ISSN: 2037-4445

Study of Fundamentals of Sustainable Cloud Computing and Available Technique for Resource Allocation

Dr. Namita Chawla
Assistant Professor, M.C.A.
ASM's Institute of Business Management & Research MCA, Pune, Maharashtra, India

Dr. Nirvikar Katiyar
Director, Prabhat Engineering College, Kanpur(D),
Uttar Pradesh, India

Abstract

In order to meet the ever-increasing demand for their computational and application services, major cloud providers like Microsoft, Google, Facebook, and Amazon depend significantly on datacenters to store and process the necessary information. Yet, the financial and carbon footprint associated expenses of operating such huge infrastructure significantly undermines the sustainability of cloud services. The majority of current initiatives largely concentrate on reducing the amount of energy that is used by servers. In this article, we propose a conceptual model and practical design principles for holistic management of all resources (including servers, networks, storage, and cooling systems) in cloud data centre, with the goals of increasing energy efficiency and lowering carbon footprints (CDCs). In addition, we examine the linked connection between energy and dependability for sustainable cloud computing, where we emphasize the accompanying research difficulties. Ultimately, our goal is to make cloud computing more environmentally friendly. Finally, we offer a number of prospective research paths in the subject and build foundations for further practical improvements.

Key Terms — Green Computing, Cloud Datacenters, Sustainable Computing, Energy Efficiency, Cloud Computing.

I. INTRODUCTION

The cloud computing paradigm is characterized by the provision of a la carte, subscription-based services through the internet for the purpose of hosting programmers' and managing user workloads. In order to guarantee the accessibility and dependability of the services, the components of Cloud Data Centers (CDCs), including network devices, storage devices, and servers, need be operated continuously around the clock. Digital activities such as data streaming, file sharing, searching, and social networking websites, e-commerce, and sensor networks generate large volumes of data, which may be effectively stored and processed by employing cloud datacenters [1]. Yet, the process of producing, processing, and storing each individual piece of data adds to the overall cost of energy, increases the amount of carbon footprints left behind, and has other negative effects on the environment [2]. The community is now confronted with the difficulty of establishing a sustainable energy economy as a result of the significant amount of power used by CDCs. As can be seen in Figure 1, the quantity of energy that is being utilized by CDCs is steadily growing, and it is anticipated that by the year 2030 it will reach 8000 Tera Watt hours (TWh) [3].

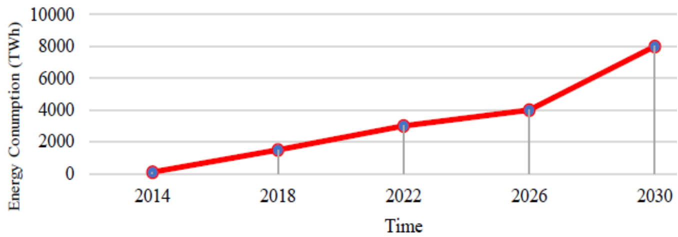

Fig. 1: The amount of energy used in cloud datacenters

Additionally, present energy-aware approaches generally concentrate on lowering the energy consumption of the servers [4]. Other aspects of CDCs, like as their networks, storage, memory, and cooling systems, are responsible for a significant percentage of the energy consumption. There is a need for an energy-aware resource management approach in order to increase the energy efficiency of CDC. This technique would be used to manage all of the resources (including servers, storage, memory, networks, and cooling systems) in a data centre.

The inefficiency of energy consumption in CDCs is caused by the underloading or overloading of infrastructure resources; in fact, the majority of the energy is consumed while some resources (such as networks, storage, memory, and processors) are in an idle state, which drives up the overall cost of cloud services [1, 2].

As things are right now, CDC service providers are looking at additional potential avenues to lessen the carbon footprints left by their own infrastructure [4]. The most important cloud service providers, including Google, Amazon, Microsoft, and IBM, are working towards the goal of employing renewable energy sources to power their datacenters [5]. Cloud computing services must be offered by CDCs of the future while minimizing their carbon footprints and the amount of heat that is released as a result of the emissions of greenhouse gases. So, what exactly are some of the concerns that are causing concern?

