Green Computing: Green Maturity Model for Virtualization

Gokulakumar.A.S[1], Venkataraghavan.C[2] and Rathika.S.K.B[3]

[1]Department of CSE, SNS College of Engineering, Coimbatore, Tamilnadu, India.

[2]Department of CSE, SNS College of Engineering, Coimbatore, Tamilnadu, India.

[3]Department of CSE, SNS College of Engineering, Coimbatore, Tamilnadu, India.

Abstract

Green Computing is one the recent trends that's growing with a tremendous phase in IT sector. Now each company are following the green principles before the design of their systems. Green computing helps in reducing the pollution and destruction of environment. The basic principles that are followed here are green disposal, green design and green manufacturing, which are achieved through a series of ways like power management, green data center, e-waste disposal and virtualization etc,. Here we look at detailed view of virtualization. Firstly we look at all the four levels in a glance, then we look more deep on each level and we see the way of achieving it with good real time examples.

I. INTRODUCTION

Green Computing is the study and practice of using computing resources efficiently. Typically, green computing systems or products take into account the so-called triple bottom line of people, planet, and profit. This differs somewhat from traditional or standard business practices that focus mainly on the economic viability of a computing solution. These focuses are similar to those of green chemistry; reduction of the use of hazardous materials such as lead at the manufacturing and recycling stages, maximized energy efficiency during the product's lifetime, and recyclability or biodegradability of both a defunct product and of any factory waste. It is "the study and practice of designing, manufacturing, using, and disposing of computers, servers, and associated subsystems—such as monitors, printers, storage devices, and networking and communications systems—efficiently and effectively with minimal or no impact on the environment. Thus, green IT includes the dimensions of environmental sustainability, the economics of energy efficiency, and the total cost of ownership, which includes the cost of disposal and recycling With increasing recognition that man-made greenhouse gas emissions are a major contributing factor to global warming,

enterprises, governments, and society at large now have an important new agenda: tackling environmental issues and adopting environmentally sound practices. Greening our IT products, applications, services, and practices is both an economic and an environmental imperative, as well as our social responsibility. Therefore, a growing number of IT vendors and users are moving toward green IT and thereby assisting in building a green society and economy.

II. ROADS TO GREEN COMPUTING

To comprehensively and effectively address the environmental impacts of computing/IT, we must adopt a holistic approach and make the entire IT lifecycle greener by addressing environmental sustainability along the following four complementary paths:

A. *Green design* — designing energy-efficient and environmentally sound components, computers, servers, cooling equipment, and data centers

B. *Green manufacturing* — manufacturing electronic components, computers, and other associated subsystems with minimal impact on the environment

C. *Green use* — reducing the energy consumption of computers and other information systems as well as using them in an environmentally sound manner

D. *Green disposal* — refurbishing and reusing old computers and properly recycling unwanted computers and other electronic equipment

These four paths span a number of focus areas and activities, including:

- design for environmental sustainability
- energy-efficient computing
- power management

www.ijreat.org

IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 1, Issue 5, Oct-Nov, 2013
ISSN: 2320 - 8791
www.ijreat.org

- data center design, layout, and location
- server virtualization
- responsible disposal and recycling
- regulatory compliance
- green metrics, assessment tools, and methodology
- environment-related risk mitigation
- use of renewable energy sources and
- eco-labeling of IT products

III. GREEN MATURITY MODEL FOR VIRTUALIZATION

Virtualization is a term used to mean many things, but in its broader sense, it refers to the idea of sharing. To understand the different forms of virtualization and the architectural implications for creating and deploying new applications, we propose a reference model to describe the differing forms of the concept.In this model we observe a number of different layers of abstraction at which virtualization can be applied, which we describe as increasing levels of maturity, shown in Table 1. We assert that higher levels of virtualization maturity correspond to lower energy consumption, and therefore architectures based on higher levels of maturity are "greener" than those at lower levels, which we discuss further on.

Table 1 Levels of virtualization maturity

Virtualization Maturity	Name	Applications	Infrastructure	Location	Ownership
Level 0	Local	Dedicated	Real	Distributed	Internal
Level 1	Logical	Shared	Real	Centralized	Internal
Level 2	Data Center	Shared	Virtual	Centralized	Internal
Level 3	Cloud	Software as a service	Virtual	Virtual	Virtual

A. Level 0 ("Local"):
Means no virtualization at all. Applications are all resident on individual PCs, with no sharing of data or server resources.

