A Review on A Greener Approach to Information Technology: Green Cloud Computing

Ashish Kumbhare*, Naveen Kumar Vaishnav, Kiran Midiyam, G. Dileep Naidu

Faculty of Science and Tchnology, The ICFAI University, Raipur, India
**Corresponding Author: Email: ashishkumbhare99@gmail.com*

ABSTRACT
Cloud computing is a popular technology because it solves big computational problems by pooling resources and provides a variety of additional services through the internet. The need to improve cloud performance has arisen as a result of computational offloading and the growing requirement for quick and suitable responses for real-time applications. Furthermore, the enormous rise in data and requests has necessitated the construction of a significant number of data centers throughout the world. As a result of the high energy consumption of data centers, this raises the economic and environmental costs. Furthermore, the production of CO2 and other greenhouse gases has opened the way for the expansion of the carbon footprint and, as a result, the greenhouse effect. Green technology is propelled by these forces. As a result, energy-efficient solutions that make the most use of existing resources are required. This article gives a summary of the cloud system, as well as the metrics that must be satisfied in order to have a green cloud computing flavor, and the existing methods.

Keywords: *Cloud Computing, Green Cloud computing, Virtualization, Workload Consolidation, Migration.*

1. INTRODUCTION

The need to create new techniques and approaches in the domain of cloud computing, which has acquired speed and relevance owing to the manner it offers information and services, is being influenced by rapid growth in all sectors, either directly or indirectly.

A. Cloud Computing

Cloud computing offers on-demand resource delivery, broad network connectivity, scalability, multi-tenancy, and a large range of services [13]. All of this is monitored, regulated, and taxed in accordance with consumption.

1. Cloud Computing Service Models: Infrastructure as a Service (IaaS), Software as a Service (SaaS), and Platform as a Service (PaaS) are the three service models (PaaS). Customers use IaaS to execute software such as the operating system and applications because it provides a vast number of computing resources such as network, processing, and storage. Customers can use the cloud provider's apps that are easily available through SaaS. PaaS provides a vast number of application programme interfaces to cloud applications, allowing customers to deploy their own apps. This paradigm provides programming languages, as well as the necessary libraries and tools for the development process.

2. Cloud Computing Deployment Models: The private cloud is a cloud that is only supplied for the use of a single organization's customers, and the company that owns it maintains its operation. The public cloud is utilized by the general public, and it is managed by a single company or a group of organizations. The cloud provider is in charge of both the private and public cloud models, as well as their maintenance and monitoring. The community cloud serves a specific community inside an enterprise and might be on-premises or off-premises. Finally, a hybrid cloud is a collection of two or more distinct cloud infrastructures that are linked together utilizing a variety of technologies to enable effective portability and interoperability.

B. Green Cloud Computing

With technological advancements, the next step will be to provide environmentally friendly, cost-effective, and energy-efficient solutions.

1. Energy Concerns in Cloud Computing: The data that is sent to the cloud goes via different devices such as the provider's router, and it passes through network devices such as the gateway router, local area network, storage servers, and Virtual Machines inside the data centre (VMs). Cisco's Global Cloud Index (GCI) [2] depicts the explosive rise in global data centre workloads and IP traffic, as well as the challenges posed by data generated by Internet of Everything (IoE) devices. By 2019, it is expected that 86 percent of all workloads will be processed on the cloud, a figure that is more than double what it is currently. The Natural Resources Defense Council (NRDC) has published a report [1] on data centre energy use, cost, and hazardous gas emissions. The power consumption will be 47 billion kWh, the cost of power will be $ 4.7 billion, and CO2 emissions will be 50 million MT, according to estimates. This information demonstrates the need for solutions that are useful in every way of cloud computing.

2. Green Cloud Computing Methods: Virtualization is a critical technology for enabling green cloud since it maximizes resource usage. For better outcomes, proportional computing is employed, which may be applied to the entire system, specific devices and components, and network protocols. Examples of proportional computing are Dynamic Voltage Scaling and Adaptive Link Rate. By analyzing the application's behavior, compilation technology optimization may be utilized to minimize system or processor power consumption. Implementing new processes for system software and computer architectures, utilizing advance clock gating, lowering cooling needs, power gating, and applying temperature control for hardware may all help save energy. To enhance the overall performance of the system, scheduling methods may be employed at both the user and system level.