To keep the temperature of data centres stable, an effective cooling system is essential; nevertheless, this necessitates additional financial investment. Furthermore, by designing methods for free cooling and the use of waste heat, one may reduce the costs associated with cooling. In order to efficiently apply strategies for producing renewable energy and providing free cooling, optimal climatic conditions that take into account the location of the building are required. In addition, the places of waste heat recovery are needed to be determined before an effective implantation of waste heat recovery possibilities can take place [6]. Datacenters could need to be moved in order to take advantage of certain chances for waste heat recovery, the availability of green resources, and the closeness of free cooling resources. This would make cloud computing more environmentally friendly. In order to tackle these problems and significantly cut the energy consumption of CDCs, there is a need for cloud computing architectures that are able to provide sustainable cloud services by using an all-encompassing resource management strategy.

A. Conceptual Model

Fig. 2 illustrates a conceptual model for sustainable cloud computing in the form of layered architecture. This design provides comprehensive management of cloud computing resources, which helps to make cloud services more energy-efficient and sustainable. The following is a discussion of the four primary components of the proposed architecture:

1. ***Software as a Service***, Platform as a Service, and Infrastructure as a Service are the three separate sub-components that make up this aspect of cloud architecture.

- ***Software as a Service,*** sometimes known as "SaaS" At this level, the application manager is put into place to manage incoming user workloads. These workloads might be interactive or batch-based, and they are sent to the workload manager so that resources can be provisioned.

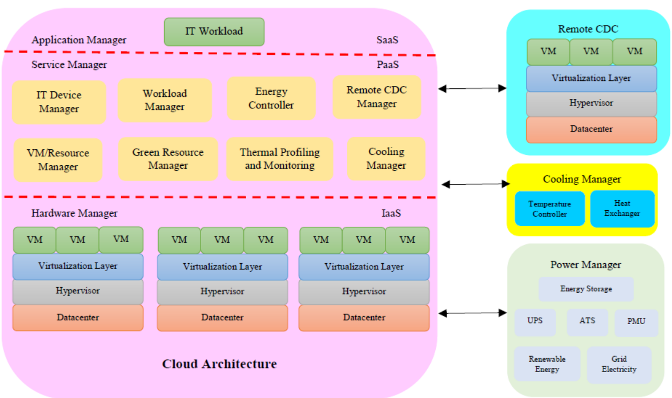

Fig. 2: A conceptual model for environmentally responsible cloud computing

- ***Platform as a Service (PaaS):*** At this tier, the controller or middleware that controls the essential components of the system is delivered. The IT device manager is responsible for managing all of the devices that are connected to the cloud datacenter. The workload manager is responsible for managing the incoming workloads that are sent by the application manager. It also determines the Quality of Service (QoS) requirement that is necessary for each workload in order to ensure the successful execution of the workload, and it transfers the QoS information of the workload to the virtual machine manager or resource manager. The use of energy in cloud datacenters is managed by energy controllers, which helps to assure the long-term viability of cloud services. The virtual machine migration as well as the workload migration between local and distant cloud datacenters is managed by the remote CDC manager for efficient usage of the available energy. VM and resource managers are responsible for provisioning and scheduling cloud resources for workload execution depending on the quality of service (QoS) needs of the workload. This may be done with either physical computers or virtual machines. Green resource manager is responsible for managing the electricity that is supplied by power management, and it gives preference to the use of renewable energy over grid electricity in order to create a cloud environment that is sustainable. The thermal profiling and monitoring approach is used to study the temperature fluctuations of the cloud datacenter based on the temperature value that is monitored by thermal sensors. This analysis is done using the thermal profiling and monitoring technique. At the infrastructure level, the cooling manager is responsible for regulating the temperature of the cloud datacenter.
- ***Infrastructure as a Service (IaaS):*** This layer includes details about cloud datacenters as well as Virtual Machines (VM). Migrations of virtual machines (VMs) are carried out in order to provide load balancing at the virtualization layer, which is necessary for the effective operation of workloads. The proactive temperature-aware scheduler is what is utilized to monitor the fluctuating temperatures of the many virtual machines (VMs) that are operating on the various cores. In order to power all of the hardware that is executing VMs, a Power Management Unit (PMU) has been integrated. The current states of virtual machines are stored in dynamic random-access memory (DRAM). Thermal sensor is used to monitor the value of temperature in order to create an alert if the temperature is higher than its threshold value and deliver the message to the heat controller for further action if the temperature is higher than its threshold value.

