International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 5 Volume No.6 Aug-Sept – 2022
DOI:10.46647/ijetms.2022.v06i05.116 ISSN: 2581-4621

Improving Energy Efficiency in Cloud Computing Data centres Using Intelligent Mobile Agents

Charan C S
PG- Computer Science ,Dr.N G P Arts And Science, Coimbatore, Tamilnadu

ABSTRACT
Cloud computing has become a profound name and also a solid bedrock for new emerging technology. Cloud technology fully supports the dynamic provisioning of computing resources as a utility service on a pay-as-you-go approach. Its numerous benefits (such as rapid elasticity, flexibility, network resource pooling) empower small, medium, and large enterprises to use other technologies through the Internet's flexibility. Despite the benefits cloud technology offers, the challenge of the high power consumption rate it incurs as it leverages its promised attributes remains a major concern. Academic research has shown that a typical 500 square meter data centre consumes about 27,048 kilowatts per hour of power per day regardless of whether it is active or not. Over the years, the most dominant energy efficient techniques for managing data centre has been Dynamic voltage frequency scaling and virtual machine consolidation, which has had a significant setback due to its inability to manage systems on overload state. Therefore, a novel paradigm based on an intelligent mobile agent approach has been proposed. This proposed approach is highly intelligent can easily detect underutilised and overloaded components of the data centre due to its unique feature. Agent technique has successfully shown it can prevent and manage overloading issues due to change in workloads and achieve a more efficient load balancing with a low power consumption rate. The mobile agent was embedded into servers and switches to regulate their activities and then shut down underutilised components. Mobile agent (Java agent) is the first of its kind used in a cloud environment. This research proposal saves a significant amount of energy and improves the entire system performance. In this thesis, the intelligent-based Agent approach is used to address energy efficiency and cost-aware related problems. Agent approach helps facilitate resource management, allocation of cloud data centre components with a significant reduction in energy usage rate with a more efficient system performance while maintaining a highly reliable system as promised by service providers.
Keywords— Cloud Computing

1. Introduction

Cloud computing has become the new norm for interconnection and communication in the Information Technology (IT) world. Cloud technology's recent growth can be attributed to the massive increase in Internet-based activities and Internet users. Cloud computing has firmly established that it is the most cost optimisation IT innovation for enterprise use. It leverages computing resources to the small, medium and struggling businesses, giving them the tremendous opportunity to strive and compete fervently with other big enterprises. It is a technology that aims for development with little or no constraints due to its flexibility of usage via virtualisation and service-oriented software. Cloud Technology brought about three dominant differing approaches to the way computing resources are used with zero-level stress and the upfront cost of purchasing infrastructure, buildings, and IT supply management. These approaches are: - pay-as-you-go, elasticity and on-demand provisioning. Cloud computing innovation provides IT resources in the form of utility such as our daily use of water and gas.

Cloud computing gives access to a great wealth of computing elements, storage, networks and software from a personal or public local server to required specific client hosted by Facebook, Amazon Web Service (AWS), Microsoft Azure, Google and others. Cloud innovation has now gained significant attention across the industry, business, research and academic sections due to its performance attributes (usage flexibility, rapid resource pooling, the ubiquity of the network, etc.) and its fastest-growing segment IT spending.

DOI:10.46647/ijetms.2022.v06i05.116 ISSN: 2581-4621

Cloud computing can be defined as delivering computing services through the Internet (simply referred to as the cloud). However, there is a standard definition of cloud computing based on the National Institute of Standards and Technology (NIST) standards which says that "cloud computing is a model that enables ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort by the service provider" (Mell & Grance). Theseservices include software, database, servers, storage, networks, etc. and can be deployed through any of these three cloud service levels models to users as- (i) Infrastructure-as-a-Service (IaaS), which provides the computing, networking and storage resources (ii) Platform-as-a-Service (PaaS) which provides the valuable tool that facilitates clouds application deployment, and (iii) Software-as-a-Service (SaaS) which gives the user access to provider's available software on cloud networks. These different service levels depict the summary of all the different services provided by the cloud. IaaS is the most common cloud computing services. With IaaS, you can rent necessary IT infrastructure (e.g. servers, virtual machines, networks, operating systems, etc.) on a pay-as-you-go basis. PaaS provides an on-demand platform for developing, testing, and managing software. This makes it easy for developers to do their job without worrying about the underlying infrastructure, which can cost a fortune to achieve. Software-as-a-service is a type of service that allows cloud customers to use any type of software provided by their chosen cloud providers. Thus, the cloud customer does not need to worry about any subscription, maintenance, security etc., apart from paying for the servicesthey used. Various cloud platforms have been built with substantial infrastructural provisions to support vast applications worldwide, assuring resources are scalable, reliable and available to clients when needed. This gives the cloud users the choice to use computing resources based on their business requirements and when the need arises.

