacenter model.² At the same time, energy costs continue to rise, along with our understanding of the environmental effects of our carbon footprint. How did we get to this point?

First, energy has been too cheap in the Western market. It has been readily available and subsidized in numerous ways in the US and other Western nations, including through the non-taxation of pollution and subsidization of energy companies' activities. This situation might change in the future, but to date, cheap and readily available energy has been assumed in the design, layout, and entire approach of the Western-style datacenter model. This is one reason why merely tinkering at the edges of the current model has proven insufficient.

Second, Western-style datacenters and Internet-hosting centers have been built with computing power and 24/7 operations as the top concerns. They're seen as having mission-critical responsibilities that are too important to take risks with in any meaningful way. As technology companies have produced increasingly powerful servers designed to be available and running 24 hours a day, the datacenters' mandate has been to play catch-up in terms of figuring out how to cool the servers and provide power for them at any cost.

Finally, datacenters have continued down this "bigger is better" path because energy responsibility hasn't been tied to those in charge of the datacenters. In 2008, *InformationWeek* reported that 78 percent of IT organizations weren't responsible for tracking—or managing—datacenter power usage, while 75 percent of them received no compensation whatsoever for reducing their power consumption.³ If more than three quarters of corporate America's IT professionals have no direct incentive to be concerned about their companies' energy costs, then we're unlikely to witness great strides in innovation or "outside the box" thinking.

We propose planning for a scenario in which energy isn't dependable, readily available, or cheap; grid power—smart or otherwise—isn't available; energy costs are a make-or-break expense instead of a bothersome outlay; and every company is responsible for reducing its environmental impact. For much of the developing world, this isn't just a scenario—it's reality.

Creativity under Constrained Conditions

Many firms that target the developing world have recently discovered that their former strategy of simply marketing their Western-targeted goods to the elites in poorer countries isn't a sustainable business model. This model fails to target the masses of poorer consumers who have different wants and needs from those who take for granted things like cheap, reliable electricity. Similarly, those who are accustomed to living in, inventing for, and marketing to those in Western countries can only guess at these wants and needs. Many Western multinationals have thus had business ventures fail in developing countries.⁴

More recently, a handful of multinational corporations have discovered that to reach the masses of consumers who live under conditions of constraint, they need to more fully comprehend the reality of their customers' circumstances. Being creative under conditions of constraint that aren't part of the everyday experience of most Western IT professionals can stimulate creative, outside-the-box thinking. Business strategies that leverage—rather than just attempt to overcome—weaknesses of developing-country markets are more likely to succeed in both the developing and developed world.⁴ By addressing issues in the context of developing-country constraints, these creative solutions often return to the developed world with added benefits, such as lower-priced products that can be sold to both developing- and developed-world markets at previously unimaginable price points.

The rising popularity of netbooks is a prime example. The nonprofit, socially and educationally oriented One Laptop per Child (OLPC) project, with its various hardware innovations, open source operating system, small size, and \$100 target price point, was perceived as a threat to for-profit PC makers. Out of fear that they might lose a prized foothold in emerging markets, PC makers stripped down, reinvented, and lowered the cost (and profit margins) of the traditional laptop, resulting in the low-end “netbook.”⁵ The netbook proved successful with both developing- and developed-world consumers and has been the lone bright spot in the PC industry during the global recession.⁶ A similar case can be made regarding innovations surrounding the tablet computer. This is the essence of the “trickle up”

effect that various technology-related corporations, such as Microsoft and Nokia, are employing as they rethink their business models and how they produce, invent, design, and market their products and services.⁷

Furthermore, change generally only comes with sufficiently high pressure (or pain) points. So the latest technologies might gain more rapid acceptance in the developing world, because they're not replacing anything. This makes the developing world a highly fertile breeding ground for the invention and implementation of many creative and innovative technologies.

