Virtualization and its impact to the process of transforming to "green" data centers

East Carolina University

Paola Stone Martinez

July, 2014

Abstract

When we hear “green” datacenter, everybody associates it to a friendly environmental building, but there are a lot of steps and contemplations before a location can be considered “green”. As new technologies are developed, the demand for different services grows as well. Sometimes the technology changes too quick we cannot physically adapt to it. Example of it is the digital era we are leaving on. The amount of data that we are producing is very big, at the beginning adding more physical resources was the best solution. Nowadays, adding more equipment is very hard not only because of space concerns, but also because of the amount of energy required for these systems to work. Data centers and cloud computing complement each other. Technologies are shared between them and the results are great. One of these technologies is virtualization. This paper is divided in three sections. First, I start explaining about data centers, what they are, how they can become “green” buildings, and about the main “green’ certification systems. Then, I cover some basic information about virtualization, as well as its advantages. Finally, both concepts are put together. The results of some experiments between physical and virtualized servers are presented in order to understand how actually virtualization affects the energy consumption in data centers.

Data Centers

The fast growth that IT technologies have experimented in the latest decade has benefitted society not only by introducing new ways of communication, but at the same time, new processes that have facilitated the way in which we do daily activities. Most of these technological advances are hosted in remote computer systems carefully configured and architected. The equipment necessary to offer these services are located in buildings called data centers.

In the past, the demand for this type of buildings was not as high as it is today. During this digital era, we have produced massive amounts of data that has to be saved somewhere, and more important, available to share at all time. Networking has also played an important role in the changes of the traditional dynamic of computing services, concepts like cloud computing require equipment to be always available, while being private, secure, and redundant. Because of this increasing need, data centers started to be built all around. At the beginning there were no too many requirements, and if there were, they were mainly related to safe construction or to optimize the performance of the equipment. Nowadays, data centers host big quantities of powerful servers, switches, storage arrays, routers, as well as other devices needed to design the infrastructure and people to support it. The equipment power and cooling requirements plus the human needs, make traditional data centers significant contributors to the levels of environmental contamination.

In order to address this, a new concept has been introduced: "green data centers". The concept is derived from the green building certification systems which are a compilation of building standards and green building certification. The National Research Council (2013) affirms that "green building certification systems differ from building standards in that they

typically take a ‘whole building’ approach. They also provide a series of increasingly stringent levels of certification to measure the overall ‘greenness’ of individual buildings” (p. 27). In green buildings it is very important to pay important attention to the resources that are going to be needed to support the regular operations within the building, how they are processed, and disposed. This includes: water, cooling, and heating, among others. There are different green building systems, in the United States the two that are mainly used are: the US Green Building Council’s (USGBC’s) Leadership in Energy and Environmental Design (LEED) and Green Globes.

The USGBC’s defines LEED as “a green building certification program that recognizes best-in-class building strategies and practices” This system has five rating systems depending on the project type (para. 2-3):

- Building Design + Construction
- Interior Design + Construction
- Building Operations + Maintenance
- Neighborhood Development
- Homes

Data centers are included in the building and design + construction, and building operations + maintenance type. For both of them, there are guidelines to follow when constructing a building from scratch or when changing the design. In each type of project, buildings get credit according to: integrative process, location and transportation, materials and resources, water efficiency, energy and atmosphere, sustainable sites, indoor environmental quality, innovation, and regional priority credits. The level of the certification is based on the

amount of points obtained: from 40 to 49, Certified; from 50 to 59, Silver; from 60 to 79, Gold; and more than 80, Platinum (USGBC, 2014).

Green Globes is given by the Green Building Initiative (GBI). It is "a web-based program for commercial buildings that includes on-site building assessment by a GBI authorized third-party assessor" (Green Globes, 2103, para. 2). It includes four modules: Green Globes for New Construction (NC), Green Globes for Continual Improvement of Existing Buildings (CIEB), Green Globes CIEB for Healthcare, and Green Globes for Sustainable Interiors (SI). In order to get the certification the following environmental characteristics are considered: project management, site, energy, water, materials and resources, emissions, and indoor environment. (Green Globes, 2014, para. 2).

Virtualization

Even though it happened some years ago, I still remember an Operating Systems class I took, that was the first time I actually used a virtual machine. At that moment, I did not understand the potential this technology would represent later on the Information Technology (IT) world. The steps to create a Virtual Machine (VM) were very easy as we followed the teacher in the process of creating our Windows and Linux VMs. The process took few minutes, and after that, we had both VMs running different Operating System (OS) on the same physical computer, which at the same time had its own OS. That would not be the end of new discoveries for me that day. At the end of the class, the teacher told us to save the VMs on our portable hard drives, and do our homework on them. As confusing as it sounded at the time, it was also very exciting. That day I had witnessed one of the main characteristics of VMs: portability for end users.

