Toward Green Cloud Computation

By Using Prediction

Nada M. A. Al Sallami [1]

1: Faculty of Information Technology and Science, Al Zaytoonah Private University of Jordan, Amman, *Email:nada.alsalami@Qyahoo.com*

Abstract:

Cloud computing is a highly scalable and cost-effective infrastructure for running High Performance Computing, enterprise and Web applications. However, the growing demand of Cloud infrastructure has drastically increased the energy consumption of data centers, which has become a critical issue. High energy consumption not only translates to high operational cost, which reduces the profit margin of Cloud providers, but also leads to high carbon emissions which is not environmentally friendly. Hence, energy-efficient solutions are required to minimize the impact of Cloud computing on the environment. In order to design such solutions, deep analysis of Cloud is required with respect to their power efficiency. In this paper we proposed an algorithm to achieve Green computing by using artificial neural networks. This paper analyzed the benefits offered by Cloud computing by studying its fundamental definitions, examples of the services it offers to end users, and its deployment model. Then Artificial N*eural Network algorithm was described to predict the workload in cloud environment.*

Keywords: Green Computing, cloud computing, workload prediction

1. INTRODUCTION

“Green” has become a popular term for describing things that are good for the environment, generally healthful and, more recently, economically sensible. “Going Green” implies reducing your energy use and pollution footprint. The technology community, specifically computer users, have popularized the term “Green Computing,” which is the reduction of the pollution and energy footprint of computers [1].Nowadays, companies continue to grow larger and larger, not only in the number of employees, but in the number of departments and type of employees. In cases such as these, cloud computing is a resource that is readily available to help companies meet their needs and accomplish their goals. Cloud computing is an excellent technological tool that can benefit the business. All businesses need to respond to competition by making better use of Internet services and offering more incentives than their competitors. Cloud computing can help business shift their focus to developing good business applications that will bring true business value [2]. Due to the exponential growth of cloud computing, it has been widely adopted by the industry and there is a rapid expansion in data-centers. This expansion has caused the dramatic increase in energy use and its impact on the environment in terms of carbon footprints. The link between energy consumption and carbon emission has given rise to an energy management issue which is to improve energy-efficiency in cloud computing to achieve Green computing [5]. In this paper we proposed an algorithm to achieve Green computing by using artificial neural networks.

Anyone trying to understand the cloud computing phenomenon need only look at how cloud services are being used to get a better picture. Over the last few years we've seen tremendous growth in cloud computing, as witnessed by the many popular Web apps used today, including: VoIP (e.g., Skype, Google Voice), social applications (e.g., Facebook, Twitter, LinkedIn), media services (e.g., Picassa, YouTube, Flickr), content distribution (e.g., BitTorrent), financial apps (e.g., Mint), and many more. Even traditional desktop software, such as Microsoft Office, has moved in part to the Web, starting with its Office 2010 Web

Apps. The following examples demonstrate the cloud being used for everything from marketing campaigns to space exploration and scientific research.

A. Web-based email services like Gmail and Hotmail deliver a cloud computing service: users can access their email "in the cloud" from any computer with a browser and Internet connection, regardless of what kind of hardware is on that particular computer. The emails are hosted on Google's and Microsoft's servers, rather than being stored locally on the client computer.

B. Data stored on your home or business computer suffers from many of the same restrictions as email and, as with email, the cloud offers a solution. Storing your MP3′s, video, photos and documents online instead of at home gives you the freedom to access them wherever you can find the means to get online. Examples of online storage services include Humyo, ZumoDrive, Microsoft's SkyDrive, S3 from Amazon, amongst others. Many offer both free and paid for storage and backup solutions.

C. Google launched a service that allowed groups of people to work on the same document, idea or proposal in real time or whenever convenient to each participant. Using Google Wave you can create a document and then invite others to comment, amend, offer opinion, or otherwise join in with the creation of the final draft.

D. Similar to instant messaging but offering much more scope it can take a project that might have taken weeks or even months to complete using other methods and potentially see it through to completion in mere minutes or hours.