The remainder of the paper is laid out as follows: Section II discusses the many study topics for green cloud and the solutions that are available. Section III outlines the metrics to be measured, as well as general methodologies and procedures. Section IV concludes the paper.

2. METHODS

Using a mixed-integer linear programming model, Federico et al [7] proposed a technique for improving network architecture in order to deliver higher QoS while minimising cost and energy usage. The Tabu search technique uses a penalty term in the objective function to minimise overloaded data centres and iterates until a workable solution is discovered. The requests that are supplied in real time vary greatly, and the technique can take this into account, as well as the current technologies used in data centres and hosts, to further improve this method.

Fei et al. [8] presented an interference aware live migration in which a decision process is used to choose the candidate VM that needs to be moved to the suitable PM, reducing both VM migration and co-location interference. The Interference-Aware VM Live Migration Strategy and a multi-resource demand-supply model are used to build this solution. This reduces performance disruption before and after VM migration, and this benefit may be extended to heterogeneous PMs as well, as only data centres with homogenous PMs are evaluated. To minimise the co-located inter-VM latency, Bei et al developed a communication aware scheduling in [3]. It is taken into account the communication between inter-VMs running on the same virtualization platform. Because this approach has only been tested on one virtualization platform, its efficacy on other systems should be assessed. The presence of multi-core architectures, as well as network connection between the migrating VM and the VM on the originating PM, might be considered for a better solution. Xiaomin et al. proposed a method for real-time activities in [15], in which scheduling is given top importance. More VMs are introduced to finish the tasks, and while the hosts are inactive, consolidation is performed to optimise resource use. The suggested technique increases scheduling quality, however it might be even better if other metrics were taken into account while creating the energy model, as current method just considers CPU energy use. To examine the energy-performance tradeoff in IaaS cloud, Dong et al presented a two-stage VM scheduling method [5], which comprises of the static VM placement scheme and the dynamic VM migration strategy. This approach presents the VM-Mig algorithm, which is a hierarchical

clustering algorithm based on the Minimum Cut, Best Fit algorithm. More precise migration cost estimates may be supplied, and optimization concerns that arise can be avoided at the same time.
Haikun et al. offer a method for correlated VM migration [9]. This process is offered for multi-tier applications, and all VMs are made to finish the migrations at the same time. The Stop and Copy phase and the Wait and Copy phase are two essential phases that need first identifying the stable phase and then allocating bandwidth. This method lowers the migration costs associated with moving the virtual machine to a new location. It's very likely that synchronisation will cause very little overhead.
For a better resource provisioning system, Weijia et al proposed a technique based on the Online Bin packing algorithm [14]. The VMs are treated as objects and are split into groups based on their sizes before being placed in bins, which are servers with a 1/3 bin gap.
Variable Item Size Bin Packing is a variation of the relaxed bin packing issue (VISBP).
There are two types of VISBP: one-dimensional and multidimensional. This technique analyses only homogeneous PMs, and a practical demonstration for the number of servers utilised may be supplied for a more detailed examination of this method. Lena et al [12] proposed an approximation and optimum method that takes into account each user's strategic behaviour and enhances the system's performance using a winner selection algorithm. The approach is simple and quick, resulting in near-optimal social welfare and increased income. Alternative approaches might be employed to improve the situation, as the greedy mechanism is the strategy used in this approach. Hui et alin [10] have proposed a workload management strategy for the hybrid cloud model that allows users to use both public cloud services and private data centres. The suggested architecture comprises two zones: a base zone in a local data centre and a cloud-based flash crowd zone.
Based on the normal mode or the factoring mode, a rapid frequent data item detection method is utilised to give the output as base or flash crowd for each incoming request. Yi-Ju et al. [16] offer a method for efficient cloud that also includes cost optimization utilising the cost function. Three power-saving policies, as well as a decision-making process and operating modes, are available. When the request arrival is a Poisson process and they are handled in order of arrival, the assumptions may need to be adjusted since the request that arrives later should be given precedence and calculated first.
A profiling-based technique for combining workloads and moving virtual machines from one host to the other has been proposed by Kejiang et al [11]. The co-location and migration problems are used to create the profiles. The foundation is made up of the co-location performance constraint, migration performance constraint, Consolidation Cases, and Consolidation Scenario. This technique reduces the number of active physical machines and the number of VM migrations that are common. For this approach to be useful to a real-world setting, it must take into account complicated workloads and failure occurrences. Chonglin et al. [4] employed machine learning techniques to build a tree model for monitoring the power consumption of VMs on the same host, using a regression-based model tree as the focus.
By comparing estimated and measured power, a model is built and assessed, and modifications are made depending on the results of the accuracy evaluation. This has a number of benefits, including simple modelling and the use of a tree trimming approach to minimise over-fitting. This technique generates a pattern that varies depending on the application, and thus necessitates a distinct set of parameters, which may result in overhead.