The term “virtualization” has been around for a while, but in the last decade it has become more important as the technology is better developed, and the benefits of implementing it are evident for users. To begin, virtualization has to be defined in order to understand how it works. Pearce, Zeadally, & Hunt (2013) affirm that Virtualization is “used to abstract resources and devices in general” (p.8). This means that by virtualizing we create the idea that there are many independent systems with its own hardware and software, when the reality is that there is just one physical device that is being shared through multiple virtual devices.

The National Institute of Standards and Technology (NIST) (2010) defined virtualization as “the simulation of the software and/or hardware upon which other software runs” (p. 6). This concept has been mainly applied for the virtualization of servers. Nowadays, this technology has been used to virtualize other IT components like networking and storage. These three mainly used types of virtualization are covered in the next paragraphs (Pawar and Bhelotkar, 2011, para. 16).

Server virtualization directly relates to the creation of Virtual Machines (VM) by using specific software (hypervisor or virtual layer) that allows a single physical server to host multiple VMs. Each one of them appears as an independent host to the users, and counts with all hardware as a physical computer. Since VMs are a compilation of files, they are independent from their physical host and can be moved from one to another within minutes and without affecting the users’ processes. This characteristic also allows the modification of the “physical” resources assigned to them like memory and storage.

Depending on how the hypervisor is used in the physical server, there are two types of virtualizing hardware: Bare metal, and hosted. On the first one, the hypervisor is directly installed on the hardware of the physical host, which means that it cannot be used for anything

else than for hosting VMs. On a hosted system, the physical hardware has its own Operating System (OS), and the hypervisor runs on top of it. In this case the physical host is capable of running other applications.

Networking virtualization has been around for many years too. All network devices, for example routers and switches, count with a way of virtualizing their ports. One well known concept in this area is virtual LANs (vLAN). When vLANs are set on single network device, it can be used as the medium between servers in different networks allowing them to communicate with each other, and without having any interference from a network different than their own. Servers that are hosting VMs also virtualize their Network Interface Card (NIC) which is called vNIC. This allows every VM to have its own Internet Protocol (IP) address.

Storage virtualization mainly refers to the use of Storage Area Networks (SAN). Norall (2007) mentions that, "storage virtualization creates an abstraction layer between host and physical storage that masks the idiosyncrasies of individual storage devices" (para. 3). When using this technology, end users perceive remote storage as it is actually attached to the system. Important element of the SAN design is the switches. They can be fiber, copper, or any other technology and speed. Because the hosts and storage arrays are connected to them (physically or wirelessly), it is within the switches where the SAN configuration takes place by joining a host to specific volumes in one or different storage array. This process is dynamic, and storage can be expanded or reduced at any time without interrupting any of the activities of the end user.

The advantages of one or more of this type of virtualization are many, and it depends on how they are being used. The following are some of the reasons to adopt virtualization (Morton and Aboud, 2012, p.22):

- Server consolidation

- Physical servers generally use only 20 percent of their capacity.
- When using virtualization the hardware utilization will increase by running different VMs in the same physical server.
- Hardware costs: Because virtualization allows for greater utilization of existing resources, fewer physical servers are required, saving money both on upfront hardware costs and maintenance costs.
- Ease of Testing and Development: Virtualization allows designers to compare application performance across different operating environments, as well testing out applications in virtual environments.
- Improved Desktop Manageability & Security: Deploy, manage & monitor secure desktop environment that users can access locally or remotely, with or without a network connection.
- Energy Efficiency: Virtualization helps in reducing the energy demands of the datacenter through server consolidation & dynamic management of computers assets across a pool of servers.

Virtualization in Green Data Centers

As stated in previous paragraphs, both of the most used "green" certification systems in the US consider energy consumption as one of their factors for a building to be "green". Before the equipment actually arrives to the data centers, it is already consuming energy; according to Uddin and Abdul Rahman (2012), "manufacturing computers and their various electronic and non-electronic components consumes electricity" (para. 26). Once in the data centers, what these systems mainly use energy for is to process data and to cool down. From this, we can discern that producing less computing systems will reduce energy consumption in three ways: one, when

they are being designed; two, when they are being used in data centers processes and cooling; and three, the waste material at the end of their life cycle. Practicing these at any degree should be the goal for all data centers to get on the path to becoming a "green" data center which is "a repository for storage, management and dissemination of data in which mechanical, lighting, electrical and computing systems are designed for maximum energy efficiency and minimum environmental impact" (Uddin and Abdul Rahman, 2012, para.32).