2. CLOUD COMPUTING PRINCIPLE

Cloud computing can mainly provide three cloud services: storage as a service, processing as a service and software as a service. Table 1 provides a summary of the location of processing, location of storage, and function of transport for each of these cloud services. In a storage service, the majority of processing occurs at the user's PC (the client) and the majority of storage is in the cloud. The transmission and switching network transports the user's files between the data center and the user. With a processing service, the user's computer processes only short tasks and the cloud processes large computationally intensive tasks. Long-term storage of data is on the user's computer and transport is required to transfer the files relevant to each large task. In a software service, processing and storage are performed in the cloud. Transport is required for all tasks to enable transmission of commands to the cloud and to return the results [4]. Clouds are deployed on physical infrastructure where Cloud middleware is implemented for delivering service to customers. Such an infrastructure and middleware differ in their services, administrative domain and access to users. Therefore, the Cloud deployments are classified mainly into three types: Public Cloud, Private Cloud and Hybrid Cloud.

To deliver technical and economic advantage, cloud computing must be deployed, as well as implemented, successfully. Deployment supersedes implementation, because merely utilizing the services of a cloud vendor does not by itself differentiate an organization from its competitors. Competitors likewise can implement cloud services, imitating resulting IT efficiencies. Successful deployment denotes the realization of unique or valuable organizational benefits that are a source of differentiation and competitive advantage. IT-related success is described through three categories of derived benefit: strategic, economic, and technological. Strategic refers to an organization's renewed focus on its core business activities that can accompany a move to cloud computing when its IT functions, whole or in part, are hosted and/or managed by a cloud vendor. Economic refers to an organization's ability to tap the cloud vendor's expertise and technological resources to reduce in-house IT expenses. Technological refers to an organization's access to state-of-the-art technology and skilled personnel, eliminating the risk and cost of in-house technological obsolescence. Deployment is defined in terms of the strategic, economic and technological benefits realized through cloud computing, setting the organization apart from its competitors. Optimizing the strategic, economic, and technological benefits derived from cloud computing is a function of an organization's ability to use its own IT-related resources and capabilities to leverage the resources of the vendor. Since cloud computing is generally characterized as an IT service (with the vendor providing and maintaining the software and hardware infrastructure), the ability of the client organization to integrate and utilize the vendor's services determines the extent IT benefits are likely to be achieved. Organization- specific capabilities related to implementation, integration, and utilization of cloud services play a key role in deployment performance [3]. There are three IT-related capabilities— technical, managerial, and relational—characterized as a major potential source of competitive advantage. Organization-Specific capabilities that can be a source of competitive advantages are [3]:

- **Technical.** IT resources giving the organization functionality, flexibility, and scalability.
- **Managerial.** Human IT resources resulting from training, experience, and insight.
- **Relational.** Ability to develop positive associations with IT providers characterized by trust.

3. GREEN COMPUTING

Green Computing, or Green IT, is the practice of implementing policies and procedures that improve the efficiency of computing resources in such a way as to reduce the energy consumption and environmental impact of their utilization. As High Performance Computing is becoming popular in commercial and consumer IT applications, it needs the ability to gain rapid and scalable access to high end computing capabilities. This computing infrastructure is provided by cloud computing by making use of datacenters. It helps the HPC users in an on-demand and payable access to their applications and data, anywhere from a cloud [6]. Cloud computing data-centers have been enabled by high-speed computer networks that allow applications to run more efficiently on these remote, broadband computer networks, compared to local personal computers. These data-centers cost less for application hosting and operation than individual application software licenses running on clusters of on-site computer clusters. However, the explosion of cloud computing networks and the growing demand drastically increases the energy consumption of data-centers, which has become a critical issue and a major concern for both industry and society. This increase in energy consumption not only increases energy cost but also increases carbon-emission. High energy cost results in reducing cloud providers' profit margin and high carbon emission is not good for the environment. Hence, energy-efficient solutions that can address the high energy consumption, both from the perspective of the cloud provider and the environment are required. This is a dire need of cloud computing to achieve Green computing. This whole scenario is depicted in Figure1 [1].

