Big Data on the GreenCloud

Shlok S Vaidya
Electrical and Systems Engineering
University of Pennsylvania
Philadelphia, US

Pranay Shah
Electrical and Computer Engineering
Polytechnic Institute of New York University
New York, US

ABSTRACT
Cloud computing currently serves massive data loads across public and private clouds. However this poses a huge demand for energy support. The current system is mainly centered on coal and non-renewable energy based sources. A transition must be made towards green data centers which depend on renewable energy sources. According to the International Data Corporation, in 2010 for every $1 spent on hardware, 50 cents are spent on power and cooling [1]. We expect that by 2013 for every $1 spent on hardware, $1 will be spent on power and cooling. This naturally prompts stakeholders of Fortune 500 companies to push towards a Green IT transition. GreenCloud based datacenters, in the long term are not only financially rewarding but also environmentally friendly. This paper describes the landscape of the current interplay between three major parameters – Big Data, Cloud Computing and Green IT. A systematic two-fold approach to bring about a transition to a GreenCloud system is suggested. The future of Big Data is on the Green Cloud.

General Terms
Big data, Dynamic Energy Allocation, Grid/Cloud computing, Distributed Systems, Green Energy

Keywords
GreenCloud, Cloud Computing, Virtualization; Datacenter, Energy Management

1. NEED FOR THE GREENCLOUD
On the 1st of February 2012, Facebook, in its SEC filing [2] estimated that it stored more than a 100 petabytes of data between its data centers. They operated at 28 megawatts each. Google's needs crown the charts with a whopping 260 megawatts of power for a single data center. In a rough sense of the word, 1 megawatt can power up to a 1000 homes. In a report [3] by the Wall Street Journal in May 2012, it estimated that Data centers consume about 1.3% of all global electricity. This is about 277 terawatt-hours per year, more than the electricity use of a multitude of countries, even prominent consumers like Australia and Mexico. To make statistics worse, Intel expects the number of electronic devices to grow to a 15 billion [4] from a current 2.5 billion by 2015. 40% of the global electricity production comes from coal - which is fast nearing depletion. It is therefore imperative in the backdrop of these circumstances, to stage a transition to migrate and support Big Data on the Green Cloud.

2. OPPORTUNITY AND THE ECONOMIC LANDSCAPE
The need for transition opens up an interesting economic landscape. Global Data's results [5] for the Solar Photovoltaic Cell driven power market are shown in Fig. 1. The cumulative installed capacity is positioned to reach a 350,000 MW by 2020.

An average of 11,000 Wind Towers are expected to be deployed by 2020 globally – typically each generating 1.5 MW of power [5]. This creates a ripe market for early stage venture capital backed companies who can create data centers powered by renewable sources. In addition to the startup landscape, two factors are driving even top cloud players to adopt a sustainable approach.

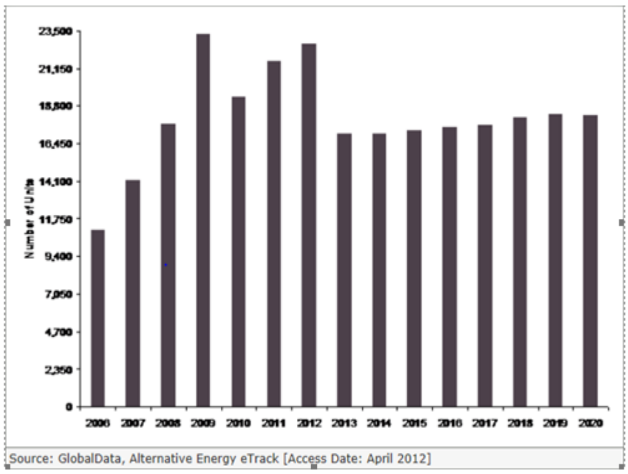

Figure 1: Solar PV Power Market, Global, Cumulative Installed Capacity (Megawatts-MW), 2001-2020.