Virtualization is one of the technologies that directly address the identified energy concerns, and because of its nature, it represents cost-reduction due to an efficient energy use, public image aligned to environmental concerns, and environmental regulatory compliance. Virtualization has become a tool to accomplish energy efficiency, which refers to "the use of technology that requires less energy to perform the same function" (Uddin and Abdul Rahman, 2012, para.36). In a study made in 2012, from the total number of data centers that were included, there were approximate 65% that confirmed that they were using virtualization as a way to become green. The change to virtualized systems represented for them almost 30% of total amount reduced on power bill for that year (Penny, Morton,and Curtland, 2012, para. 8).

When comparing the power consumption between virtual servers and physical machines, the results varied and can be affected by air temperatures. On an experimental setup for comparing two VMs against a physical server, it was found that while in idle state physical servers consume two-thirds of the same amount of energy that when totally active; on the other hand, a server hosting VMs consumes one-third. It was also found that a server hosting two VMs consumed between 15% and 20% less energy than a server running three VMs. The energy consumption is relative to the number of VMs (Jin, et al., 2012, p.11). In another experiment, it was showed that server virtualization improved power efficiency. Here, two virtualize servers

consumed 103.1 Watts less than two physical server. Most of the energy was consumed when the servers were inactive. This said, if all this servers are consolidated as VMs in a single physical server, it can represent reducing the energy cost of four servers. Virtualized servers also showed better power consumption in warm environments, the difference between them was almost of 100 Watts. According to the author, this is mainly related to the structure of the servers use for virtualization and how the workloads are divided depending on available resources (Liu, Masfary, and Antonopoulos, 2012, para. 28-30).

In general, the advantage of adopting virtualization is a great idea in order to start reducing environmental contamination, but it is also important to consider that it is not the panacea to the IT footprint. Others factors like better building practices, energy efficient equipment, and education have to be relearned in order to fit the new technologies. As all good lasting education, the one we learn at home, it is necessary that we change the mentality of having a single physical computing device for every single member of the family. Today, with virtualization, families can share a physical server and install VMs on them. I believe that culturally we are not ready for the quick technological advances that the society is experimenting. Probably when we see the effect of our actions, then we will understand that the free resources we have today might not be there tomorrow.

References

Jin,Y., Wen, Y.,Chen, Q., & Zhu, Z. (2012). An Empirical Investigation of the Impact of Server Virtualization on Energy Efficiency for Green Data Center. doi 10.1093/comjnl/bxt017

Liu,L., Masfary, O., & Antonopoulos,N. (2012). Energy Performance Assessment of Virtualization Technologies Using Small Environmental Monitoring Sensors. doi: 10.3390/s120506610

Green Globes. (2013). Green Globes – A Practical, Web-based Alternative to LEED. Retrieved from http://www.datacenterjournal.com/dcj-magazine/green-globes-practical-webbased-alternative-leed/

Green Globes. (2014). Green building programs. Green Globes Overview. Retrieved from http://www.thegbi.org/green-globes/

Morton, R & Aboud, R. (2012). VMware vSphere: Install, Configure, Manage ESXi 5.1 and vCenter Server 5.1. Lecture Manual. Vol 1.

National Research Council. (2013). Energy-Efficiency Standards and Green Building Certification Systems Used by the Department of Defense for Military Construction and Major Renovations. Washington, DC: The National Academies Press. Retrieved from http://www.usgbc-sd.org/resources/documents/dod_leed_cost_study.pdf

Norall, S. (2007). Storage virtualization. InfoWorld, 29(7), 24-27. Retrieved from http://search.proquest.com.jproxy.lib.ecu.edu/docview/194380632?accountid=10639

Pawar, U., & Bhelotkar, M. (2011). Proceedings from ICWET 2011: Virtualization: A Way Towards Dynamic IT. Mumbai, India. doi 10.1145/1980022.1980082

Pearce, M., Zeadally, S., & Hunt, R. (2013). Virtualization: Issues, security threats, and solutions. ACM Computing Surveys, Vol.45 Iss: 2, Article No. 17 doi 10.1145/2431211.2431216.

Penny, J., Morton, J., & Curtland, C. (2012). Virtualization and consolidation dominate data center energy solutions. Buildings, 106(6), 10. Retrieved from http://search.proquest.com.jproxy.lib.ecu.edu/docview/1024158143?accountid=10639

Scarfone, K., Souppaya, M., & Hoffman, P. (2011). Guide to Security for Full Virtualization Technologies. *Recommendations of the National Institute of Standards and Technology, NIST special publication 800-125.* Retrieved from http://csrc.nist.gov/publications/nistpubs/800-125/SP800-125-final.pdf

Uddin,M. & Abdul Rahman, A. (2012). Energy efficiency and low carbon enabler green IT framework for data centers considering green metrics. Renewable and Sustainable Energy Reviews. Vol. 16 Iss 6, doi 10.1016/j.rser.2012.03.014.

US Green Building Council. (2014). LEED. Retrieved from http://www.usgbc.org/leed