Journal of Network Communications and Emerging Technologies (JNCET)
Volume 10, Issue 6, June (2020)

Efficient Strategies for Green Cloud Computing

P Aishwarya Naidu
Department of Information Science and Engineering, B.M.S College of Engineering, Bangalore, India
1bm16is062@bmsce.ac.in

Pooja Chadha
Department of Information Science and Engineering, B.M.S College of Engineering, Bangalore, India
1bm16is064@bmsce.ac.in

Nalina V
Department of Information Science and Engineering, B.M.S College of Engineering, Bangalore, India.
nalinav.ise@bmsce.ac.in

Published online: 30 June 2020

Abstract – Due to increased energy demand, the rising economic and environmental costs of data centres are becoming a real concern. 'Green Data Centres' applies to energy-conscious, energy-efficient, and CO2 emissions reducing architectures, protocols, tools, infrastructures, and algorithms in data centres. Today's data centres are equipped for peak processing. But, most of the time, servers are seen to be idle. Idle servers and related components of the network consume a significant amount of resources. We need to keep track of the Data Centres to avoid these problems. There are also ways of reducing energy wastage, the power costs. In today's world, establishing the Green Data Centre is needed. Through utilizing a green computing method, the data centres become green to reduce the electricity consumed by the data centres. The facilities in the data centres can be constructed in such a way that they use green computing to consume only minimal energy. In this paper, we describe potential research enablers for the green data centre and green metrics specific to data centres. We discuss energy-saving solutions for servers, network infrastructure, and additional green solutions.

Index Terms – Green Computing, Cloud, Data Centre, Virtual Machine, Energy Efficient.

1. INTRODUCTION

Green technology has become a major concern for governments and corporations all across the world because of its environmental, economic, and political aspects. Due to the increase in internet usage and the various advantages that the cloud offers such as scalability, increased collaboration, fault tolerance, and high availability, companies are starting to shift to using Cloud services. Cloud computing offers shared pools of services and resources like servers, software, databases, monitoring, and more from a centralized location to its users worldwide rapidly and with minimal management [1]. It is an alternative to an on-premise data centre which is considered more expensive because of infrastructure setup expenditures [2]. Some of the popular players of Cloud include Amazon Web Services, Google Docs, and Azure.

Cloud computing is an emerging field that makes efficient use of computing resources. Data centres which host cloud services consume large amounts of energy, resulting in high operating expenses and atmospheric carbon footprints. The growing use of video-on-demand and cloud storage technology has led to the worldwide development of many new data centers. Data centres should be easily available, and fault-tolerant. Within a data center, cooling, network components and computing services are the main sources of power consumption [3, 4]. Such requirements apply to high data centre energy use.

Therefore, Green Cloud computing reforms are critical not only to conserve environmentally sustainable energy but also to reduce operational charges. A growing cloud host is a computing node in the cloud setting that executes a function or subtask that allows user-based programs to deploy in. Cloud users only pay organizations when they use computing resources. Cloud consumers, irrespective of where the applications are being stored, may access software or tools depending on their demands.

The prevalent energy-conscious techniques in cloud environments must be emphasized. Research and development of processes and methods for consolidation of energy-aware activities for data centres must be undertaken in such a manner as to make cloud computing a feasible ecological technology for future generations to achieve cost-effective, coordinated, and technological progress.

Changes in networking, computing, cooling, clouding, and virtualization will have to be done to get green data centres. Energy-based efficient resource scheduling frameworks can help companies reduce power use, and directly contribute to the growth and success of the business. Servers can achieve improved efficiency through resource scheduling and complex optimization of the workload. In this paper, we review a few strategies to apply green IT using cloud computing.

2. RELATED WORK

In [5], the literature proposes a VM-based framework in cloud computing for the adaptive management of virtualized resources, using feedback control theory and an adaptive controller model that dynamically controls the use of several virtualized resources to satisfy the cloud computing service level objective (SLO) criterion. KVM is chosen as a virtual machine monitor (VMM) to implement the architecture in comparison to Xen. The evaluation of the proposed controller model showed that the model could effectively allocate resources to achieve SLO applications in response to continually changing resource requirements for various applications distributed over multiple VMs within the virtual resource pool.