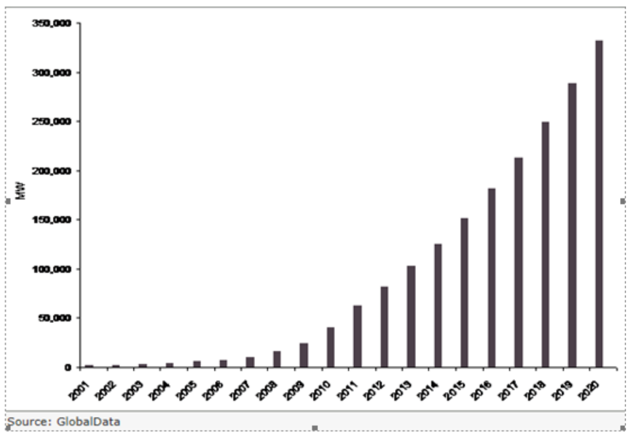

Figure 2: Wind Tower Market, Global, Units Installed, 2006-2009.

Analyst reports estimate that the top cloud computing companies performed 94.4% better than the S&P 500 [6]. With profits skyrocketing, these companies are increasingly interested in investing in the Green Cloud transition. The Green Cloud is a data center based on the principles of Green IT. In the US and most parts of the European Union, government regulations require companies to reduce their carbon footprint. This clearly establishes a motivation for the transition. The issue that needs to be addressed next is that of bridging the gap. IBM's solar-power array in Bangalore, India provides 50 kilowatt electricity for 330 days a year, an average of five hours a day. The mathematics quickly crystalizes into only a 20% powering of the data center by solar energy [7]. The difference in the power that renewable sources can deliver and the actual requirements for data centers is huge. A synthesis of top down and bottom up approaches needs to be adopted to address this disparity. The top down approach involves using increasing onsite energy generation using wind energy, solar arrays and fuel cells. However market penetration and successful acceptance of on a global scale require not only easy availability and access of renewable energy but also a superior technological backing. The bottom up involves Green IT. What the Green IT industry essentially seeks to do is to rapidly empower carbon neutral cloud services while essentially balancing high loads of data. A technological transition is needed to effectively enforce Green IT principles.

The question of how this energy is harnessed, made economical, scaled globally and how technology is commissioned to support this evolving paradigm is what we turn our attention to in the rest of this paper. We begin, by having a closer look at the fundamentals of what really makes a cloud system.

3. BIG DATA ON CLOUD COMPUTING

Cloud computing includes both the applications delivered as services over the Internet and the hardware and systems software in the data centers that provide these services. The services themselves are often referred to as Software as a Service (SaaS) [8]. Two other service models also exist with respect to this. In the IaaS model, the vendor only provides the underlying infrastructure whereas in the PaaS model, the vendor provides the application development platform as well. The data center hardware and software is what we will call a cloud. Essentially, it is an idea that the technological capabilities should "hover" over everything and at the same time be available whenever a user or a business need wants [9]. Ontology wise, cloud computing systems fall into one of five layers: applications, software environments, software infrastructure, software kernel, and hardware. At the bottom of the stack is the hardware layer with the physical components of the system and at the top of the stack is the cloud application layer which is the interface between the cloud and the common computer users [10]. The essence of the discussion is simply this- the business data resides on the cloud and clients can connect to that server in order to access this data. Huge amounts of information can be effectively utilized by allowing them to reside on the cloud. Interconnecting and analyzing all pieces of information collected within a firm offers limitless business potential. As Cloud computing is becoming a de facto standard, companies are now trying to mine this immense source of information.

4. TOP DOWN – LEVERAGING RENEWABLE ENERGY

Based on an alternative energy solution, this approach requires a significant amount of initial capital investment, but gives equivalent high monetary returns in the future. At the most fundamental level, the primary challenge with solar and wind energy is that, unlike brown energy drawn from the grid, it is variable. [11] The model shown below is just one of the solutions developed to addresses this problem and provide continuous power to support the servers.