B. Level 1 ("Logical Virtualization"):
Introduces the idea of sharing applications. This might be, for example, through the use of departmental servers running applications that are accessed by many client PCs.

This first appeared in the mainstream as mainframe and then "client/server" technology, and later with more sophisticated N-tier structures. Although not conventionally considered virtualization, in fact, it is arguably the most important step. Large organizations typically have a large portfolio of applications, with considerable functional overlaps between applications. For example, there may be numerous systems carrying out customer relationship management (CRM) functions.

C. Level 2 ("Data Center Virtualization"):
Is concerned with virtualization of hardware and software infrastructure. The basic premise here is that individual server deployments do not need to consume the hardware resources of dedicated hardware, and these resources can therefore be shared across multiple logical servers. This is the level most often associated with the term virtualization. The difference from Level 1 is that the hardware and software infrastructure upon which applications/ servers are run is itself shared (virtualized).

For server infrastructure, this is accomplished with platforms such as Microsoft Virtual Server and VMware among others, where a single physical server can run many virtual servers. For storage solutions, this level is accomplished with Storage Area Network (SAN) related technologies, where physical storage devices can be aggregated and partitioned into logical storage that appears to servers as dedicated storage but can be managed much more efficiently. The analogous concept in networking at this level is the Virtual Private Network (VPN) where shared networks are configured to present a logical private and secure network much more efficiently than if a dedicated network
were to be set up.

D. Level 3 ("Cloud virtualization"):
In the virtualization maturity model extends Level 2 by virtualizing not just resources but also the location and ownership of the infrastructure through the use of cloud computing. This means the virtual infrastructure is no longer tied to a physical location, and can potentially be moved or reconfigured to any location, both within or outside the consumer's network or administrative domain. The implication of cloud computing is that data center capabilities can be aggregated at a scale not possible for a single organization, and located at sites more advantageous (from an energy point of view, for example) than may be available to a single organization. This creates the potential for significantly improved efficiency by leveraging the economies of scale associated with large numbers of organizations sharing the same infrastructure. Servers and storage virtualized to this level are generally referred to as Cloud Platform and Cloud Storage, with examples being Google App Engine, Amazon Elastic Compute Cloud, and Microsoft's Windows Azure. Accessing this infrastructure is normally done over the Internet with secure sessions, which can be thought of as a kind of virtualizeddiscrete VPN. Each level of maturity has a number of significant technology "aspects" of the computing platform that may be virtualized. A summary of the virtualization layers as they map to the server, storage, and network aspects is shown in Table 2.

IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 1, Issue 5, Oct-Nov, 2013
ISSN: 2320 - 8791
www.ijreat.org

Table 2: Technology aspects for virtualization

Virtualization Maturity	Technology Aspects			
	Name	Server	Storage	Network
Level 0	Local	stands one PC	Local disks	None
Level 1	Departmental	Client Server, N-tier	File server, DB server	LAN, Shared services
Level 2	DataCenter	Server virtualization	SAN	WAN/VPN
Level 3	Cloud	Cloud platform	Cloud storage	Internet

IV. LEVELS IN DETAIL

A. Starting at Home – Level 0

Level 0 ("Local") in the virtualization maturity model means no virtualization at all. Even with no virtualization, there is plenty of scope for energy savings. Traditional design and development approaches may lead to applications that are less efficient than they could be. There are also a number of other design issues that can be readily recognized in applications, and therefore, a set of rules, or principles, can be recommended to be implemented by architects and developers for all applications.

4.1 Enable power saving mode

Most PCs are idle for the majority of time, and theoretically can be turned off when idle. This can generate enormous energy savings. This has some implications for application design, as applications designers need to consider the platform (client and/or server) going to sleep and waking up. For example, if a client or server goes to sleep while an application user session is still active, what are the session timeout implications and policies when the platform wakes up?

- Testing to ensure that applications do not restrict a computer from entering sleep mode.
- Testing to ensure that an application can execute successfully when a computer has left sleep mode.
- Making sure that applications do not unnecessarily hold network connections open and do not require full-screen animation, for example. Both of these may stop a computer from entering sleep mode.
- Make sure that your application uses disk access sparingly as constant disk access will stop hard drives from powering down automatically.

4.2 Minimize the amount of data stored

Data uses power because data stored in application databases or file systems needs disks to be operating to store the data. Therefore, reducing the amount of data stored can reduce the amount of power used by an application by reducing the amount of data storage infrastructure. Efficient data archiving approaches can assist here.