In [6], the authors propose two adaptive scheduling algorithms for allocation of resource mechanisms. Cloud users can request multiple cloud services at the same time. In this scenario, the efficiency of parallel computing in the cloud network may be improved. When implementing parallel processing in cloud computing, a system for allocating resources and scheduling the execution order of tasks must be implemented. Additionally, a resource management system with pre-emptive task execution will maximize cloud usage. In this paper, the authors suggested an adaptive resource management algorithm with pre-emptive tasks for the cloud system. The algorithms adjust the allocation of the resource to the actual mission executions based on the changes. Results have shown that models have succeeded even when resource dispute is extreme.

Walsh et al. [7] present a distributed architecture and illustrate how the utility functions can enable the collection of decentralized elements to continuously optimize the use of computational resources in a complex environment. To achieve this, the authors used a functional prototype data centre. Architecture is typically a two-tier system with separate, autonomous components that encourage versatility. The authors present statistical data showing the effectiveness of the scheme in handling realistic, fluctuating web-based transactional workloads operating on a Linux cluster.

In [8], the authors present a novel framework for dynamic autonomous cloud resource management. The main contribution of this research is twofold. First, a distributed approach in which strongly tied autonomous node agents in a data centre conducts resource management that is decomposed into individual activities. Second, the Autonomous Node Agents execute simultaneous modifications using the Multiple Criteria Decision Analysis PROMETHEE method. The findings of the analysis demonstrate that the suggested approach offers scalability, feasibility, and flexibility.

In [9], Kim et al. discuss how reorganizing virtualized servers into fewer physical servers leads to flexible resource provision and improved energy efficiency. A model for estimating each virtual machine's energy consumption without dedicated metering hardware. The authors also proposed a virtual machine scheduling algorithm based on an estimation model. According to the energy budget of each virtual machine, the algorithm provides computing resources. Xen virtualization system was used for implementing the model, and an estimated error of less than 5 percent of the total energy consumption and provided energy consumption was recorded.

[10] Emphasizes the increasing demand for computing and memory and distributed computing systems. As the net network benefit depends on how the network can reach the Service Level Agreements (SLA) agreed upon with the consumers, resource management becomes one of the most critical problems in distributed systems. The paper considers an SLA-based resource allocation problem in cloud computing for multi-tiered applications. A force-directed search algorithm and an upper limit on the overall income is proposed to solve the problem. The three aspects on which optimization is carried out are Processing, power capacity, and communication capabilities. Simulation results show that the proposed algorithm is efficient.

Computing and cooling consume the greatest amount of energy in data centres. But these two aspects are not looked into combined. There needs to be some coordination between these aspects to optimize energy consumption to the maximum extent in data centres since they lack an existing system that jointly controls their operations. Torrens et al. [11] propose an architecture called GENiC, an integrated energy management system that addresses this issue for data centre wise optimization. It captures monitoring, cooling, waste heat recovery, control of the IT workload, and local power generation to achieve energy efficiency. The optimization happens at different estimated lengths of times: short term predictions and long term predictions. The authors also propose a strategy to optimize the allocation of Virtual Machines (VMs) while unused servers are switched off.

Virtual Machine (VM) migration is a technique that has recently emerged in data centres. It shifts the virtual machine from one physical hardware to another in case of unscheduled server downtime or because of some fault in the server. This is done so that there is high availability. VM migration can also be performed for disaster recovery purposes. In [12], the authors propose that energy should be considered an important factor in virtual machine migration. They describe a system architecture that provisions the lives migration of virtual machines based on a favourable number of servers and also carefully switches off underutilized services. The architecture consists of modules Service Request Processing, Cloud Controllers which includes the Trigger Engine and pre-processed data. The Trigger Engine uses pre-processes data to automatically initiate migration or to turn off the server.