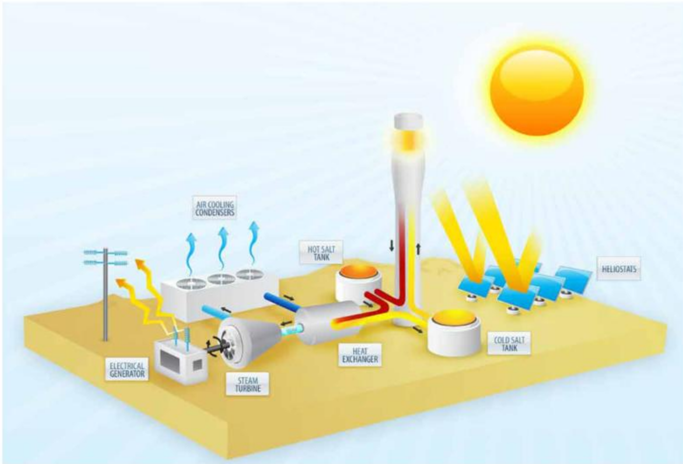

Figure 3: Diagram showing a concentrating solar thermal power plant

Mirrors called "heliostats" are used to track the Sun and focus sunlight on to a central "power tower". This energy is then stored in molten salt as heat, warming the salt to 565°C. This energy storage has efficiency nearly up to 93%. To generate electricity, the hot salt is pumped into a generator, where the heat is transferred to steam which drives a turbine. Once the salt is cooled to 290°C (while it is still warm enough to be molten), it returns to the tank to be reheated. The solar thermal plant has a store of energy ready to go during night time too [12]. Overall, to mitigate the variability, data centers could "bank" green energy in batteries or in some other form. Scheduling systems like GreenSlot and GreenHadoop are essentially software packages that aim at maximizing the use of solar energy in a data center and switching to conventional energy only when the need arises [11]. Modulate the servers' duty cycle using fast sleep states were also suggested in this context.

5. BOTTOM UP - THE TECHNOLOGY TRANSITION

5.1 Efficiency

At the heart of supporting current data center infrastructure and allowing it to scale lays the golden axiom to optimize current energy efficiency. Efficiency in a data center is measured as computation per total energy drawn. The power usage effectiveness (PUE) [13] accounts for ratio of total power to the IT power i.e. servers, network equipment and memory. The benchmark value is 1 but current systems are facing PUE of 1.7-1.8 on an average [14].

The main sources of high PUE are due to multiple sources of energy overheads like cooling equipment, UPS systems, lighting or power distribution units, switches, computing servers and others. By leveraging on an efficient designing of infrastructure, this overhead can be reduced [15]. Intel's datacenter 2020 design suggest reducing this PUE to 1.4 by careful placement of server racks and supplying just enough energy to sustain cooling for the servers by analyzing the hot and cool airflow in the data center [16]. Cooling using water from nearby river and sea is one option. Another big source of energy dissipation takes place during AC to DC & DC to AC conversion for UPS systems which accounts an efficiency of 89%. This efficiency can be increased to 99% by placing by having an individual UPS per server than Datacenter wide UPS.

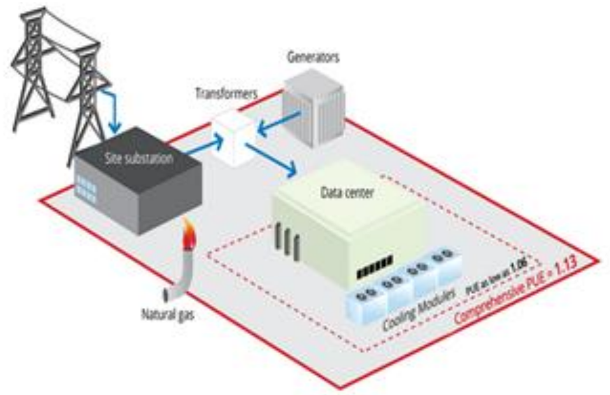

Figure 4: A Typical Google Datacenter (Source: Google)

Google's data center uses highly efficient voltage regulator per server which saves up to 500 kWh per server. The Server PUE, another factor, accounts for computation energy losses like motherboard, CPU, disk, DRAM and I/O cards. The suggested design indicates to virtualize as much as possible and to remove all not usable video cards and peripherals to achieve maximum efficiency [15].