3. RESULTS AND DISCUSSION

A. Data Center Energy Measurements:

The electricity utilised by IT equipment accounts for approximately 30% of the total power consumed by gadgets, according to research. The Green Grid Consortium provides essential indicators such as Power Usage Effectiveness (PUE), Data Center Infrastructure Efficiency (DCiE), and Carbon Usage Effectiveness (CUE). The green versions of the metrics are also available, and the ‘G' in GPUE stands for the total of energy sources and their lifecycles, which is further deduced by weight in kilogrammes of CO_2 per kilowatt hour of energy from power sources.

Table I. Energy Efficient Metrics

Metric	Formula
PUE: Power Usage Effectiveness	Total Facility Power / IT Equipment Power
GPUE: Green Power Usage Effectiveness	G * PUE
CUE: Carbon Usage Effectiveness	Total CO2 Emission / IT Equipment Energy
WUE: Water Usage Effectiveness	Annual Usage of Water / IT Equipment Energy
DCP: Data Center Productivity	Useful Work Done / Total Power Consumed
CPE: Compute Power Efficiency	IT Equipment Utilization / PUE
DCiE : Data Center infrastructure Efficiency	1 / PUE

B. Green Cloud Solutions in General:

1. Renewable Energy Sources: Renewable energy sources such as wind, solar, and hydropower, which are derived from natural resources, are developing energy sources. These may be used to cool the systems and generate the necessary electricity to keep the data center's numerous components running smoothly.
2. Data centre airflow and cooling processes: In data centres, cooling equipment such as chillers and air conditioners, as well as lighting equipment, account for around 35% of total power usage. As a result, effective solutions are offered for reducing waste caused by air movement. Because less air can be provided and the temperature distribution is more evenly distributed, this promotes a welcome shift in the energy consumed.
3. Consolidation and Migration: Consolidation is the most common approach, in which the number of servers is reduced and unneeded servers are turned down to conserve energy. Virtual machines are moved from one machine to another during migration. This may be accomplished by turning off the machines at the source and then turning them back on once the transfer is complete. Cold migration is the term for this. The following step is a hot or live migration, in which the computers are in operation at the time of the switchover. This technique eliminates the downtime caused by cold migration while still delivering superior results. However, the most crucial aspect to remember is that this should occur without causing any disruption to the regular routine.

C. Algorithmic Approaches:

1. First Come, First Serve: The job that comes first will be scheduled first, and resources will be assigned to it according to its requirements. After the job is finished, the next process in the queue is started.
2. Round-Robin: The job will be completed until the time slice, or quantum, that has been set is achieved. The job will then be placed last in the queue, and when it reaches the front end, it will be taken again and the remaining execution will take place.
3. Min-Min Method: This algorithm is focused on the shortest completion time, and minor jobs are completed first, causing bigger work to be delayed or not scheduled at all.
4. The Max-Min Algorithm is used to assess the jobs and create a matrix that represents the execution time. By allocating the available resources to the jobs with the longest completion time, they are completed first.
5. Swarm Optimization: This technology offers clever multi-dimensional solutions for energy-conscious virtual machine consolidation. These are adaptive methods in which the load from heavily utilised machines is transferred to less heavily used devices.

In general, this is a green option since it reduces operating costs through proper job assignment. Swarm optimization has shown to be more reliable and efficient. Ant colony optimization, artificial bee colony optimization, particle swarm optimization, and differential evolution are just a few examples.

4. CONCLUSION

Many companies have adopted the cloud as a result of its widespread appeal and various benefits. As a result, numerous devices are used more often, resulting in higher power dissipation. As a result, the demand for environmentally acceptable solutions to combat the negative impacts of the toxic gases released by the various components of the cloud is increasing. Many algorithms are available as a starting point, and many novel scheduling and allocation strategies are being created. Virtualization, consolidation, and migration techniques are used to deploy virtual machines on real machines and decrease the number of hosts in order to handle over-utilization and under-utilization.