2. ***Cooling Manager:*** Thermal alarms will be created if the temperature is greater than the threshold value, and heat controllers will take action to manage the temperature while having as little of an effect as possible on the performance of the CDC. The energy provided by the Uninterruptible Power Supply (UPS) is put to use in order to keep the temperature under control by powering the cooling units. The use of a chiller plant, an outside air economizer, and a water economizer are included into the administration of the district heating system. This allows for accurate temperature regulation.
3. ***The Power Manager*** is responsible for managing the power that is produced from both fossil fuels and renewable energy sources (grid electricity). Renewable energy is chosen over grid electricity for use in cloud computing because it helps maintain the environment of the cloud. Grid energy may be utilized to ensure the dependability of cloud services if there is execution of workloads that are deadline focused. Solar and wind power are two examples of renewable energy sources. Batteries are used to store the energy that is generated from renewable sources. The Automatic Transfer Switch, also known as ATS, is utilized to control the energy that is coming from both sources (renewable energy and grid power), and it then transfers that energy to the UPS. In addition, the Power Distribution Unit is what moves the energy around to all of the CDCs and cooling devices throughout the facility.
4. ***Remote CDC:*** In order to efficiently balance the load, virtual machines (VMs) and workloads may be transferred to a remote CDC.

B. The Influence of Dependability on Environmental Sustainability

Cloud computing is becoming more lucrative as a result of the increasing number of clients attracted by sustainable cloud services. Increasing energy efficiency, which in turn leads to lower monthly power bills and overall operating expenses, allows more sustainable cloud computing. [13] [14]. On the other hand, in order to deliver dependable cloud services, the business operations of various cloud providers like Microsoft, Google, and Amazon are duplicating services. This requires extra resources and results in an increase in the amount of energy used. It is necessary to make a compromise between the amount of energy used and the level of dependability maintained in order to provide cloud services that are economical. The research challenges that are associated with energy and dependability for sustainable cloud computing are shown in Figure 3. In executing workloads, the energy-efficient resource management approaches that are now available use a significant amount of energy, which results in a reduction in the resources that are leased from cloud datacenters. However, response time and service delay increased as a result of the switching of resources between high scaling and low scaling modes. Dynamic Voltage and Frequency Scaling (DVFS) based energy management techniques reduced the amount of energy that was consumed; however, these techniques also reduced energy consumption. In addition, the frequency with which servers are powered on and off might reduce their overall dependability as a component of the system. The dependability of server components such as storage devices, memory, and so on is decreased when power modulation is used. We are able to increase the server's resource usage, dependability, and performance if we lower the CDCs' energy consumption.

As a result, innovative energy-aware resource management strategies are required in order to lessen the amount of power used by cloud services while maintaining their level of dependability.

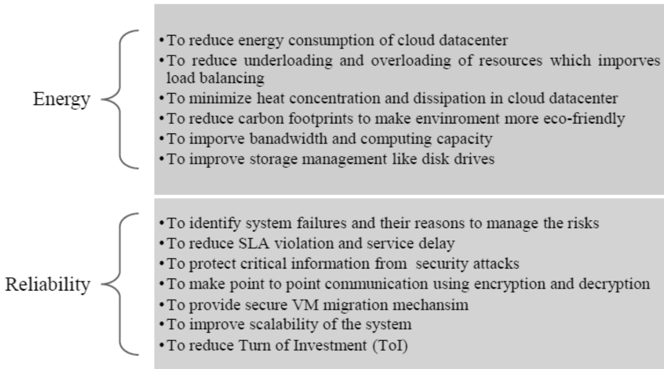

Fig. 3: Problems in Study Concerning Energy and Reliability

C. Potential Investigable Regions

The ever-increasing demand for cloud computing services that are spread over several CDCs consume considerable amounts of electricity, which in turn results in high carbon emissions and has an impact on the environment. In sustainable cloud computing, the cloud datacenters are powered by renewable energy resources, effectively reducing the amount of carbon emissions that are produced [7]. This is accomplished by replacing the conventional fossil fuel-based grid electricity or brown energy with electricity generated from solar or wind power. Cloud computing may be made more environmentally friendly by using processes that improve energy efficiency, which in turn reduces carbon footprints to a significant degree. The usage of waste heat generated by the servers and the implementation of technologies that allow for free cooling of the servers contribute to the sustainability of the CDCs [8].