Low-Power Computing Platforms

The first technological innovation we highlight focuses on the viability of small cloud computing farms deployed in small-scale datacenters using low-power processors, such as the Intel Atom, ARM Cortex, and Via Nano. These CPUs, originally designed for netbooks, require only 2 to 10 watts, versus the 80 to 160 watts typically required of processors used in datacenter servers. These low-power processors also have quiescent states that consume little energy, unlike the standard datacenter CPU, which consumes 50 percent of its peak energy even while idle.⁸

These quiet states were designed to let netbooks hibernate but quickly wake and provide on-demand processing. This model is perfect for cloud or specialized computing.⁹ By using less energy and producing little waste heat, datacenters can improve the efficiency, reliability, availability, and usability of computing systems. Although using low-powered CPUs sacrifices some raw computing performance, the integration of software, CPU clusters, and solid-state storage enables overall computing power that can more than support typical Internet Web-based services and applications while providing more computing performance per watt of energy consumed.

Both industry and academia are contributing to research and innovation in this area. Microsoft Research has an organizational division called Cloud Computing Futures (CCF), focused on reducing the operational costs of datacenters and increasing their adaptability and resilience to failure. Their first project aimed to lower hardware costs, power consumption, and the environmental effect of datacenters using low-power processors in the datacenters. They found that “these processors offer substantial fractions (33 percent to 50 percent) of the performance of the high-performance processors used in Microsoft datacenters but consume a disproportionately smaller amount of power (5 percent to 10 percent).”¹⁰ Similar results were obtained during comparable studies performed at Carnegie Mellon University,¹¹ the Green-scale Center for Energy-Efficient Computing at the University of California at Santa Barbara (<http://iee.ucsb.edu/greenscale>), and the ARM Corporation.

Although the modern datacenter has undergone a major increase in size, scope, and power usage—including dramatic changes in demand, the type of application processing required, and the economics of operation—the actual server has remained architecturally unchanged over the past 12 years, and server manufacturers have made little progress in reducing their servers' energy consumption. The same general-purpose servers are designed to support everything from specialized workloads to general Web servers, which typically consume the same amount of energy regardless of workload. Instead of just modifying existing server architectures to redesign power supplies or improve airflow,¹² we need a new server architecture for low energy usage. We propose a green technological approach: combining low-power CPU design, processor

virtualization, and networking to create a new server architecture optimized for an ultra-energy-efficient infrastructure but still capable of performing at the level required for Internet and cloud-related services.¹³

A server specifically designed for low energy use could use 8 percent of the electrical power required by standard servers, at 28 percent of the cost, while delivering more than 60 percent the processing power of standard servers. When clustered, these new servers with low-power processors can deliver more computing power, with less energy, and at a lower cost.¹³

Energy-Aware Cloud Computing

Cloud computing has been transformational in that it lets the computing cloud grow and shrink based on application demand. However, a more valuable use would be to grow and shrink the cloud based on available energy. In normal off-grid solar photovoltaics (PV) systems, we design for the worst case, which increases the total system size for periods of little or no sunshine. When we “precision manage” energy usage in cloud computing—what we term “energy-aware cloud computing”—then we can make an educated trade-off by prioritizing energy savings instead of operating on a smaller, more cost-effective solar PV system. Using energy-aware cloud computing also lets us manage the computing load during extreme weather conditions by shutting off nonessential activities.

Again, this is an educated, financial trade-off: we can install either a large and expensive PV system and batteries to handle worst-case scenarios, or a smaller, cost-appropriate PV system to handle normal, everyday loads while accepting that there might be times when essential services must be prioritized. Because developing-world resources are scarce, to address the large initial capital investments associated with solar technology, we're harnessing software to optimize circumstances and manage the computing load when the weather is less than ideal. Because of this improved energy management, a smaller solar system can be used, which then saves on the initial capital investment.

This is where the reconceptualization of the Western mindset comes in: prioritization can become necessary under circumstances of constraint. For example, if an annual two-week

sandstorm or monsoon blocks the sun so that little photovoltaic energy is generated, then the energy-aware cloud computing software can shrink the cloud and run only critical services. In other words, the possibility exists for enabling the cloud software to address issues of unpredictable or constrained energy supplies. The interface between the energy-management control system and the cloud computing environment is critical for identifying and monitoring how much energy the cloud computing can use (see Figure 2).