In [13], the authors present a new approach to schedule resource allocation for Cloud computing based on energy optimization based on a genetic algorithm. In this case, the literature focuses on the geographical location of the data centre and also schedules allocations dynamically based on predictions instead of static allocations methods such as FCFS. The algorithm primarily focuses on Infrastructure as a Service (IaaS). The energy required (Ecal) is calculated based on the model of CMOS processors by adding constraints: run time and number of processors required to run the application. The cooling energy factor (Eaux) is a ratio between the energy required and the energy performance coefficient. Therefore, the total energy to minimize is the sum of energy required (Ecal) and cooling energy (Eaux). The experiments show that the approach gives a reduction in consumption of energy and CO2 emission when compared to static resource allocation techniques.

Virtual machines must be assigned to the correct host nodes on the network in such a manner that a greater percentage of service or resource demands are met in less time. Therefore, the hypervisor will run a qualified algorithm for resource allocation purposes. In [14], Lee et al. purpose an algorithm that efficiently allocated VM requests to the physical nodes in a best-fit strategy. This goal is attained by designing a performance analysis scheme for each node which takes into consideration the specifications on cores, CPU, and memory.

The literature [15] focuses on optimizing the process of resource allocation using an improved Clonal Selection Algorithm (CSA) since it offers a global exploration ability in a viable solution range and runs in less time. This algorithm is effective in optimizing resource allocation when compared against existing algorithms based on the evaluation studies. The energy optimization model is built on dynamic voltage and frequency scaling (DVFS). It specifies that the power of a resource depends on resource frequency and its voltage supply. The ICSA is used for resource allocation as follows: nodes are measured over an affinity function and are thus sorted in that affinity order. The nodes are chosen based on the maximum affinity.

Due to the increase in internet traffic we see a rise in network energy consumption as well. This must be tackled if we intend to guarantee sustainability. Because data centres are responsible for the majority amount of consumption of energy, we need to improve the energy efficiency at the architectural and service levels. In [16], a unified network architecture is proposed by Fiorani el at. It combines intra-data-centre and inter-data-centre connectivity. This architecture is most appropriate for carrier cloud models. Both the telecom infrastructure and the data centre are operated and owned by the same organization in the carrier cloud. The paper also talks about the impact of distributed video servers and their related energy consumption levels.

3. ENERGY SAVING STRATEGIES

There are various factors that need to be considered for energy efficiency in data centres. The first step to green data centres is to define certain green metrics based on which algorithms can be quantified. The most common metric is Power usage efficiency (PUE) which is calculated by total system energy to energy used by data centre equipment.

3.1. Servers

The servers are the primary cause of energy use in data centres. Usually, only 30 percent of the time servers are busy. The resource usage of the computer server is equal to the use of the CPU. The idle server uses approximately two-thirds of its peak load usage to maintain memory, disks, and I / O services running, while one-third of its usage decreases linearly with the use of CPU. All servers available are either in mode or idle to respond to the immediate peak load.

Metric	Description	Formulation
PUE	Power Usage Effectiveness	$PUE = \frac{Total\ facility\ energy}{IT\ equipment\ energy}$
CUE	Carbon Usage Effectiveness	$CUE = \frac{Total\ CO_2\ emission\ from\ total\ data\ center\ energy}{IT\ equipment\ energy}$
WUE	Water Usage Effectiveness	$WUE = \frac{Annual\ site\ water\ usage}{IT\ equipment\ energy}$
ERF	Energy Reuse Factor	$ERF = \frac{Reuse\ energy}{IT\ equipment\ energy}$
ERE	Energy Reuse Effectiveness	$ERE = \frac{Total\ energy - Reused\ energy}{IT\ equipment\ energy} = (1 - ERF) \times PUE$
DCiE	Data Center Infrastructure Efficiency	$DCiE = \frac{1}{PUE} = \frac{IT\ equipment\ power}{Total\ facility\ power} \times 100\%$
DCP	Data Center Productivity	$DCP = \frac{Useful\ work}{Total\ facility\ power}$
ERP	Energy-Response time Product (Energy delay product)	$ERP^{\pi} = E[P^{\pi}] \times E[T^{\pi}]$ ($E[P^{\pi}]$ = average power consumed under policy π, $E[T^{\pi}]$ = mean customer response time under policy π)