Hence, sustainable cloud computing encompasses all of the following components that contribute to the creation of a sustainable datacenter [9]: I employing renewable energy as an alternative to grid electricity produced from fossil fuels; ii) making use of the waste heat created by heat-dissipating servers; iii) utilizing free cooling mechanisms; and iv) utilizing energy efficient mechanisms. Each of these aspects has a role in lowering the carbon footprints, the operating costs, and the energy consumption of the business. Concerns about environmentally responsible cloud computing have been sorted into seven areas, as indicated in Figure 4: application model; resources targeted in energy management; thermal-aware scheduling; virtualization; capacity planning; renewable energy;

and waste heat usage. Despite the fact that a few works [3-6] [8-14] have investigated problems associated with sustainable cloud computing, there are still a great deal of unanswered questions in the context of models for application composition, resources targeting for energy management, scheduling, capacity planning, and the harnessing of renewable energy and heat generated by resources.

II. RELATED WORK

[53] The ever-changing needs of cloud users make comprehensive resource management in the cloud a difficult challenge to undertake. The vast majority of currently available works focus only on the energy management of servers, disregarding other components that account for a significant portion of total energy use [9], [55]. In this part, we will discuss the resource management strategies that are already in use, as well as compare those strategies to the strategy that we have presented.

A. Scheduling of Cloud Resources Taking Energy Considerations Into Account

Energy-Aware Resource Scheduling, often known as EARS, is a method that was introduced by Li et al. [6] to run workloads inside virtual cloud environments. The EARS technique models the power and failure profiles for CDC and implements them using an event-based cloud simulator. It is effective in reducing the energy cost of a cloud data centre and improving the rate at which tasks are completed, but it only takes into consideration workloads that are identical in nature. In a similar vein, Li et al. [8] introduced a strategy called VM Scheduling (VMS) to cut down on the amount of energy used by servers. Moreover, they identified the influence that energy usage had on the number of SLA violations, which led to an increase in user satisfaction. In order to cut down on the amount of money spent on execution and the amount of energy used, Balis et al. [7] suggested a Holistic Approach (HA) for the management of IT infrastructure. Holistic Workload Scaling (HWS) is a technique that was proposed by Perez et al. [10] to enable scaling of resources vertically as well as horizontally simultaneously. This technique also helps to reduce latency by using a multi-scaling approach, and it does all of this without taking into account the energy consumption of cloud resources. Luo et al. [11] formulates energy consumption as a task-core assignment and scheduling problem. They then proposed a Holistic Energy Optimization Framework (HEOF) to reduce thermal effect as well as cooling cost simultaneously. The HEOF framework places an emphasis on powerful computation capability. In order to increase the energy efficiency of various pieces of hardware, such as memory, storage, and CPUs, Liu and colleagues [12] suggested a Server-based Cloud-enabled Architecture (SCA). A case study of a video tracking application is used to analyze the performance of SCA, and the experimental findings demonstrate that SCA performs better in terms of memory use. Guzek et al. [13] suggested a Holistic Model for Resource Management (HMRM) in virtualization-based cloud datacenters in order to cut down on the amount of energy used by various resources such memory, storage, and networking. Holistic Approach for Cloud Service Provisioning (HACSP) was suggested by Ferrer et al. [14] to satisfy planned and unforeseen changes in resource needs dynamically and optimizes energy cost. It was hypothesized by Feller et al. [15] that a Holistic approach for Energy Management (HEM) strategy may be used for the efficient management of virtual cloud resources by making use of dynamic web workloads, all while conserving a significant amount of energy. Dinkar et al. [60] suggested an Energy Efficient Data Center Management (EEDCM) approach under availability restrictions. In their paper, the authors highlight the significance of availability and construct a hill climbing algorithm in order to avoid failure zones. The experimental result demonstrates that the EEDCM approach lowers the amount of energy that is used by the datacenter; nevertheless, the trade-off between the amount of energy consumed and other significant QoS metrics, such as dependability, cost, and execution time, is not taken into account.

B. Scheduling Taking Reliability Into Account

Cloud Service Reliability Enhancement (CSRE) is the name of the method that Zhou and colleagues [16] proposed for making use of available network and storage resources. This method saves the current state of the VM by using service checkpointing while it is in an active state.