Energy-aware cloud computing also works during normal datacenter operations to reduce idle time on servers, decreasing costs and the environmental footprint. Traditional, silo-oriented architectures use noninterchangeable server, storage, and network elements, with each application component hosted on its own set of components. Each set of servers, storage, and network elements is traditionally sized to handle the worst-case-scenario workload with no possibility of resource sharing between silos. The result is low utilization rates that are no longer affordable because of the excess power used and heat generated. Virtualization technologies let datacenter servers consolidate multiple workloads onto a single server to more efficiently use each server and reduce the overall power demand.¹⁴

Virtualization technologies help datacenters economize in two ways. First, they can be used to consolidate multiple workloads onto a smaller number of servers, raising utilization levels and increasing energy efficiency. Then, the cloud-virtualized datacenter can choose which workload to run and which resources to use. This is much more efficient, because cloud virtualization technologies facilitate dynamic resource allocation within a single server, making it possible to move workloads from server to server.¹⁵ Using these advanced virtualization features in both the computing cloud and the interface to the power generation subsystems reduces server power usage by 80 percent.

DC Power Distribution

DC-only power distribution in the datacenter supply can eliminate the need for inefficient AC-DC transformers in power supplies, which reduces not only the cooling costs but also the total datacenter footprint.¹⁶ DC power could

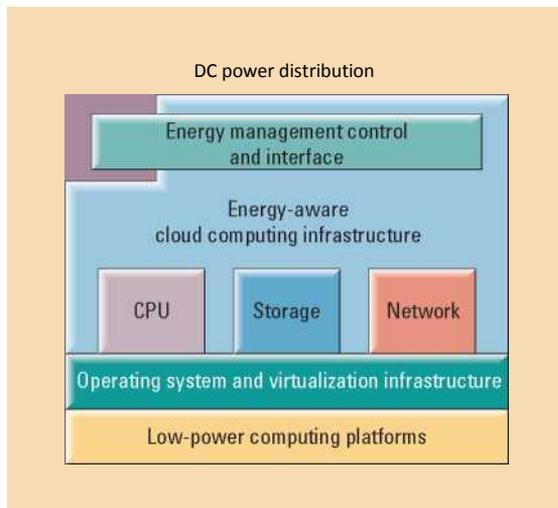


Figure 2. The interface between the energy management control system and the cloud computing environment identifies and modifies how much energy the cloud computing can use.

play an important role in ultra energy-efficient datacenters, because many alternative energy sources—such as solar and fuel cells—naturally produce DC power.

Telecommunication carriers have decades of experience in powering central switching offices using DC power and have thus perfected the reliability and efficiency of DC power-distribution systems and DC-powered servers and network equipment. DC power can significantly simplify datacenter power distribution. Because multiple DC power sources can be joined using only a diode system, expensive technology, such as static-transfer switches, can be eliminated. DC-only power can also save both power and money.

Traditional datacenter power architectures convert power back and forth between AC and DC at least three times (see Figure 3). Incoming AC power is converted to DC power to charge the uninterruptible power supply (UPS) batteries. DC power from the batteries is converted to AC and is distributed to racks and then to servers. Finally, AC power is converted in each server's power supply, first typically to 380 volts DC and then to the various voltages needed within the server. Each conversion between AC and DC introduces inefficiencies and generates heat (which must be cooled).

Converting datacenters to run on DC power only can reduce server power supply consumption by 25 to 28 percent.¹⁷ This approach can