Figure 1 Data Centre Green Metrics

Leaving servers in idle mode contributes to a needless waste of energy. Two main approaches can be used to decrease energy usage in data servers. One of these is dynamic voltage frequency scaling (DVFS) [17]. The DVFS system modifies the power of the CPU (performance level) to suit the load provided. This is meant mainly to improve the energy utilization of CPUs, whereas current computer system modules remain unchanged and continue to operate at their regular energy cost. The second approach which can be used is dynamic power production (DPM). The DPM conserves more resources by powering down all of the computer server components. You save a lot of resources by shutting down the idle server. However, switching on and off the computer needs a considerable setup price. Setup costs take the form of both a delay in time (setup time) and an energy penalty. Another choice is putting servers idle in sleep mode. The server in sleep mode requires more energy than the off-server, but the installation cost is lower than the off-server installation cost. Algorithms are needed to measure the amount of servers that would have to be operational to fulfill current needs. This is dynamically referred to as right-sizing.

It is feasible to apply the Distributed Robust Autoscaling Policy (DRAS) in computationally expensive application farms. DRAS depends on dynamic resource changes that are load-dependent without the need for any potential load estimation or feedback control. The objective is to reduce the average power consumption and mean response time. The difference between when the task comes and when it comes to an end is called the response time. Servers are not switched off as soon as they go idle as the findings from a few studies show that waiting for some time before shutting down inactive computers results in a 40 percent and 83% decrease in energy consumption in response time. The reasoning for this observation is that, if we quickly shut down idle servers, fresh requests do not hit the idle site. A new request will have to spend a long time waiting for an already busy server to go idle, or for it to turn on.

Similarly, when a request comes and no idle servers are detected, powered off servers are not switched on until more requests occur. This would result in several servers being switched on as a burst of jobs occurs in the network, some of which might not be used later. So at the cost of response time, DRAS offers energy savings.

3.2. Networking

The Network infrastructure is the second major cause of energy use in data centres. The data centre's network absorbs approximately 30 percent of the overall computing resource usage. The network architecture consists of the links and the switches. To get green data centres we should understand the energy consumption of all links and switches. A link's power use isn't strictly proportional to its usage [18]. A link's energy consumption depends more on its efficiency than its use. Adaptive link rate methods can be used to decrease the power consumption of links. It would dynamically change the links' data rate to match traffic requirements. Sleep mode technique can also be used to gain energy efficiency by turning off the system or switching a subset of idle modules to sleep mode. The thing to be kept in mind is that quality of service should not be affected and fault tolerance for the data centre networks should not be compromised. A switch's power usage varies with the amount of ports and line cards used. Energy-aware algorithms can be made to minimize the power consumption of switches. Instead of putting the network device completely in sleep mode, components may be placed into sleep mode. Disabling a port switch while no traffic is being sent, adjusting dynamically the forwarding capability (10 Mbps, 100 Mbps or 1 Gbps) of individual ports depending on their load, shutting offline cards that do not have active ports in multi-line card systems, shutting off the light that is not in operation are some of the ways that this may be accomplished.

There is an approach called DENS which takes account of network knowledge while allowing energy-efficient planning in data centres. The DENS solution seeks to minimize energy usage by scheduling jobs according to the best match method, taking into consideration the network components' load level and connectivity capacity when certain jobs involve low computational load while delivering large data streams powered from the data centre. Hence the main goal is to prevent congestion while reducing the amount of servers on the network. DENS maintains the QoS standards for jobs with marginal energy usage rise.

3.3. Additional Green Solutions

3.3.1. Virtualization

Virtualization is one of the most notable costs and energy-saving technology that is rapidly evolving in the IT landscape. Virtualization is based on installing multiple virtual machines (VMs) onto a physical server. This results in reducing the amount of hardware used and increasing resource utilization. Virtualization is possible by using a software called a hypervisor. Some vendors that provide virtualization tools include VMware and Xen. Virtualization can help to reduce the physical servers by organizing the workloads while maintaining the same computing capability. This can provide energy savings.