Principle: Minimize the amount of data stored by the application and include data archiving in application design.

4.3 Design and code efficient applications

In the past, developers were forced to code carefully because computers were so slow that inefficient code made applications unusable. Moore's Law has given us more powerful computer hardware at a faster rate than we can consume it, resulting in applications that can appear to perform well even though internal architecture and coding may be wasteful and inefficient. Inefficient code causes greater CPU cycles to be consumed, which consumes more energy, as we describe in more detail later. Tools such as code profilers are available that can automatically review the performance profile of code.

B. Sharing is Better – Level 1

Level 1 ("Logical Virtualization") in the virtualization maturity model introduces the idea of sharing applications. This might be for example through the use of departmental servers running applications that are accessed by many client PCs. This first appeared in the mainstream as "client/server" technology, and later with more sophisticated N-tier structures. Although not conventionally considered virtualization, in fact, it is arguably the most important step. Large organizations typically have a large portfolio of applications, with considerable functional overlaps between applications. For example, there may be numerous systems carrying out "CRM" functions.

Moving to Level 1 ("Logical Virtualization") is all about rationalizing the number of applications and application platforms where there is overlapping or redundant functionality, and increasing the use of common application services so that shared application components can be run once rather than duplicated multiple times. For large organizations, this will produce much bigger payoffs than any subsequent hardware virtualization. The best way to do this is to have a complete Enterprise Architecture, encompassing an Application Architecture identifying the functional footprints and overlaps of the application portfolio, so that a plan can be established for rationalizing unnecessary or duplicated functions. This may be accomplished by simply identifying and decommissioning unnecessarily duplicated functionality, or factoring out common components into shared services. As well as solving data integrity and process consistency issues, this will generally mean there are fewer and smaller applications overall, and herefore, less resources required to run them, lowering the energy/emissions footprint and at the same time reducing operational costs. The increased use of shared application services ensures that common functions can be centrally deployed and managed rather than unnecessarily consuming resources within every application that uses them.
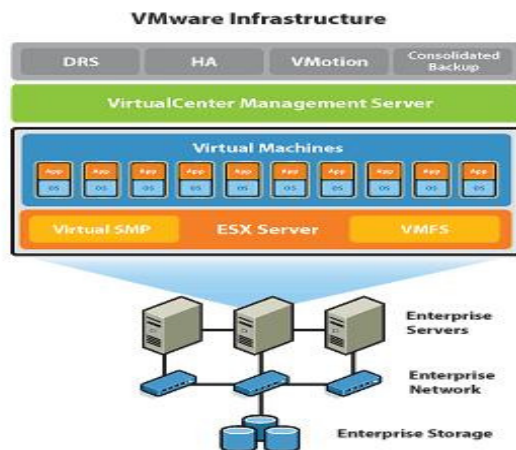
Fig-1 VW Ware Infrastucture

Fig-1 Workload of servers

1.Server Efficiency

Servers are more energy efficient per unit of work done when run at higher utilization. This is because even though each processor consumes more electricity when it is working harder, there is a base level of power that is used regardless of what a computer is doing. Increasing the utilization does not increase power consumption at the same rate. Moreover, most servers operate under quite low utilization. Therefore there are huge gains to be had from using up spare server capacity.

As described in *Windows Server 2008 Power Savings*

(see References), multiple virtual machines can run on a single physical machine without consuming significantly more power than would have been consumed if that machine was operating as a standalone server. This means that for comparable levels of throughput, you can add virtual machines at very low power cost, up to the load limits of your equipment. So long as there is excess peak CPU capacity, the savings continue to scale with the number of servers you are able to virtualize. According to the Windows Server 2008 document, "Running 4 virtual machines means saving the equivalent power output of three physical servers; running 10 virtual machines means saving the equivalent power output of 9 physical servers." In addition, physical computers take up more space (even if they are blades) and need more cooling than their virtualized equivalents. Additional hardware is a considerable waste of power and resources if it can be avoided.

A single computer running at 50 percent CPU usage will use a considerably less power than two similar computers running at 25 percent CPU usage (see sidebar, "Server Efficiency, page 11). This means that single-application servers are not efficient and that servers should ideally be used as shared resources so they operate at higher utilization levels. When aiming for this end result, it is important to ensure that sociability testing is executed to make sure that applications can work together. Also, performance testing should be executed to ensure that each application will not stop the others on the server from running efficiently while under load. In effect, it is important to ensure that the available CPU can be successfully divided between the applications on the server with sufficient capacity for growth.