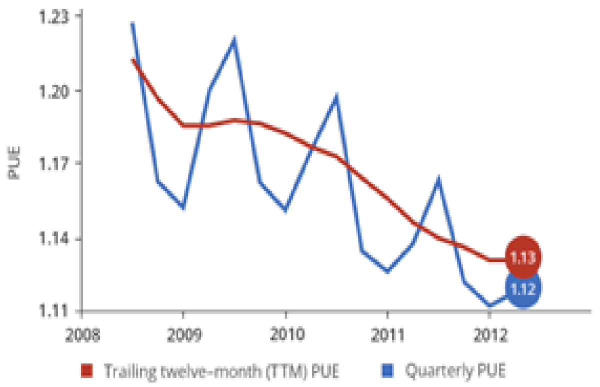

Figure 5: Graph shows continuous improvement of Energy efficiency in large Google datacenters (Source: Google.com)

5.2 Dynamic Energy Allocation

In real-world dynamics, service requests in the cloud are highly time dependent. By leveraging this phenomena, dynamic allocation of the cluster resources have proven to provide 33% energy savings under medium loads and 54% under light loads [17]. Idle machines consume up to 60% of a fully utilized computational power. Switching off servers or setting them on a sleep mode has thus proved to be highly effective [18]. The MapReduce framework has been designed to incorporate functions such as reliability, load balancing fault tolerance and resilience within the cluster in the data center. MapReduce essentially relies on the underlying Hadoop Distributed File System (HDFS) to divide large applications into smaller blocks of work and large data into smaller data blocks. They are then allowed to execute in parallel, in a fault tolerant way by creating multiple copies of data as well as the computation. The key idea here is to implement power control in conjunction with MapReduce to improve cluster utilization. Thus, when the node is not required for computation, the node can shut down [17].

5.3 Virtualization

In this growing big data era, there has been a dynamic resource request arriving at the data centers, so in order to fulfill the time dependent requests and to avoid the scaling out of the current data center the most efficient technique applied is virtualization. Virtualization by its virtue means to generate a virtual sense of either a server, operating system, storage or a networking device of a system on the same physical hardware [19]. Many current server CPU utilization rates typically hover between 5% and 15%. Direct-attached storage utilization sits between 20% and 40%, with network storage between 60% and 80%. Virtualization can increase hardware utilization by five to 20 times and allows organizations to reduce the number of power-consuming servers [9].