Buyya et al. [19] provided scheduling algorithms and energy-efficient allocation of resources that take into account the QoS criteria. In cloud applications, the authors concentrated on heterogeneous workloads. Algorithms were suggested for mixing and mapping VM to cloud resources, thus resolving energy efficiency concerns. It is an approach of two sections. New requests are accepted in the first part and can be compared with a problem of bin packing where the bin sizes and prices differ. The second part is directed at maximizing the current allocation of VMs to hosts. The authors have applied dynamic server modification to allow energy-conscious resource management by shutting out idle servers. Dynamic reallocation of VMs offers higher energy savings as opposed to static allocation according to simulations.

Consolidation of dynamic workload while growing energy usage in cloud environments can also be accomplished by aggregating the workload with energy-per-transaction parameters dependent on disk use and CPU. This approach is appropriate even to heterogeneous environments but is not known because of the overhead of migration.

The results indicate that cloud systems provide energy-efficient approaches by evaluating all cloud resources such as Storage as a Service, Processing as a Service and Software as a Service in terms of energy consumption in the data centre as well as in the network between the data centre and the customer for transportation and switching. But, where significant transport and switching is involved, Storage as a Service uses more energy than storing on local hard disks.

3.3.2. Cooling

Cooling is also another major source of power consumption from the data centre. To minimize the usage of cooling electricity, corporations tend to take advantage of the natural pre-cooling processes. Facebook, for example, has set up its data centre in Sweden which naturally has dry and cold weather. Google's Belgian data centre lacks a full chiller, running solely on "free" cooling all the time. It has also put its data centre near a hydroelectric dam since when transmitted over long distances with high voltage, there is so much electricity leakage. Microsoft has come up with another solution that is to enable servers to cool easily in the open air.

3.3.3. Nano Data Centres

For nano data centres, it is built on a distributed architecture instead of a conventional client-server model and is better suitable to provide peer-to-peer material [20]. A large number of internationally distributed nano data centres are being planned, rather than a few major data centres. The infrastructure consists of tiny, integrated data centres that are spread along network edges. Data centres are vulnerable to over-supplying high levels of heat dissipation, and rising distance to end-users. The data centres are over-supplied and, considering that the daily load for most of the day is much lower, they must satisfy the peak requirement. In some cases, data centres are expensive to cool off. Consequently, centralization reduces the gap between the data centre and customers. The greater gap from end-users increases the bandwidth-mileage criteria and contributes to the networking equipment's power usage. Those issues will be addressed by nano data centres.

4. CONCLUSION

A significant part of internet traffic is now focused on data centres as a lot of organizations are turning towards the cloud for storage and computations. Energy consumption should be considered while managing servers and networking. Businesses can save energy and money through Green Cloud computing. Various problems related to cloud computing and possible ways to achieve green cloud computing were discussed in the paper. Further research in this area is very important to reduce the economic and environmental issues related to increasing numbers of growing data centres.

REFERENCES

[1] H. Tabrizchi, MK. Rafsanjani. “A survey on security challenges in cloud computing: issues, threats, and solutions.” The Journal of Supercomputing (2020) Feb 28:1-40.

[2] R. Balasubramanian, and M. Aramudhan. “Security Issues: Public vs Private vs Hybrid Cloud Computing.” International Journal of Computer Applications. (2012) Jan 1;55(13).

[3] A. Uchechukwu, K. Li, and Y. Shen. “Energy consumption in cloud computing data centers.” International Journal of Cloud Computing and Services Science (IJ-CLOSER) (2014) Jun 1;3(3):145-62.

[4] D. Cavdar, and F. Alagoz. “A survey of research on greening data centers”. IEEE Global Communications Conference (GLOBECOM) (2012) Dec 3 (pp. 3237-3242). IEEE.

[5] Qiang Li, Qinfen Hao, Limin Xiao, and Zhoujun Li. "Adaptive management of virtualized resources in cloud computing using feedback control." First International Conference on Information Science and Engineering. IEEE, 2009.

[6] Jiayin Li, Meikang Qiu, Jian-Wei Niu, Yu Chen, and Zhong Ming. "Adaptive resource allocation for preemptable jobs in cloud systems." 10th International Conference on Intelligent Systems Design and Applications. IEEE, 2010.

[7] William E. Walsh, Gerald Tesauro, Jeffrey O. Kephart, and Rajarshi Das. "Utility functions in autonomic systems." International Conference on Autonomic Computing, 2004. Proceedings. IEEE, 2004.