In addition, the node failure predicator is intended to make the most efficient use of the resources provided by the network. A convergent dispersion based multi-cloud storage (CDStore) system was created by Li et al. [17] in order to provide a cloud service that is dependable, secure, and efficient in terms of cost. A deterministic-based deduplication strategy is provided by the solution that has been offered in order to conserve both network traffic and storage space. In addition to that, CDStore employs a two-stage deduplication process in order to defend the system from malicious assaults.

An technique known as Multi-Objective Resource Scheduling (MORS) was introduced by Azimzadeh et al. [18] in order to improve the dependability of cloud services and make the most efficient use of execution time. In addition, the authors identify a trade-off between reliability and execution time for the purpose of efficient execution of workloads associated with HPC (High Performance Computing). An Adaptive and Just-In-Time (AJIT) scheduling technique was offered by Poola et al. [19], who suggested employing spot and on-demand instances as a means of providing a fault management mechanism. This strategy utilizes resource consolidation to optimize execution time and cost, and the results of experimental testing suggest that AJIT is effective for the execution of workloads that are deadline-oriented. A Heterogeneous Spot Instances-based Auto-scaling (HSIA) solution was presented by Qu et al. [20] to run web applications in a dependable management. In addition, the HSIA method created a fault-tolerant system in order to increase the availability of cloud services while simultaneously lowering their execution costs and improving their response times. During the execution of real-world and synthetic streaming applications, the replication-based state management (E-Storm) solution that was introduced by Liu et al. [21] actively maintains numerous state backups on various VMs. E-performance Storm's is compared to that of the checkpointing approach, and the results of the experiments show that E-Storm produces effective outcomes in terms of latency and throughput. A Dynamic Clustering League Championship Method (DCLCA) was presented by Abdulhamid et al. [22] to reduce fault reduction in task failure during resource allocation for workload execution. A Resource Optimization approach for Cloud Data Center (ROCDC) was suggested by Liang et al. [61]. This method builds a conceptual model to maximize the performance parameters reliability and energy while scheduling resources. Nevertheless, no simulation nor experiments were used to verify the efficacy of this method.

C. Resource Scheduling Method Based on the Cuckoo Optimization Algorithm

A Group Technology-based model and Cuckoo Optimization (GTCO) technique was suggested by Shahdi-Pashaki et al. [23] to allocate resources for the effective mapping of virtual machines (VMs) to cloud workloads. GTCO is superior than round robin based resource scheduling in terms of both performance and cost savings in regards to energy consumption when VMs are being executed. A Cuckoo Optimization based Task Scheduling (COTC) approach was presented by Sundarrajan and Vasudevan [24] to schedule the tasks in cloud computing. This algorithm optimizes the energy usage for the execution of homogenous workloads. Abbasi and Mohri [25] presented a technique for task scheduling called Cuckoo Optimization based Resource Management (CORM), which improves load balancing to decrease energy cost. CORM is superior than round robin based resource scheduling in terms of performance and increases energy efficiency when cloud resources are being executed. In their paper [26], Navimipour and Milani introduced a method called Cuckoo Search Algorithm based Task Scheduling (CSATS), which is designed to make efficient use of cloud resources. The authors have just assessed the fitness value (execution time) of CSATS with various values of probability in order to discover the cloud resource that is best suited for the execution of workloads. Cuckoo Search Meta-Heuristic (CSMH) is an algorithm that was presented by Madni et al. [27], and its purpose is to optimize the amount of energy that cloud workloads use. Both COTC [24] and CSMH [27] have had their performance analyzed by CloudSim [30], and the results showed that they both cut the energy cost of servers without concentrating on any other aspects of the cloud datacenter.

D. Evaluation of CRUZE in Relation to Previously Used Resource Scheduling

The current methods of resource scheduling are outlined in further detail in Table 1, which contrasts them with our suggested method (CRUZE).