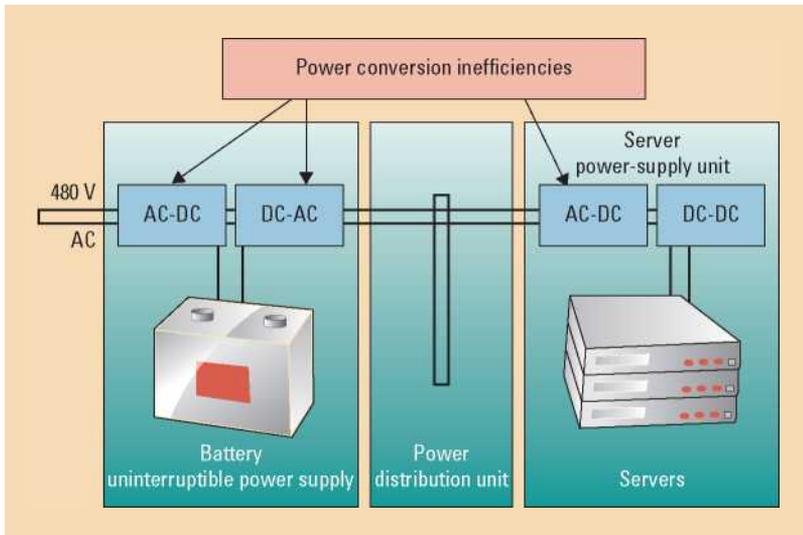


Figure 3. Traditional datacenter power architectures convert power between AC and DC at least three times. Converting datacenters to run on DC power only can reduce server power supply consumption and cooling requirements.

lower the energy needs and reduce complexity (while increasing reliability) by eliminating a layer of components that can fail.

Cooling of the Datacenter

The Green Grid Consortium estimates that even in an efficient datacenter, 42 percent of the total power is used by the air conditioning systems.¹⁸ Two conventional methods for addressing this waste of energy are to invest in more efficient mechanical cooling systems or to move the datacenter to colder parts of the world. Our innovative approach is twofold: eliminate the majority of the heat so we can use ambient cooling (augmented with spot rack cooling), and run the datacenter at higher temperatures.

We can trace ambient cooling back for centuries in the Middle East and across Africa. People from these regions harnessed the cooler temperatures found underground to provide cooler air to vent aboveground structures, as in a thermal chimney. Using ambient cooling in the datacenter context has a significant upside in terms of energy savings, offering another example of how developing-country constraints can produce solutions useful around the globe.

Employing ambient cooling as the main method of heat dissipation lets the datacenter operate at higher temperatures. This warmer approach to datacenter operations has recently found support in the mainstream datacenter industry.

Experts working for major industry players, including Microsoft, Google, and HP, have begun to advocate for operating datacenters at higher temperatures.¹⁹

The critical factor to consider when raising the datacenter temperature is the risk of IT equipment component failure. Our approach to mitigating this risk is again twofold. First, we suggest using servers, storage, and network equipment specifically designed to operate at higher temperatures. Second, we would increase temperature control precision through spot rack cooling—not by cooling the entire datacenter. Still, the less heat that must be removed, the better. Ambient cooling exploits the combined

effect of the previously described innovations in terms of reducing heat using efficient, lower-power servers with quiescent states, using virtualization and cloud computing, and avoiding the multiple heat-generating transformations of energy currents. Reducing heat in these ways obviates the need for the energy and heat produced by constantly running HVAC (heating, ventilation, and air conditioning) systems, thereby eliminating the expense, expertise, and energy required to operate them in the first place as well as a further layer of non-fail-safe technology.

The developed world has much to learn from developing countries, and when products are designed realistically, taking into consideration the constraints facing the latter, both worlds benefit from greater innovation, less energy use, and a reduced carbon and pollution footprint. The three technologies we've discussed—low-power processors, energy-aware cloud computing, and DC-power distribution—are feasible today in areas where power reliability and availability hold back growth and development. Compact in size and energy use—but not in processing power—small solar-powered cloud computing datacenters can be the next generation of growth in our global Internet infrastructure fabric.