As a by-product of executing this form of design, it is recommended that a set of architectural standards be introduced to require that applications install cleanly into an isolated space and not impact other applications, and that testing takes place to ensure that this is the case. However, we have all seen examples where this is not the case regardless of how simple this task would appear.

C. Level 2: Infrastructure Sharing Maximized

As defined earlier, Level 2 ("Data Center Virtualization") is the level most often associated with the term "virtualization." Through platforms like Microsoft Virtual Server, VMware and others, server and storage virtualization does provide more efficient solutions for organizations that have the size and capability to develop a virtualized infrastructure. The bar for virtualization is low and is becoming lower all the time as virtualization software becomes easier to manage and more capable. The price/performance of servers has now reached the point where even smaller organizations hosting only 3 or 4 departmental applications may reduce costs through deployment into a virtualized environment. It would appear straightforward, by extending the previous arguments about avoiding single-task computers, that data center virtualization would provide a simple answer and approach to share resources. This is indeed a good starting point, but is not as simple as it appears. The lowest-hanging fruit in the transition to virtualization are in test, development, and other infrequently used computers. Moving these machines into a single virtual environment reduces the physical footprint, heat

produced and power consumed by the individual servers. Each physical server that runs a virtualized infrastructure has finite resources and consumes energy just like any other computer. It is therefore apparent, from extending the same set of rules outlined above, that the aim is to load as much as possible onto a single physical server, and to make use of its resources in the most efficient way possible.

Creating virtual servers does not come at zero energy or management cost. Computers should not be running unless they are needed, even in a virtual environment. This will extend the limit of the available resources. It is particularly efficient to do this in a virtual environment as virtual machines can be easily paused, restarted, and even moved. This adds, as would be expected, the requirement for applications running on the virtual machines to be able to be paused, along with the base operating system. It could be possible, for example, to pause the operation of a file and print server at night while application updates run on another virtual machine, making use of the now available resources.

Fig-2 Data Center Virtualisation

2.Data Center Efficiency

Bigger data centers can be made much more energy efficient than smaller data centers. Standards are emerging for measuring this, such as the concept of Power Usage Effectiveness (PUE). PUE is defined as the ratio of total facility power divided by IT equipment power. Thus, it is a measure of how much of the power being consumed by the facility is actually being used to power the IT equipment itself rather than all the other things. By IT equipment, we mean the equipment actually delivering useful value, including servers, storage devices, switches, routers, telco network equipment, workstations/PCs, and so on.
Other elements of the facility that consume power but can be considered overhead include UPS, transformers, distribution frames, lighting, cooling, fire suppression equipment, security devices, building management, and so on. Typical IT facilities have a PUE of about 2. Best practice is currently considered to

be about 1.2. This represents a huge difference, with a potential savings of about 40 percent, however, it can only be achieved with sophisticated equipment and on a large scale. For all but the largest organizations, the best way to achieve these levels is to aggregate IT infrastructure across many organizations to reach the required economies of scale.

D. Level 3: Cloud Virtualization

Cloud virtualization in the virtualization maturity model can significantly improve efficiency by leveraging the economies of scale associated with large numbers of organizations sharing the same infrastructure. Obtaining energy efficiencies in data centers is highly specialized and capital intensive. Standard metrics are emerging such as Power Usage Effectiveness (see sidebar, "Data Center Efficiency") which can be used to benchmark how much energy is being usefully deployed versus how much is wasted on overhead. There is a large gap between typical data centers and best practice for PUE. The other major advantage of Level 3 is the ability to locate the infrastructure to best advantage from an energy point of view.

Note that applications need to be specifically designed for Level 3 to take full advantage of the benefits associated with that level. This is an impediment to migrating existing functions and applications that may limit the degree to which organizations can move to this level.

3.Making Sure Your Cloud Has A Green Lining

In applying the principles provided in this article, it is apparent that some cloud computing models are more attractive than others, keeping in mind that even running applications on servers that are located "somewhere else on the Internet" and are owned by someone else still produces an environmental cost. Cloud data centers may be more efficient to cool, but CPU and disks still need power, and therefore, the less used, the better. There are four aspects of efficiency that should be considered in cloud computing:

• The placement and design of the cloud data center
• The architecture of the cloud platform
• The architecture and development approach of the applications that are hosted.