[8] Yagiz Onat Yazir, Chris Matthews, Roozbeh Farahbod, Stephen_Neville ,Adel_Guitouni, Sudhakar Ganti, and Yvonne Coady. "Dynamic resource allocation based on distributed multiple criteria decisions in computing cloud." 3rd International Conference on Cloud Computing. 2010.

[9] Kim, Nakku, Jungwook Cho, and Euiseong Seo. "Energy-credit scheduler: an energy-aware virtual machine scheduler for cloud systems." Future Generation Computer Systems 32 (2014): 128-137.

[10] Goudarzi, Hadi, and Massoud Pedram. "Multi-dimensional SLA-based resource allocation for multi-tier cloud computing systems." IEEE 4th International Conference on Cloud Computing. IEEE, 2011.

[11] Torrens, J. Ignacio, Deepak Mehta, Vojtech Zavrel, Diarmuid Grimes, Thomas Scherer, Robert Birke, and D. Pesch. "Integrated Energy Efficient Data Centre Management for Green Cloud Computing." In Proceedings of the Proceedings of the 4th International Conference on Cloud Computing and Services Science, Rome, Italy, pp. 23-25. 2016.

[12] Arthi, T., and H. Shahul Hamead. "Energy aware cloud service provisioning approach for green computing environment." In 2013 International Conference on Energy Efficient Technologies for Sustainability, pp. 139-144. IEEE, 2013.

[13] Azaiez, Meriem, Walid Chainbi, and Hanen Chihi. "A Green Model of Cloud Resources Provisioning." In CLOSER, pp. 135-142. 2014.

[14] Lee, Hwa Min, Young-Sik Jeong, and Haeng Jin Jang. "Performance analysis based resource allocation for green cloud computing." The Journal of Supercomputing 69, no. 3 (2014): 1013-1026.

[15] Long, Zhuqian, and Wentian Ji. "Power-efficient immune clonal optimization and dynamic load balancing for low energy consumption and high efficiency in green cloud computing." JCM 11, no. 6 (2016): 558-563.

[16] Fiorani, Matteo, Slavisa Aleksic, Paolo Monti, Jiajia Chen, Maurizio Casoni, and Lena Wosinska. "Energy efficiency of an integrated intra-data-center and core network with edge caching." Journal of Optical Communications and Networking 6, no. 4 (2014): 421-432.

[17] Pouwelse, Johan, Koen Langendoen, and Henk Sips. "Energy priority scheduling for variable voltage processors." In Proceedings of the 2001 international symposium on Low power electronics and design, pp. 28-33. 2001.

[18] Chabarek, Joseph, Joel Sommers, Paul Barford, Cristian Estan, David Tsiang, and Steve Wright. "Power awareness in network design and routing." In IEEE INFOCOM 2008-The 27th Conference on Computer Communications, pp. 457-465. IEEE, 2008.

[19] Buyya, Rajkumar, Anton Beloglazov, and Jemal Abawajy. "Energy-efficient management of data center resources for cloud computing: a vision, architectural elements, and open challenges." arXiv preprint arXiv:1006.0308 (2010).

[20] Valancius, Vytautas, Nikolaos Laoutaris, Laurent Massoulié, Christophe Diot, and Pablo Rodriguez. "Greening the internet with nano data centers." In Proceedings of the 5th international conference on Emerging networking experiments and technologies, pp. 37-48. 2009.

Authors

P Aishwarya Naidu is currently in her final year pursuing Bachelor of Engineering in Information Science and Engineering in B.M.S College of Engineering, Bangalore affiliated to Viveswaraya Technological University.

Pooja Chadha is currently in her final year pursuing Bachelor of Engineering in Information Science and Engineering in B.M.S College of Engineering, Bangalore affiliated to Viveswaraya Technological University.

Nalina V is working as Assistant Professor in the Department of Information Science and Engineering at BMS College of Engineering, Bangalore, India. She has received BE (VTU Belgaum) in Computer Science and Engineering in 2008 and M.Tech in Computer Network Engineering from Dayananda Sagar College of Engineering (VTU), Bangalore in 2010.