According to our findings, the currently available strategies for holistic resource management only take into account one or two components concurrently. The vast majority of the currently available work schedules resources for the execution of homogeneous workloads, although other types of work, such as EARS [6], HEOF [11], and AJIT [19], schedule resources for the execution of heterogeneous workloads as well. None of the currently available works takes into account the clustering of workloads for the purpose of resource allocation. Only in CSRE [16] is the provisioning-based scheduling of resources taken into consideration. Cuckoo Optimization (CO) based scheduling is accomplished in GTCO [23], COTC [24], CORM [25], CSATS [26] and CSMH [27], however, scheduled resources are only for homogeneous cloud workloads and there is no provisioning of resources. The energy cost and latency are both optimized by GTCO [23], while COTC [24], CORM [25], CSATS [26], and CSMH [27] only optimize the energy cost and the execution time. This is the first research study that takes an all-encompassing approach to the management of basic CDC components with the goal of delivering cloud services that are dependable and environmentally friendly. Scheduling the allocated resources for the execution of clustered and heterogeneous workloads is the goal of the suggested method, which is referred to as CRUZE. This helps to allow dependable and sustainable cloud services.

III. APPLICATION MODEL

When it comes to environmentally responsible cloud computing, the application model is of the utmost importance, and having an application that is well-structured and efficient may help increase the energy efficiency of cloud datacenters. There are three possible models for application architectures: data parallel, function parallel, and message passing. [1] [2]. The data parallel model is a kind of parallelization that may take place across numerous processors in parallel computing environments. This type of parallelization places an emphasis on spreading the data among various nodes so that it can be worked on simultaneously by several processors. Map-Reduce and parameter sweep models are two examples of data parallel models. Another example is the bag of tasks model. The function parallel model is a type of parallelization of computer code across multiple processors in parallel computing environments. This type of parallelization places an emphasis on the concurrent distribution of tasks, which are carried out by processes or threads that are distributed across various processors. Both the thread model and the task model are examples of data parallel models. It facilitated the creation of portable and scalable large-scale parallel applications by providing a communication feature between a collection of processes that are mapped to nodes or servers in a manner that is independent of the language being used.

A. The Resources That Will Be Focused on While Managing Energy

There are a variety of potential options available to enhance the energy efficiency of CDCs. It is stated that the energy usage of cloud datacenters for the CPU, memory, storage, network, and cooling is 45%, 15%, 10%, and 10% accordingly [1] [4] [10]. The CPU is the primary source of energy consumption, followed by the cooling management system. The power regulation techniques cause an increase in the amount of energy that is used when the task is being executed, which in turn impacts the resource usage of CDCs. In addition, DVFS was successful in resolving the issue of resource usage; however, the fact that it may transition between high and low scaling modes for resources results in an increase in response time and service delay, which is a violation of the SLA. Placing servers into a "sleeping mode" or manually turning them on and off might have an impact on the components of the system's dependability, such as storage. The goal of enhancing the energy efficiency of cloud datacenters has an impact on the usage of server resources, as well as the dependability and performance of such servers.

The solution of bin packing has been used in energy-conscious resource management approaches that are now in use in order to distribute resources for the completion of workloads. There are two issues that can arise with resource allocation: I under-utilization of resources (which occurs when resources are reserved in advance, but resource requirement is lower than resource availability, which increases cost), and ii) over-utilization of resources (which occurs when resource requirement is higher than resource availability, which also increases cost) (a large number of workloads are waiting for execution due to unavailability of sufficient amount of resources). Using the stall time is going to provide the greatest results out of all the potential strategies for reducing the high voltage supply in order to reduce energy consumption. Several strategies have been put forth.

B. Scheduling With Consideration of Temperature

Architecture and scheduling techniques are two key aspects of thermally aware scheduling that need to be considered. Both single-core and multi-core architectures are viable options, and the scheduling technique may either be reactive or proactive. The inefficiency of cloud datacenters is increased when there is a heating issue during the execution of workloads. Thermally aware scheduling seeks to decrease the cooling setpoint temperature, hotspots, and thermal gradient in CDCs in order to alleviate the issue of excessive heating in these types of facilities [5]. In comparison to heat modelling, thermally aware scheduling results in significant cost savings and improved performance. The

By activating servers that are near to each other in a rack or chassis, it is possible to limit the energy consumption of CDCs; nevertheless, this results in a rise in power density, which in turn causes a concentration of heat. In order to find a solution to this issue, a cooling system is necessary. There is a need for efficient thermally aware scheduling approaches that are able to carry out tasks with a minimal amount of heat concentration and dissipation [6]. This not only saves money on power but also makes it easier for the cooling mechanism to do its job. Because of the fluctuating temperatures of the servers in CDC, the complexity of scheduling and monitoring is enhanced. This volatility in temperatures is also the reason of ambiguity in thermal profiling. If we are going to be successful in resolving this issue, we will need dynamically updated thermal profiles as opposed to static ones. These thermal profiles will be automatically updated and will produce temperature readings that are more accurate. Current thermally aware strategies that are centred on lowering Power Use Efficiency (PUE) may be discovered; however, lowering PUE alone might not be enough to bring down the Total Cost of Ownership (TCO).