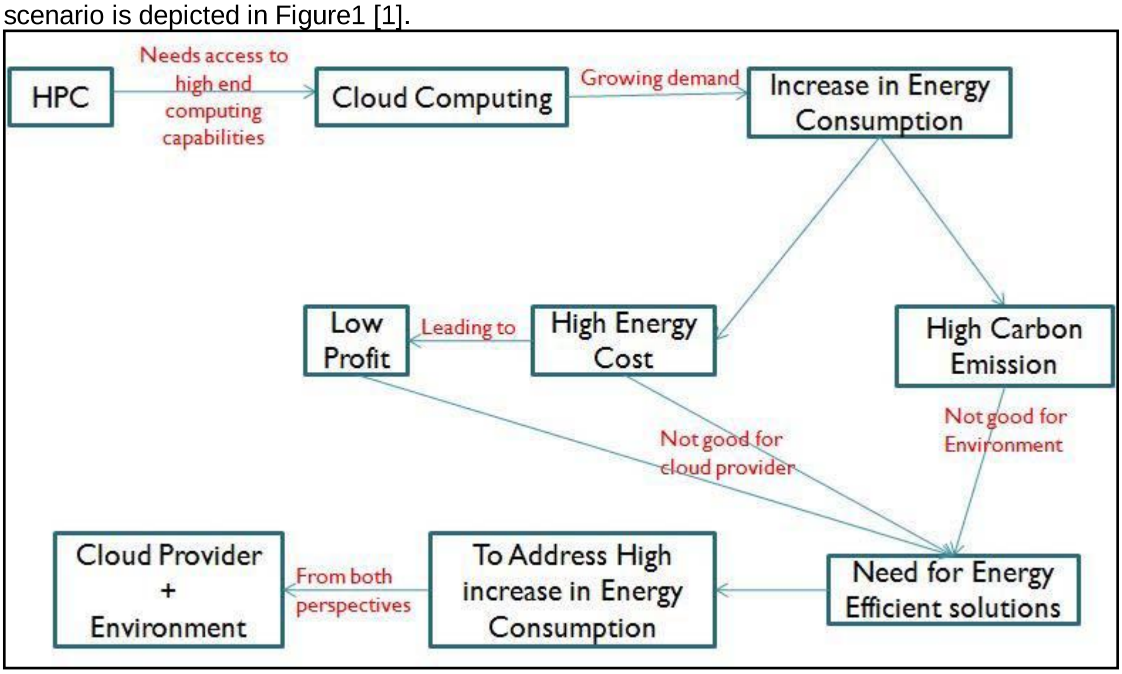
HPC
Needs access to high end computing capabilities
Cloud Computing
Growing demand
Increase in Energy Consumption
Low Profit
Leading to
High Energy Cost
High Carbon Emission
Not good for cloud provider
Not good for Environment
Cloud Provider + Environment
From both perspectives
To Address High increase in Energy Consumption
Need for Energy Efficient solutions

Figure 1: Green Computing in Clouds

According to reference [9], there are following four key factors that have enabled the Cloud computing to lower energy usage and carbon emissions from ICT. Due to these Cloud features, organizations can reduce carbon emissions by at least 30% per user by moving their applications to the Cloud. These savings are driven by the high efficiency of large scale Cloud data centers.

A. **Dynamic Provisioning:** In traditional setting there are two reasons for over-provisioning: first, it is very difficult to predict the demand at a time and second, to guarantee availability of services and to maintain certain level of service quality to end users. Cloud providers monitor and predict the demand and thus allocate resources according to demand. Those applications that require less number of resources can be consolidated on the same server. Thus, datacenters always maintain the active servers according to current demand, which results in low energy consumption than the conservative approach of over-provisioning.
B. **Multi-tenancy:** The smaller fluctuation in demand results in better prediction and results in greater energy savings. Using multi-tenancy approach, Cloud computing infrastructure reduces overall energy usage and associated carbon emissions. The SaaS providers serve multiple companies on same infrastructure and software. This approach is obviously more energy efficient than multiple copies of software installed on different infrastructure.
C. **Server Utilization:** High utilization of server results in more power consumption, server running at higher utilization can process more workload with similar power usage. Using virtualization technologies, multiple applications can be hosted and executed on the same server in isolation, thus lead to utilization levels up to 70%.
D. **Datacenter Efficiency:** By using the most energy efficient technologies, Cloud providers can significantly improve the PUE of their datacenters. We shall explain this factor more in the next section.

3.1 Towards "Green" Data Centers

The Cloud datacenters are quite different from traditional hosting facilities. A cloud datacenter could comprise of many hundreds or thousands of networked computers with their corresponding storage and networking subsystems, power distribution and conditioning equipment, and cooling infrastructures.. A data center hosts computational power, storage and applications required to support an enterprise business. A data center is central to modern IT infrastructure, as all enterprise content is sourced from or passes through it. Data centers can be broadly classified, on the basis of power and cooling layout, into one of the 4 tiers [8]:

- Tier 1: Single path for power and cooling; no redundant components;
- Tier 2: Redundancy added to Tier 1, thereby improving availability;
- Tier 3: Multiple power and cooling distribution paths, of which one is active;
- Tier 4: Two active power and cooling paths, and redundant components on each path.