When a customer elects to purchase electricity from a supplier, in most cases it is possible for the customer to elect to buy green electricity, and in doing so, it is possible to verify just how green that electricity is. Hopefully, your organization has made a decision to buy green power.

In the same vein, you should make an informed decision about which cloud provider to use. That is not our purpose here as this article is being written when cloud services are evolving. We can, however, outline the environmental facets of cloud data center design that an organization should evaluate in selecting a platform. Firstly, the location of the data center is

5

important. All other things being equal, a data center in a relatively hot climate such as Central Australia will require far more resources to cool than a data center that is located in a colder environment, like Iceland. Of course, there may be other considerations such as access to other "green" mechanisms for site cooling — the point is that the characteristics of a site can help reduce the energy footprint substantially. Similarly, a location near renewable power sources, such as hydro or wind power allows for a significant reduction in carbon emissions. Different vendors are approaching the design of their data centers in different ways. Different approaches that can be used to reduce power consumption in data centers include:

- Buying energy-efficient servers
- Building energy efficient data centers that use natural airflow, water cooling (ideally using recycled water and cooling the water in an efficient manner)
- Efficient operation by running lights out, by moving load to the cooler parts of the data center and by recycling anything that comes out of the data center, including equipment.

Some data center operators already publish statistics on their power usage, such as Google. These operators use an industry standard for measuring the efficiency of a data center through the ratio of power used to run IT equipment to the amount of power used to run the data center itself (PUE). As this space grows, it is expected that other organizations will do the same, allowing a comparison to be made. Another area that can be considered is the technical architecture of the cloud platform, as different organizations provide different facilities and these facilities can determine the efficiency of the application and therefore impact the efficiency of the overall platform. Some cloud vendors, for example, provide services that are controlled by the vendor; such is the case with SaaS vendors. In this case it is up to the architect of the calling application to ensure the efficiency of the overall architecture.

4.A Brighter Shade Of Green Cloud Computing
Cloud computing provides the next big thing in computing — some interesting architectural constructs, some great potential from a monetary aspect, and a very real option to provide a more environmentally friendly computing platform. Fundamentally, cloud computing involves three different models:

- Software as a Service (SaaS), refers to a browser or client application accessing an application that is running on servers hosted somewhere on the Internet.
- Attached Services refers to an application that is running locally accessing services to do part of its processing from a server hosted somewhere on the Internet.
- Cloud Platforms allow an application that is created by an organization's developers to be hosted on a shared runtime platform somewhere on the Internet.

All of the above models have one thing in common — the same fundamental approach of running server components somewhere else, on someone else's infrastructure, over the Internet. In the SaaS and attached services models the server components are shared and accessed from multiple applications. Cloud Platforms provide a shared infrastructure where multiple applications are hosted together.

V. MOVING TO INCREASING LEVELS OF VIRTUALIZATION

Referring to the model in Table 1, most IT organizations are now at Level 1 with more advanced organizations moving in whole or in part to Level 2. Only a small proportion of organizations are at Level 3. Although we argue that these levels of increasing maturity in virtualization correspond to reduced energy footprint, we note that Level 3 is not necessarily an endpoint for all organizations — in some cases there may be good business reasons for not moving to Level 3 at all. We omit the transition from Level 0 to Level 1 as the vast majority of medium to large organizations have already taken this step. Moving from Level 1 to 2 involves replacing individual dedicated servers with larger platforms running virtual servers. If more than one physical server is required, additional benefits can be achieved by grouping applications on physical servers such that their peak load profiles are spread in time rather than coincident. This enables statistical averaging of load since normally the sizing of server capacity is driven by peak load rather than average load. The trade-off here is the embodied energy cost of the new server to host the virtualized environments (see the sidebar, "Getting a Return on the Embodied Energy Costs of Buying a New Virtualization Server"). Storage infrastructure can be approached in a similar manner.

In general, with regard to server virtualization at least, the gains in virtualizing multiple servers onto a physical server are substantially greater than the embodied energy costs associated with buying a new server to host the virtualized platform.

Level 2 virtualization can also be used to leverage the embodied energy in existing infrastructure to avoid the need for procuring more infrastructure.This may seem counter intuitive but it may be better to use an existing server as a virtualization host rather than buy a replacement for it.