C. Virtualization

Migration of virtual machines is necessary while workloads are being executed in order to efficiently balance the load and make use of renewable energy resources in decentralized cloud data centres (CDCs). Virtual machine approaches shift the workloads to other computers that are geographically dispersed since there is no renewable energy available on the premises. VM technology also enables the movement of workloads from cloud datacenters that are powered by renewable energy sources to cloud datacenters that are powered by waste heat located at another location. VM-based workload migration and consolidation solutions give virtual resources utilizing just a small number of physical servers. This helps to strike a balance between the workload demand and renewable energy. It is possible to relocate storage from one operating server to another in order to improve the speed of virtualization without having an impact on the way in which VMs carry out their workloads [7]. The exploitation of waste heat and other alternative forms of renewable energy resources may be tapped into via the employment of VM migration methods to allow sustainable cloud computing. Improving energy usage and reducing network latency during the transfer of workloads across geographically distant resources presents a significant challenge for virtual machine (VM) migration solutions. Raising the size of a VM causes it to use more energy, which in turn might cause a delay in the provided service. When migrating virtual machines across a wide area network (WAN), point-to-point communication is essential for finding a solution to this issue [8].

D. Preparing for Available Resources

In order to achieve a satisfactory return on investment, cloud service providers are required to engage in efficient and well-organized capacity planning (ROI). Planning may be done for the capacity of the electrical infrastructure, the information technology resources, and the workloads. SLA should establish service quality requirements to assure backup and recovery, storage, and availability. This will boost user happiness, which will in turn attract more consumers in the future [9]. It is necessary to take into account crucial usage factors for each application in order to optimize the consumption of resources via virtualization. This may be accomplished by locating apps that can be combined into a single one. The consolidation of applications results in increased resource utilization and decreased costs associated with capacity. Cloud workloads should be analyzed before their execution in order to ensure that capacity planning is carried out in an effective manner. This is especially important for time-sensitive workloads. In order to effectively manage power infrastructure, virtual machine migration should be made available for the migration of workloads or machines. This enables cloud datacenters to improve their energy efficiency by successfully completing the execution of workloads while using a minimum amount of resources. There is a pressing need for efficient capacity planning in order to store data and process it in a way that is both effective and cost-efficient.

E. Renewable Energy

Some aspects of renewable energy that are amenable to improvement include the energy source (such as the sun or the wind), the energy storage technology (such as net metering or batteries), and the location (such as off-site or on-site). Unpredictability and the high cost of capitalizing green resources are two of the primary obstacles faced by renewable energy. The problem of unpredictability in the supply of renewable energy was addressed by using strategies such as energy-aware load balancing and workload relocation [10]. The majority of commercial cloud datacenters are situated in areas that are far far from plentiful sources of renewable energy. As a result, movable cloud datacenters are needed to be placed closer to renewable energy sources in order to make the most of their economic potential [13]. In addition, the Carbon Use Efficiency (CUE) may be improved by increasing the amount of resources for renewable energy. The significant initial investment required for renewable energy adoption in cloud datacenters is a hurdle for research.

F. The Use of Waste Heat

When it comes to making efficient use of waste heat, the cooling mechanism and heat transfer model play a significant role. CDCs are producing heat because they use a significant quantity of energy, which causes them to behave as a generator. CDCs may employ waste heat in their cooling systems that are based on vapour absorption; the system then makes use of the heat while it is evaporating [11], [12]. By offsetting the cost of cooling, vapor-absorption-based free cooling solutions have the potential to bring the PUE value to its optimal level. The heat produced by CDC may be used to warm up buildings in regions with low average temperatures. The use of stacked and multi-core server designs is increasing the power density of servers, which in turn further raises the expenses associated with cooling the servers. It is possible to enhance the energy efficiency of CDCs by lowering the amount of energy that is used for cooling. There is a pressing need to relocate cloud datacenters in order to cut down on the expenses associated with cooling them, and this may be accomplished by positioning the CDCs in regions that have access to free cooling resources.