This classification however, is not precise and commercial data centers typically fall between Tiers 3 and 4 [8]. A higher tier implies an improvement in resource availability and reliability, but it comes at the expense of an increase in power consumption. Data centers host services that require high availability, close to 99.99%. Fault tolerance, therefore, becomes imperative. The loss of one or more components must not cause the data center to terminate its services to clients. Consequently, data centers feature hardware redundancy. Furthermore, data centers are designed for a peak load, which might be observed only occasionally, and for short bursts of time. This conservative design results in over provisioning of hardware in a data center. All these factors combined together contribute to the high power consumption of data centers. So the electricity usage is the most expensive portion of a data center's operational costs. There are two major and complementary methods [8] to build a green data center:

A. Utilize green elements in the design and building process of a data center.
B. Greenify the process of running and operating a data center in everyday usage.

A data center is composed of several components involved in the power consumption (Figure 1):

A. Cooling equipment (e.g., computer room air-conditioner units);
B. Power equipment (e.g., uninterruptible power supplies and power distributions units);
C. IT equipment (e.g., rack optimized and non-rack optimized enterprise servers, blade servers, storage and networking equipment);
D. Miscellaneous equipment (e.g., lighting).

The cooling system occupies a significant part, in terms of energy consumption, in a data center. The degree of redundancy in the system is a very important consideration in the infrastructure. The data center should not run out of energy so for this there are a number of sources of supply and battery backup; the more the system is redundant, the greater the fault tolerance, but at the same time, the costs of installation and maintenance of this system will be higher. The IT equipment is the main part of a data center, which makes possible its basic operations, and contains the electronics used for computing (servers), storing data (data storage) and communicating (network). Collectively, these components process, store and transmit digital information.

3.2 Power/Energy Metrics

The popular power related performance metrics are shown in figure 2 [8]

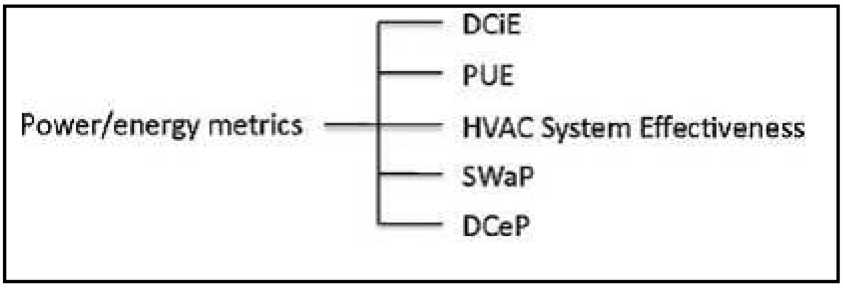

Figure 2: Power consumption metrics taxonomy

3.2.1 DCiE

The Data Center Infrastructure Efficiency, DCiE or DCE is an industry accepted metric. It is defined as:

$$\text{DCiE} = \frac{\text{ITEquipmentPower}}{\mathit{TotalFacility\ \ Power}} \qquad \text{…………Equation (1)}$$

where *IT Equipment Power* includes the load associated with all of the IT equipment, i.e., computation, storage and network equipment, along with supplemental equipment such as switches, monitors, and workstations/laptops used to monitor or otherwise control the data center. *Total Facility Power* includes all IT Equipment power as described above plus everything that supports the IT equipment load such as: power delivery components i.e., UPS, switch gear, generators, PDUs, batteries and distribution losses external to the IT equipment; Cooling system components i.e., chillers, computer room air conditioning units, direct expansion air handler units, pumps, and cooling towers. Other miscellaneous component loads such as data center lighting are also included. A DCiE value of about 0.5 is considered typical practice and 0.7 and above is better practice. Some data centers are capable of achieving 0.9 or higher.

3.2.2 PUE

An organization for green computing provides the following standard metric for calculating the energy efficiency of the data centers, Power Usage Effectiveness (PUE) [8]:

$$\text{PUE} = \frac{1}{\text{DCiE}} = \frac{\text{TotalFacilityPower}}{\text{ITEquipmentPower}} \qquad \text{…….. Equation(2)}$$

PUE shows the relation between the energy used by IT equipments and energy used by other facilities, such as cooling needed for operating the IT equipment. For example, a PUE of 2.0 indicates that for every watt of IT power, an additional watt is consumed to cool and distribute power to the IT equipment. At the present time, the PUE of a typical enterprise data center falls between the ranges.