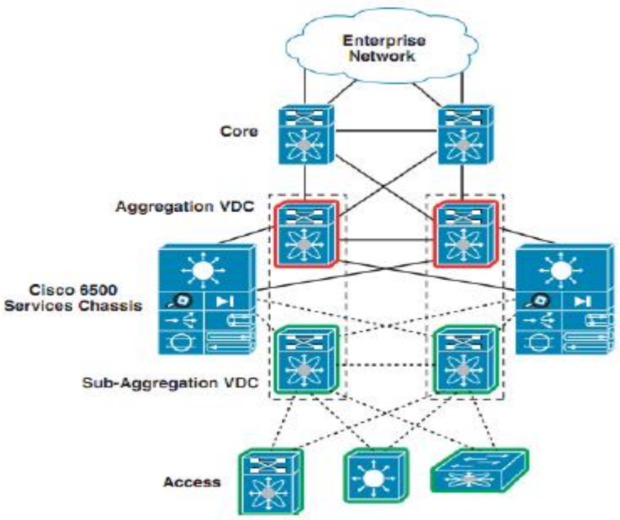

Figure 6: A Cisco Datacenter 2.5 Network Architecture with a virtualized router

To elaborate, a report by the US Environmental Protection Agency (EPA) states that a processor sits idle for 85-95% of the time. Thus by consolidating 30 to 40 virtual servers in the single physical server cuts power, cooling costs and provides more computation power in less space [20]. In conclusion, virtualization of server saves up to 111 MW-h of electricity and 26 metric tons of CO2 emissions per year. Another type of virtualization takes place for networking devices such as Layer 3 switches. The figure above shows the virtualization of aggregation layer switch which allows service modules to incorporate within the data center. Service modules include NAT, IDS, IPS and SSH which allows load balancing, security and resilience within the data center. This feature allows remote clients to work from home using VPN services, reducing pollution caused by driving to work every day; it is estimated to eliminate 23 metric tons of greenhouse gases emission per year [21].

5.4 Power and Energy Management

Many hardware products have built-in power management features that are never used. Most major vendors have been implementing such features for quite some time. These features include the ability of the CPU to optimize power by dynamically switching among multiple performance states. The CPU will drop its input voltage and frequency based on how many instructions are being run on the chip itself. These types of features can save organizations up to 20% on server power consumption [9].

In a study on energy aware cloud computing, power management schemes like voltage scaling and dynamic shutdown were successfully applied on the computing as well as on the Networking components of a data center [22].

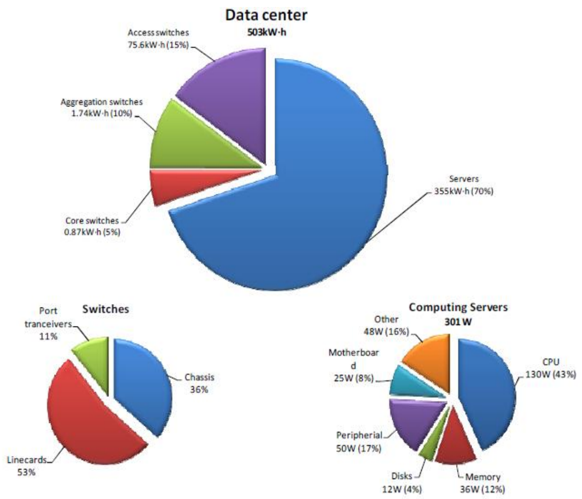

Figure 7: Distribution of energy in a data center

6. BIG DATA ON THE GREEN CLOUD

Figure 8: IBM "Big Green" Approach to data center efficiency

Given the backdrop of these technological and economic propellers, it becomes imperative to design an integrated way to re-engineer the Big Data process to fit the Green cloud Implementation. The central theme of the solution lies in creating a Green Data center by implementing the technological transition points discussed above. The Green Data center should integrate virtualization, energy management, cooling and efficiency solutions and dynamic energy allocation principles. IBM's five factor model, shown in Fig. 8, articulates this understanding well.

7. CURRENT GREEN INITIATIVES AND DEVELOPMENT

To state it simply, the concept of bringing big data on the GreenCloud has gone viral. From small businesses to well-known giants, all are investing significant resources and time towards an energy efficient and sustainable tommomow. Below, we list a few developments in this field:

- In 2007, the largest search engine Google (NASDAQ-GOOG), they installed the largest corporate solar panel of its kind generating 1.7MW in Mountain View, CA. This solar plant generates 30% of the current datacenter energy requirements. In September 2012, Google signed an agreement with the Grand River Dam authority to supply 48MW of wind powered energy to the west Oklahoma datacenter [23].
- Intel (NASDAQ – INTC) has begun testing using solar energy to power a data center in New Mexico, with a photovoltaic array generating 10 kilowatts of power. This will offset an estimated 907,000 pounds of carbon dioxide, the leading greenhouse gas, by 2025, according to the U.S. Environmental Protection Agency. That is equivalent to as much carbon dioxide produced by 919,005 miles driven in an average car. [24].
- Apple's (NASDAQ:AAPL) data center in Maiden, North Carolina, draws about 20 megawatts of power at full capacity. Apple plans to produce 60 percent of this power onsite. A combination of solar arrays, on-utility fuel cell installation and huge improvements in technological drivers has enabled Apple to promise a 100 percent renewable energy for its data centers by 2013 [25].
- In 2008's, CeBIT trade show, Sun Microsystems (NASDAQ – JAVA) demonstrated a solar Black box shipping container hooked to a 700 square foot array of solar panels, which produced about 10 kilowatts of power [22].
- The new Cisco (NASDAQ-CSCO) data center in Allen, Texas has solar cells on the roof, which can generate 100 kilowatts of power for the office spaces in the building. The facility also captures rainwater for use in its irrigation systems, and is seeking LEED Gold certification [26].
- BendBroadband Vault's new facility in Bend, Oregon includes a solar array that can generate up to 152 kW of power, or about 18 percent of the expected power capacity of 900 kW for the first phase of the building. The Vault also purchases renewable energy credits through its utility Pacific Power & Light that offset the data center's electricity use [27].

8. CONCLUSIONS

The answer is clear. Google, Facebook, Amazon, Intel, eBay all have joined the 'green' race. Overall, trends suggest that solar and/or wind power will become increasingly attractive, especially for small and medium data centers as they require smaller and cheaper installations [22]. Government incentives and the economic landscape are encouraging. By using a combination of the top down and bottom up approaches outlined, the authors believe a successful transition can be achieved. The future of Big Data shall be on the Green Cloud.

9. ACKNOWLEDGMENTS

The authors of this paper are deeply grateful to The University of Pennsylvania and the New York University for their vast library resources. We would also like to thank Ms. Komal Vora for her vision and guidance.

10. REFERENCES

[1] White Paper :"Solutions of the Datacenter's Thermal Challenges", Jed Scaramella Matthew Eastwood, Sponsored by IBM and Intel,January 2007.p1

[2] United States Securities and Exchange Commission."FormS-1". Internet:http://sec.gov/Archives/edgar/data/1326801/000119312512034517/d287954ds1.htm, Feb.1, 2012 [Oct. 18, 2012]

[3] Robert Bryce ,"Renewable Energy Can't Run the Cloud", The Wall StreetJournal.Intenet: http://online.wsj.com/article/SB10001424052702303610504577420251668850864.html, May28, 2012 [Oct. 14, 2012]

[4] Sherri Negrea,"Intel in Everything" Internet: http://www.engineering.cornell.edu/magazine/features/intel.cfm, [Oct. 18, 2012]

[5] Copyright 2012 Alternative Energy eTrack- GlobalData

[6] Tickerspy,"Cloud Computing Stocks", tickerspy Internet: http://www.tickerspy.com/index/Cloud-Computing-Stocks Oct 13, 2012 [Oct. 13, 2012]

[7] IBM press releases, "IBM Rolls Out First Solar Array Designed For High-Voltage Data Centers and Industrial Use", IBM, Internet: http://www-03.ibm.com/press/us/en/pressrelease/35891.wss, Nov 3, 2011 [18 Oct, 2012]

[8] Armbrust, M. et al. "A View of cloud computing.Commun". ACM 53, 4 April. 2010.

[9] John Lamb, Ph.D, "Green IT and use of Private Cloud Computing in South Africa", Emerging Technologies for a Smarter World (CEWIT), 2011 8th International Conference & Expo, New York,US, 2-3 Nov. 2011

[10] Lamia Youseff, Maria Butrico, DilmaDaSilva, "Toward a Unified Ontology of CloudComputing", Grid Computing Environments Workshop, 2008. GCE , Nov. 2008.