IV. CONCLUSION

We noticed the need of cloud computing environments as well as the challenges that come with maintaining them. We put up a conceptual model for the comprehensive management of resources with the goal of reducing the carbon footprints of cloud datacenters. This results in cloud services that are more energy-efficient and sustainable. Integrating power infrastructure and cooling devices into an energy-aware resource management approach that is used with information technology equipment is one way that holistic management may increase the energy efficiency of these two categories of devices.

REFERENCES

[1]. Zhenhua Liu, Yuan Chen, Cullen Bash, Adam Wierman, Daniel Gmach, Zhikui Wang, Manish Marwah, and Chris Hyser. “Renewable and cooling aware workload management for sustainable datacenters.” ACM SIGMETRICS Performance Evaluation Review, vol. 40, no. 1, pp. 175-186, 2012.

[2]. Hui Dou, Yong Qi, Wei Wei, and Houbing Song. “Carbon-Aware Electricity Cost Minimization for Sustainable Datacenters.” IEEE Transactions on Sustainable Computing, vol. 2, no. 2, pp. 211-223, 2017.

[3]. Anders SG Andrae, and Tomas Edler. “On global electricity usage of communication technology: trends to 2030.” Challenges, vol. 6, no. 1, pp. 117-157, 2015.

[4]. Abbas Mardani, Ahmad Jusoh, Edmundas Kazimieras Zavadskas, Fausto Cavallaro, and Zainab Khalifah. “Sustainable and renewable energy: An overview of the application of multiple criteria decision making techniques and approaches.” Sustainability, vol. 7, no. 10, pp.13947-13984, 2015.

[5]. Sukhpal Singh, Inderveer Chana and Rajkumar Buyya. “STAR: SLA-aware autonomic management of cloud resources.”

[6]. IEEE Transactions on Cloud Computing (2017).

[7]. Xiang Li, Xiaohong Jiang, Peter Garraghan, and Zhaohui Wu. “Holistic energy and failure aware workload scheduling in Cloud datacenters” Future Generation Computer Systems, vol. 78, pp. 887-900, 2018.

[8]. Peter Garraghan, Ismael Solis Moreno, Paul Townend, and JieXu. “An analysis of failure-related energy waste in a large- scale cloud environment”, IEEE Transactions on Emerging topics in Computing, vol. 2, no. 2, pp. 166-180, 2014.

[9]. Altino M. Sampaio and Jorge G. Barbosa. “Energy-Efficient and SLA-Based Resource Management in Cloud Data Centers.” Advances in Computers, vol. 100, pp. 103-159, 2016.

[10]. Xiang Li, Peter Garraghan, Xiaohong Jiang, Zhaohui Wu, and JieXu. “Holistic virtual machine scheduling in cloud datacenters towards minimizing total energy.” IEEE Transactions on Parallel and Distributed Systems (2017).

[11]. Cinzia Cappiello, PacoMelia, Barbara Pernici, Pierluigi Plebani, and Monica Vitali. “Sustainable choices for cloud applications: a focus on CO2 emissions.” In 2nd International Conference on ICT for Sustainability, pp. 352-358, 2014.

[12]. Mueen Uddin and Azizah Abdul Rahman. “Energy efficiency and low carbon enabler green IT framework for datacenters considering green metrics.” Renewable and Sustainable Energy Reviews, vol. 16, no. 6, pp. 4078-4094, 2012.

[13]. Tianyi Chen, Yu Zhang, Xin Wang, and Georgios B. Giannakis. “Robust workload and energy management for sustainable datacenters.” IEEE Journal on Selected Areas in Communications, vol. 34, no. 3, pp. 651-664, 2016.

[14]. Sukhpal Singh Gill and Rajkumar Buyya. "A Taxonomy and Future Directions for Sustainable Cloud Computing: 360 Degree View." arXiv preprint arXiv:1712.02899 (2017).

[15]. Yogesh Sharma, Bahman Javadi, Weisheng Si, and Daniel Sun. “Reliability and energy efficiency in cloud computing systems: Survey and taxonomy.” Journal of Network and Computer Applications, vol. 74, pp. 66-85, 2016.