3.2.3 HVAC System Effectiveness

The HVAC (Heating, Ventilation, and Air Conditioning) system of a data center typically includes the computer room air conditioning and ventilation, a large central cooling plant, and lighting and other minor loads. The HVAC System Effectiveness is the fraction of the IT equipment energy to the HVAC system energy. The HVAC system energy is the sum of the electrical energy for cooling, fan movement, and any other HVAC energy use like steam or chilled water. The HVAC System Effectiveness is calculated as follows:

HVAC -Effectiveness = $\frac{IT}{HVAC+(FUel+Steam\ +ChilledWater\ )*293}$ …….Equation (3)

where *IT* is the annual IT Electrical Energy Use, *HVAC* is the annual HVAC Electrical Energy Use, *Fuel* is the annual Fuel Energy Use, *Steam* is the annual District Steam Energy Use and *Chilled Water* is the annual District Chilled Water Energy Use. The HVAC System Effectiveness denotes the overall efficiency potential for HVAC systems. A higher value of this metric means higher potential to reduce HVAC energy use .

3.2.4 SWaP

he metric of SWaP (Space, Watts and Performance) [20] characterizes a data center's energy efficiency by introducing three parameters of space, energy and performance together. The SWaP is calculated as follows:

SWaP = $\frac{Performance}{(Space\ *PowerConsumption\ )}$ ……… Equation (4)

where *Performance* is computed using industry standard benchmarks, *Space* measures the height of the server in rack units (RUs) and *Power* Determines the watts consumed by the system, using data from actual benchmark runs or vendor site planning guides. The SWaP metric gives users an effective cross comparison and total view of a server's overall efficiency. Users are able to accurately compare the performance of different servers and determine which ones deliver the optimum performance for their needs. The SWaP can help users better plan for current and future needs and control their data center costs.

3.2.5 DCeP

The DCeP is proposed to characterize the resource consumed for useful work in a data center. The DCeP is calculated as follows:

DCeP = $\frac{Useful\ -Work\ -Produced}{Total\ -energy\ -consumed\ \ -to-produce\ \ -that-work}$ …….. Equation (5)

Useful work is the tasks performed by the hardware within an assessment window. The calculation of total energy consumed is the kWh of the hardware times the PUE of the facility.

4. CASE STUDY: LOAD PREDICTION

There are a few common approaches to predict network loads such as linear prediction, neural networks and fuzzy logic. The distinction in these prediction approaches is that the models are based on either stochastic or discrete modeling. Stochastic models use distributions to service a user request, and the time it takes for the device to transition between its power states (i.e. sleep, idle and active). In discrete models, the transition state times are fixed. However, assuming that the time it takes to change from one state to another is fixed may not true in all scenarios. For example, the equipment may be degrading over time and therefore the transition times may vary. [7][8]. Neural network prediction estimates the dynamic incoming load, and submits it to controller that takes a decision for turning servers on and off. These decisions work to minimize the number of currently running servers. Multi-layer neural network have been applied to predict the future load in cloud data centers. We predict the forecasted load by training ANN in a supervised manner with a backpropagation algorithm. A three- layer neural network predictor is illustrated in fig. 2. The neural model receives five inputs (from I_1 to I_5) from external information source that represent the number of requests in five continues times and is applied to the input layer. The predicted load is then calculated at the output node (the number of request at time six I_6). The number of neurons in the input layer is five to represent the actual load in five continues time. The value for the look ahead intervals used in the simulations was one second. There is one neuron in the output layer. The values for the look ahead intervals used in the simulations and the number of inputs in the output layer of ANN were selected by experiences works on Math Lap software, see figure 8. Once the network was trained within a tolerable error, the network is tested with different data, and both the Mean Squared Error (MSE) and the Root Mean Squared Error (RMSE) were calculated from the predicted results.