For example, if you have four servers that could potentially be virtualized on a single new large server, there may be advantages in simply retaining the best two servers, each virtualizing two server instances (thus avoiding the embodied energy in a new server and reducing operational consumption by about a half) rather than throwing all 4 out and buying a new server. This can be possible because the existing servers are probably running at such low utilization that you can double their load without impact on system performance.

Obviously, this will ultimately depend on the antiquity of the existing servers and current and projected utilization levels.

Note also that if you can skip Level 2 altogether, rather than deploy to level 2 and then migrate to Level 3 later, you can save on the embodied energy costs of the entire Level 2 infrastructure.

Moving from Level 2 to 3 is fundamentally about two things — sharing the data center infrastructure across multiple organizations, and enabling the location of the data center infrastructure to shift to where it is most appropriate. Sharing the infrastructure across multiple organizations can deliver big benefits because achieving best practice efficiencies in data center energy usage requires complex, capital intensive environments.

Another advantage of Level 3 is that the infrastructure can be dynamically tuned to run at much higher levels of utilization (and thus energy efficiency) than would be possible if dedicated infrastructure was used, since the dedicated infrastructure would need to be provisioned for future growth rather than currently experienced load. In a cloud structure, hardware can be dynamically provisioned so that even as load for any individual application grows, the underlying hardware platform can be always run at optimal levels of utilization.

The trade-offs here are:

- Increased load and dependence on external network connectivity, although this is largely energy neutral
- (Perceived) loss of local control because the infrastructure is being managed by an external organization (although they may commit to service levels previously unachievable through the internal organization)
- (Perceived) security or privacy concerns with having the data hosted by an external party (such as managing hospital records, for example).

In summary, we argue that energy consumption will reduce at each increase in virtualization level, therefore organizations should be looking to move up the levels to reduce their energy footprint.

VI. CONCLUSION

There is a compelling need for applications to take environmental factors into account in their design, driven by the need to align with organizational environmental policies, reduce power and infrastructure costs and to reduce current or future carbon costs. The potential reduction in energy and emissions footprint through good architectural design is significant. The move to more environmentally sustainable applications impacts software and infrastructure architecture. The link between the two is strong, driving a need for joint management of this area of concern from infrastructure and software architects within organizations. These issues should

be considered at the outset and during a project, not left to the end. An interesting observation is that our principles also align well with the traditional architectural drivers. Does this mean that energy reduction can be used as a proxy for all the other drivers? An architecture designed solely to reduce energy over the full lifecycle would seem to also result in a "good" architecture from a broader perspective.

Can we save a lot of time and effort by just concentrating solely on energy efficiency above all else?

REFERENCES

[1].Emissions of Greenhouse Gases in the United States 2006, Energy Information Administration ftp://ftp.eia.doe.gov/pub/oiaf/1605/cdrom/pdf/ggrpt/05 7306.pdf

[2].Google Data Center Efficiency Measurements, Commitment to Sustainable Computing http://www.google.com/corporate/datacenters/measuri ng.html

[3]"The Green Grid Data Center Power Efficiency Metrics: PUE and DCiE," The Green Grid http://www.thegreengrid.org/gg_content/TGG_Data_C enter_Power_ Efficiency_Metrics_PUE_and_DCiE.pdf

[4]."Guidelines for Energy-Efficient Datacenters," The GreenGrid http://www.thegreengrid.org/gg_content/Green_Grid_ Guidelines_WP.pdf

[5].Microsoft Environmental Principles, Corporate Citizenship http://www.microsoft.com/About/CorporateCitizenshi p/US/

[6].ResponsibleLeadership/EnvironmentalPrinciples.m spx Microsoft Windows Server 2008 Power Savings http://www.microsoft.com/downloads/details.aspx?Fa milyID=61d493fd-855d-4719-8662-a40ba3a0a5c&displaylang=e

[7].Moore's Law, Wikipedia http://en.wikipedia.org/wiki/Moore%27s_law [8]Object Consulting http://www.objectconsulting.com.au/beliefs.aspx

[9]."The Path to Green IT," Mike Glennon http://www.fujitsu-siemens.com/ps/itfuture2008/presentations_benefr/GA RTNER_Green-IT.ppt

[10].Performance Analysis, Wikipedia http://en.wikipedia.org/wiki/Performance_analysis [11].Telstra Corporation http://www.telstra.com.au/abouttelstra/csr/environmen t.cfm