[11] Ricardo Bianchini, "Leveraging Renewable Energy in Data Centers: Present and Future" Department of Computer Science, Rutgers University, HPDC '12 Proceedings of the 21st international symposium on High-Performance Parallel and Distributed Computing, New York, US, 2012

[12] James Wight, "Zero Carbon Australia: We can do it", Skeptical Science, Internet: http://www.skepticalscience.com/Zero-Carbon-Australia-2020.html. 19 March 2011 [19 Oct. 2012]

[13] Wikipedia,"Power Usage Effectiveness",Wikipedia.Internet: http://en.wikipedia.org/wiki/Power_usage_effectiveness, Sept. 13, 2012 [Oct. 19, 2012]

[14] "A1: Power Usage Effectiveness (PUE)" Self Benchmarking guide for high-tech buildings,.Internet: http://hightech.lbl.gov/benchmarking-guides/data-a1.html, 2012 [Oct. 19, 2012]

[15] Luiz André Barroso and Urs Hölzle, Google Inc. The Datacenter as a Computer - An Introduction to the Design of Warehouse-Scale Machines. Morgan&Claypool, Madison, 2009, pp. 39-67.

[16] Whitepaper: "Data Center 2020: Delivering high density in the Data Center, efficiently and reliably by 2020", published: Mar. 2nd, 2012.

[17] Nitesh Maheshwari, Radheshyam Nanduri, Vasudeva Varma. "Dynamic Energy Efficient Data Placement and Cluster Reconfiguration Algorithm for MapReduce Framework." submitted to Future Generation Computer Systems (FGCS), India, July 2011

[18] Qi Zhang, Mohamed Faten Zhani, Shuo Zhang, Quanyan Zhu,Raouf Boutaba, Joseph L. Hellerstein. "Dynamic Energy-Aware Capacity Provisioning for Cloud Computing Environments" Copyright 2012 ACM 978-1-4503-1520-3/12/09.

[19] Wikipedia,"Virtualization", Wikipedia.Internet: http://en.wikipedia.org/wiki/Virtualization ,Oct. 18,2012 [Oct. 19, 2012]

[20] Whitepaper: "How VMware Virtualization Right-sizes IT Infrastructure to Reduce Power Consumption", VMware, Inc Item number respective companies. Item No: VMW-WP-REDUCE-PWR-CONSUMPTN-USLET-103, www.vmware.com

[21] Rob DiStefano ," Building an Energy-Efficient Data Center Using Virtualization Technology", Environmental Leader. Internet: http://www.environmentalleader.com/2009/11/19/building-an-energy-efficient-data-center-using-virtualization-technology/ Sept. 19, 2009 [Oct. 16, 2012]

[22] Dzmitry Kliazovich ,Pascal Bouvry, Yury Audzevich, Samee Ullah Khan,"GreenCloud: A Packet-level Simulator of Energy-aware Cloud Computing Data Centers" , Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE, Miami, FL, US 6-10 Dec. 2010.

[23] Google,"Renewable Energy",Goole, Internet: http://www.google.com/green/energy/ [Oct. 13, 2012]

[24] Preston Gralla, "Can Solar Energy Power a Data Center?", GreenBiz.com, Internet: http://www.greenbiz.com/blog/2009/01/18/can-solar-energy-power-data-center, Jan 18th 2009 [19 Oct. 2012]

[25] Apple, "Data Centers and Renewable Energy", Apple, Internet: http://www.apple.com/environment/renewable-energy/ [Oct. 19,2012]

[26] Rich Miller,"Cisco Opens Doors on Texas Data Center", DataCenterKnowledge,Internet: http://www.datacenterknowledge.com/archives/2011/04/18/cisco-opens-doors-on-new-texas-data-center/. Apr. 18,2011 [Oct. 19,2012]

[27] Rich Miller, "Closer Look: The BendBroadband Vault" DataCenterKnowledge.Internet: http://www.datacenterknowledge.com/archives/2011/04/26/closer-look-the-bendbroadband-vault/. Apr.26,2011 [19 Oct. 2012]