Cloud computing has established a robust network communication flow classified into two main parts; the cloud-to-user flow and the intra-cloud flow. Compare this traffic flow to the era of fewer cloud activities when legacy local computers were the norm that holds all the data and software for the user. Now, correspondences are received on request from the cloud data centre. The volume of data processing per second globally is getting so massive, which corresponds to Cisco's whitepaper report that network communication traffic flow is the fastest-growing data centre component, rising to 4.3 zebibytes (ZiB) in 2016 with a combined annual growth rate of 44% (Cisco, 2013). Therefore, the data centre is the backbone of cloud activities because cloud service providers use data centres to deploy their services in different geographical locations. However, it is worth noting that cloud computing is not the same as a data centre. While cloud stores data on the Internet, data centres do the same but within the enterprise's local network. A typical data centre is comprised of a plethora of smaller networks that contain processing servers and storage devices. These networks are interconnected through many network switches to form an overarching data centre (DC) architecture. The design of DC architecture must provide a means to being efficient, dynamic, resilient and function 24/7, which it sometimes fails to achieve.

This brings me to explaining the component contentment of DC. A typical data centre consists of the following components: core router, processing servers, layered network switches, and Virtual Machines (VM), each strategically positioned and connected to meet clients' needs based on the DC requirements. Currently, the cloud data centre contains tens of thousands of servers connected to the storage and then service the entire globe. These connection processes put massive pressure on the data centre networks to maintain a certain low cost, high bandwidth and low power usage level. To ensure that the cloud data centre system operates with a minimal service cost without violating the promised system reliability, service providers acquire specialised server boards and networking equipment built by Original Design Manufacturer (ODM) with specific workload requirements. For instance, Facebook now uses Open Compute Project (OCP) with better optimises server while Microsoft uses Performance Optimised Data Centre modules (PODs). Figure1.1 below shows how cloud platforms accommodate many computing components, applications, and deliveries to many customers and enterprises worldwide.

2. Results and Discussion

2.1 Introduction

This chapter discusses the results of the experiment during this research work. In order to compare the effectiveness and the performance of the proposed model used in this research work to other existing approaches, the energy consumption level of using agent technology on data centre switches, virtual machine, servers and the other non-critical components were captured, measured and evaluated.

In this chapter, the experiment was conducted on various scenarios using the agent code to ascertain the effectiveness of integrating agent attributes to different components of the data centres. Therefore, this research work considers the shutting down of switches and servers intelligently, underutilised or over utilised based on a set threshold defined in the java agent code. Secondly, the agent code monitors the virtual machine activities, hence migrating VMs based on the same agent technique but focusing on the VM capacity and limitations attached to each VM and its associated host. Then the system performance was examined based on the QoS's reliability during the cause of using this approach on a cloud data centre by evaluating the behavioural impact of agent code on the system as it communicates with the entire network. Finally, the analysis of the result obtained in this experiment with other existing valid findings is discussed.

2.2 Virtual Machine Migration Result Discussion

From the observations based on the conducted research, the mobile agent encapsulates the virtual machine's behaviour, then acts on it through the migration of VMs. A competitive analysis of an agent-based migration of VM in an intelligent way to ensure dynamic VM consolidation problem got a solvable solution was conducted and then compared to other existing approaches. From the obtained results, it is evident that migrating the VM instances at the approximate time to avoid underutilization and overutilization of virtual machines on top of the physical servers through a hypervisor's aid improved the system performance metric.

This also aids in shutting down network components that were supplying electrical current to the instances, which saves a significant amount of power during the processing time compared with other research works. Figure 4.1 shows that at each traffic flow, which is from the time the cloudlets were sent to the processing queue to the time it was returned to the client, as a complete executed task. The virtual instances receive and allocate cloudlets based on the set capacity. The agent system monitors the transitional flow and decides which VM to migrate based on the defined thresholds and constraints. This intelligent method has access to the entire system and with very high-speed performance level. This resulted in the system's increased agility and the amount of energy saved at each virtual machine's migration period using our algorithm was significant. At each phase of the traffic metric, a substantial amount is saved. As shown in Figure 4.1, the system has maximum energy savings of 90% and a minimum saving of 49%. This shows a very high chance of using mobile agent technology to break the deadlock on the cloud system complexity challenges. Despite the savings observed at each stage, there is minimal disruption to the system's performance due to the agent's attribute to flexibly and autonomous movement around the network without interrupting the system distribution process or traffic flow.