REFERENCES

[1] “Data centers are the new polluters”, http://www.computerworld.com/article/598562/ data-center/data-centers-are-the-newpolluters.html
[2] “Global Cloud Index”,http://www.cisco.com/c/en/us/solutions/service - provider/ global-cloud-index-gci/index.html
[3] Bei Guan, Jingzheng Wu, Yongji Wang and Samee U. Khan (2014), ‘CIVSched: A Communication-Aware Inter-VM Scheduling Technique for Decreased Network Latency between Co-Located VMs‘, IEEE Transactions On Cloud Computing,Vol.2,No.3,pp.320-332.
[4] Chonglin Gu, Pengzhou Shi, Shuai Shi, Hejiao Huang and Xiaohua Jia (2015),‘A Tree Regression-Based Approach for VM Power Metering’, IEEE Access, Vol. 3, pp. 610-621.
[5] Dong Jiankang, Wang Hongbo, Cheng Shiduan (2015), ‘Energy-Performance Tradeoffs in IaaS Cloud with Virtual Machine Scheduling’,ChinaCommunications, Vol. 12, No. 2, pp. 155-166.
[6] Fahimeh Farahnakian, Adnan Ashraf, Tapio Pahikkala, Pasi Liljeberg, Juha Plosila, Ivan Porres, and Hannu Tenhunen (2015), ’Using Ant Colony System to Consolidate VMs for Green Cloud Computing’, IEEE Transactions On Services Computing, Vol. 8, No. 2, pp. 187-198.
[7]. Federico Larumbe and Brunilde Sanso`(2013),’A Tabu Search Algorithm for the Location of Data Centers and Software Components in Green Cloud Computing Networks’,IEEE Transactions On Cloud Computing,Vol.1,No.1,pp. 22-35.
[8] Fei Xu, Fangming Liu, Linghui Liu, Hai Jin, Bo Li and Baochun Li (2014), ‘iAware: Making Live Migration of Virtual Machines Interference-Aware in the Cloud ‘, IEEE Transactions on Computers, Vol. 63, No. 12, pp. 3012-3025.
[9] Haikun Liu and Bingsheng He (2015), ’VMbuddies: Coordinating Live Migration of Multi-Tier Applications in Cloud Environments’, IEEE Transactions on Parallel and Distributed Systems,Vol.26,No.4,pp.1192- 205.
[10] Hui Zhang, Guofei Jiang, Yoshihira k , Haifeng Chen (2014), ’Proactive Workload Management in Hybrid Cloud computing’, IEEE Transactions on Network and Service Management , Vol. 11, No. 1, pp. 90-100.
[11] Kejiang Ye, Zhaohui Wu, Chen Wang, Bing Bing Zhou, Weisheng Si, Xiaohong Jiang and Albert Y. Zomaya (2015), ’Profiling-Based Workload Consolidation and Migration in Virtualized Data Centers’, IEEE Transactions On Parallel And Distributed Systems, Vol. 26, No. 3, pp. 878-890.
[12] Lena Mashayekhy, Mahyar Movahed Nejad and Daniel Grosu (2015), ’Physical Machine Resource Management in Clouds: A Mechanism Design Approach’, IEEE Transactions On Cloud Computing,Vol.3,No.3,pp.247-260.
[13] Peter Mell, Timothy Grance, ‘The NIST Definition of Cloud Computing’, NIST Special Publication 800-145.
[14] Weijia Song, Zhen Xiao, Qi Chen, and Haipeng Luo (2014), ’Adaptive Resource Provisioning for the Cloud Using Online Bin Packing’, IEEE Transactions On Computers, Vol. 63, No. 11, pp. 2647-2660.
[15] Xiaomin Zhu, Laurence T. Yang, Huangke Chen, Ji Wang, ShuYin and Xiaocheng Liu (2014), ‘Real-Time Tasks Oriented Energy- Aware Scheduling in Virtualized Clouds’, IEEE Transactions On Cloud Computing, Vol. 2, No. 2, pp. 168-180.
[16] Yi-Ju Chiang, Yen-Chieh Ouyang and Ching- Hsien (Robert) Hsu (2015), ‘An Efficient Green Control Algorithm in Cloud Computing for Cost Optimization’,IEEE Transactions On Cloud Computing, Vol. 3, No. 2, pp. 145-155.