References

1. "Data Center Report to Congress FINAL 7-27-07," Environmental Protection Agency, 2 Aug. 2007; www.energystar.gov/ia/partners/prod_development/downloads/EPA_Datacenter_Report_Congress_Finall.pdf.
2. J.G. Koomey, "Worldwide Electricity Used in Data Centers," *Environmental Research Letters*, vol. 3, no. 3, 2008; doi:10.1088/1748-9326/3/3/034008.
3. A. Wittman, "The Cold, Green Facts," *Information Week*, 1 Sep. 2007; www.informationweek.com/the-cold-green-facts/201803326.
4. T. London and S.L. Hart, "Reinventing Strategies for Emerging Markets: Beyond the Transnational Model," *J. Int'l Business Studies*, Sept. 2004, pp. 350-370.
5. K. Kraemer, J. Dedrick, and P. Sharma, "One Laptop Per Child: Vision vs. Reality," *Comm. ACM*, vol. 52, no. 6, 2009, pp. 66-73.
6. B. Stone and A. Vance, "\$200 Laptops Break a Business Model," *New York Times*, 26 Jan. 2009; www.nytimes.com/2009/01/26/technology/26spend.html?_r=0.
7. R. Jana, "Innovation Trickles in a New Direction," *Bloomberg BusinessWeek*, 11 Mar. 2009; www.businessweek.com/magazine/content/09_12/b4124038287365.htm.
8. Y. Watanabe, J. Davis, and D. Wood, "WiDGET: Wisconsin Decoupled Grid Execution Tiles," *Proc. 37th Ann. Int'l Symp. Computer Architecture (ISCA 10)*, ACM, 2010, pp. 2-13.
9. V.J. Reddi et al., "Web Search Using Small Cores: Quantifying the Price of Efficiency," white paper, Microsoft Research, Aug. 2009; http://research.microsoft.com/pubs/101821/search_hardware_techreport.pdf.
10. "TechFest 2009 Demos," Microsoft Research, 2009; <http://research.microsoft.com/en-us/events/techfest2009/demos.aspx>.
11. D. Andersen et al., "FAWN: A Fast Array of Wimpy Nodes," Carnegie Mellon Univ. and Intel Labs, Oct. 2009; www.cs.cmu.edu/~dga/papers/fawn-pdl-tr-08-108.pdf.
12. P. Bemis and L. Marshall, "Improving Data Center PUE Through Airflow Management," white paper, Applied Math Modeling, Feb. 2010; www.coolsimsoftware.com/LinkClick.aspx?fileticket=KE7C0jwcFYA%3D&tabid=65.
13. A. Vance, "Big Web Operations Turn to Tiny Chip," *New York Times*, 20 June 2010; www.nytimes.com/2010/06/21/technology/21chip.html.
14. L. Grit et al., "Virtual Machine Hosting for Networked Clusters: Building the Foundations for 'Autonomic' Orchestration," *Proc. 2nd Int'l Workshop on Virtualization Technology in Distributed Computing (VTDC 06)*, ACM, 2006, p. 7.
15. A. Berl et al., "Energy-Efficient Cloud Computing," *Computer J.*, vol. 53, no. 7, 2010, pp. 1045-1051.
16. M. Ton, B. Fortenbery, and W. Tschudi, "DC Power for Improved Data Center Efficiency," Lawrence Berkeley National Laboratory, Mar. 2008.
17. S. Bergqvist, "Energy Efficiency Improvements using DC in Data Centers," master's thesis, Uppsala Universitet, 2011; www.netpower.se/documents/Energy%20efficiency%20DC%20in%20data%20centers.pdf.
18. "Guidelines for Energy-Efficient DataCenters," white paper, The Green Grid Consortium, 16 Feb. 2007; www.thegreengrid.org/~media/WhitePapers/Green_Grid_Guidelines_WP.pdf?lang=en.
19. R. Miller, "Google: Raise your Data Center Temperature," Data Center Knowledge, 14 Oct. 2008; www.datacenterknowledge.com/archives/2008/10/14/google-raise-your-data-center-temperature.

Laura Hosman is an assistant professor of political science and associate chair of the Department of Social Sciences at the Illinois Institute of Technology. Her current research focuses on the role of information and communications technology (ICT) in developing regions, particularly in terms of its potential effects on sociocultural factors, human development, and economic growth. Hosman has a PhD in political economy and public policy from the University of Southern California. She's a member of IEEE and is a recent recipient of an IEEE Foundation grant for an educational solar energy project in Haiti. Contact her at laurahosman@gmail.com.

Bruce Baikie engineers and implements solar-powered ICT4D projects and heads Green WiFi, a solar-powered data and Wi-Fi network technology company. His research interests include wireless networking, eco datacenters, DC power, and solar power. Baikie has a BS in mechanical engineering from Michigan Technological University. Contact him at bruce@green-wifi.org.