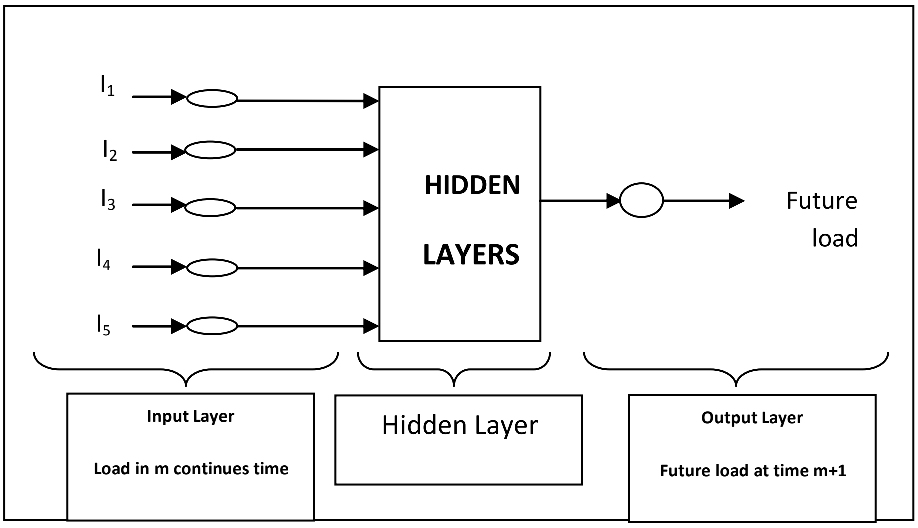

Figure 3: Neural network prediction

4.1 Prediction Algorithm

The training phase process consist the following steps:

1. **Define the ANN architecture, number of hidden layers, number of input nodes(5) in input layer, and number of output node (1) in output layer.**
2. **Define training, test, and validation sets. Each set is represented as N X 6 matrix, where N is the number of examples in that set, and the columns are used as follow: the first 5 columns specify the inputs to ANN (cloud’s workload at five consecutive times), whereas the last column specifies the target output(the predicted cloud’s workload at time six).**
3. **The training process will initiate to train the ANNs. The training process will continue till the extracted outputs reached to target outputs.**
4. **When the error percent reached to acceptable value, the training process will stop and the ANN will consider trained.**

5. CONCLUSION

Cloud computing cannot be claimed to be Green. What is important is to make its usage more carbon efficient both from user and provider’s perspective. Cloud Providers need to reduce the electricity demand of Clouds and take major steps in using renewable energy sources rather than just looking for cost minimization. Neural network prediction estimates the dynamic incoming load, and submits it to controller that takes a decision for turning servers on and off. These decisions work to minimize the number of currently running servers. Multi-layer neural network have been applied to predict the future load in cloud data centers.

REFERENCES

1. Abdulaziz Aljabre, “ Cloud Computing for Increased Business Value”, *International Journal of Business and Social Science, Vol. 3 No. 1; January 2012.*

2. Mujtaba Talebi and Thomas Way, “Methods, Metrics and Motivation for a Green Computer Science Program”, SIGCSE’09, March 3–7, 2009, Chattanooga, Tennessee, USA. Copyright 2009 ACM 978-1-60558-183-5/09/03...$5.00.
3. Bhatt, G. and Grover, ”Types of information technology capabilities and their role in competitive advantage: An empirical study”, Journal of Management Information Systems 22, 2 (2005), 253–277.
4. Saurabh Kumar Garg and Rajkumar Buyya, “ Green Cloud computing and Environmental Sustainability”, Cloud computing and Distributed Systems (CLOUDS) Laboratory
5. K. Dinesh, G. Poornima, K.Kiruthika, " Efficient Resources Allocation for Different Jobs in Cloud", *International Journal of Computer Applications (0975 – 8887) Volume 56– No.10, October 2012*
6. Gary Garrison, Sanghyun Kim, and Robin L. Wakefield, " Success Factors for Deploying Cloud Computing", communications of the ACM | September 2012 | vol. 55 | no. 9, doi:10.1145/2330667.2330685.
7. Jayant Baliga, Robert W. A. Ayre, Kerry Hinton, andRodney S. Tucker, “Green Cloud Computing:Balancing Energy in Processing, Storage, and Transport “,Vol. 99, No. 1, January 2011 | Proceedings of the IEEE.
8. Osvaldo Marra, Maria Mirto, Massimo Cafaro and Giovanni Aloisio, “Green Computing and power saving in HPC data centers ”, Research Papers Issue 2011 December 2011 *Scientific Computing and operation (SCO)*
9. Nidhi Jain Kansal and Inderveer Chana, “ Existing Load Balancing Techniques in Cloud Computing: A Systematic Re-View”, Journal of Information Systems and Communication ISSN: 0976-8742, E-ISSN: 0976-8750, Volume 3, Issue 1, 2012, pp- 87-91. Available online at http://www.bioinfo.in/contents.php?id=45
10. Nidhi Jain Kansal1, Inderveer Chana2, “Cloud Load Balancing Techniques : A Step Towards Green Computing “,IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 1, January 2012 , ISSN (Online): 1694-0814, www.IJCSI.org