4.3 Server and switch Performance Evaluation

In this section, the mobile agent's results of various simulation outputs were evaluated for both time- and daemon- base. The observed simulation results are based on five set goals, which are to observe the agent's behaviour on the components of the switches and servers, test the effectiveness of implementing agent technique on the cloud system with the proposed mobile agent called Java agent, check the number of energy-savings from the agent technique at each triggered scenario, compare the performance of agent approach to other existing methods to validate its performance strength and finally, to observe the impact of the agent on the entire system when it is deployed based on the system reliability test metric.

In this phase of the simulation results, the time-based, demon-based and no agent strategy were deployed simultaneously with various hosts, VMs and cloudlets. Table 4.2 shows the different data

DOI:10.46647/ijetms.2022.v06i05.116 ISSN: 2581-4621

centres' configuration used for the number of hosts, VMs and Cloudlets, and the total power consumption used by each agent type.

3. CONCLUSION

3.1 Introduction

This chapter summarises research contributions on agent-based approaches for an energy-efficient data centre. In the section, we will summarise the research outcome and the research working experience trying to extend the existing practices of managing cloud data centre energy-related challenges. We encountered some obstacles during this work that will now be the future direction for upcoming research works in this field.

6.2 Summary and Contributions

The irresistible benefits of using cloud technology can be in high jeopardy if the challenge of the energy consumption rate it faces continues without a workable permanent solution. The cloud vendors such as Amazon, Google, Salesforce, IBM, Microsoft has always used data centre as the backbone for their cloud activities and applications globally. Vendors always aim to satisfy customers' needs by ensuring they receive a flexible, accurate, suitable and timely service on request. Unfortunately, the challenge of high energy consumption rate threats the vendors promise to the customers. The data centre can consume high energy level as much as the power consumed by 25,000 households in the US, according to (Darathna et al.,2014).

Compared to developing countries with significantly lower energy generation level, it is better not to imagine the demand the high energy consumption can cause in such areas. Furthermore, the rapid growth of the use of the Internet, smart device, the need to use different complex data-driven applications, and sudden switch to the remote working environment caused by the pandemic has increased the data centre's activities heightens its energy consumption level. Therefore, it is very expedient that cloud vendors find a soluble pattern to minimise the cloud data centres' energy consumption rate to improve its efficiency, system reliability and finally reduce the operational cost of running the data centre.

Original research has suggested VM consolidation and Dynamic Voltage Frequency Scaling (DVFS) to solve this energy consumption challenge. However, the solution provided by this approach was minimal and lacks efficiency, especially when there is a sudden surge on the system with a high level of system overload.

From Literature, we have stated in previous chapters that inefficient resource utilisation causes a major cause of cloud data centre high energy consumption rate by application on server and switches. To utilise the data centre infrastructure resources efficiently, data centre components should be observed to have an explicit understanding of its operational capacity, encapsulate its logic, and monitor its interface functionality to ensure flexible manipulation and migration of components based on its performance rate.

This research work is the first research documented work to have used the mobile agent approach in the cloud computing-related field and also to use its features to leverage energy efficiency-related challenges in the cloud data centre. As we have previously pointed out, the agent technology was used in 2012 for the wired and wireless network during traffic regulation and sensor monitoring. The agent-based technique was used to shut down inactive switched and server and migrate VMs during their underutilised phase. In this thesis, we have conducted a rigorous literature search through a survey, proposed the agent-based approach to fill the literature gap, implemented the prototype of our finding through simulation, and monitored the activities of the data centre under agent controller traditional design settings.

Chapter 1 gave a background understanding of cloud computing, its enabling technologies, its challenges and why an intelligent agent system should be welcomed. We then stated the objective and the contribution of our work to the literature and research world. Chapter 2 was mainly a literature review on existing literature based on energy efficiency techniques, approach and the ways it was implemented. Different taxonomy based on this research work was produced to prove the significance

of the research which was classified as followed (a)cloud data centre design, (b) workload scheduling (c) VM migration and system maintenance and (d) monitoring - discussing their advantages and disadvantages.
Chapter 3 discussed the methodology adopted during the process of investigating the impact of using an agent-based approach on cloud data centre network.

References

1. A. Bieszczad, B. Pagurek, T. White, "Mobile Agent for Network Management" IEEE communication Surveys,1, 2-9, 1998.
2. A. M. Al-Qawasmeh, S. Pasricha, A. A. Maciejewski, and H. J. Siegel, "Power and thermal-aware workload allocation in heterogeneous data centres," IEEE Transactions on Computers, vol. 64, no. 2, pp. 477–491, 2015.
3. A. Verma. Ahuja, and A. Neogi, "Mapper: Power and migration cost aware application placement in virtualized systems", in Proceedings of the 9th ACM/IFIP/USENIX International Conference on Middleware. Springer-Verlag New York, Inc., 2008, pp. 243–264.