International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 12
August 2016

Balancing Load and Scaling Application Performance with Minimized Energy Consumption in cloud computing

[1]SANKARA BABU VEMULAPALLI, [2] Mrs.N.SUJATHA
[1] M.Tech (CSE) from JAGRUTI INSTITUTE OF ENGINEERING AND TECHNOLOGY
[2]Associate Professor, Department of Computer Science and Engineering, JAGRUTI INSTITUTE OF ENGINEERING AND TECHNOLOGY, Telangana State, India.

ABSTRACT: Cloud computing offers low price and services using cloud any user can access cloud services as pay per use. The rapidly growing rate of the usage of big-scale data centers on cloud has demand for computational power. In order to maintain applications in cloud the cloud has to maintain resource for applications. Therefore minimization of energy consumption and balance the resources are issues in cloud computing. Load balancing in cloud computing Environment is the primary consideration. Its main aim is to distribute workloads across various computing resources and optimize the usage of resources, increase efficiency. Load balancing provides use satisfaction and also the ratio of resource utilization after ensuring the allocation and efficiency of every resource being computed. This paper presents a survey of load balancing mechanism in order to provide the efficient and optimized utilization of resources and overall cost minimization.

Keyword: - Cloud computing, Load balancing, Consolidation, Energy-aware scheduling, Energy proportional systems, scheduling, energy efficiency.

INTRODUCTION:

Cloud Computing can be simply defined as computing in remote location or location independent with shared and dynamic resource availability on demand. The paramount motive behind more organizations moving to cloud is the mitigation in cost and dynamic provisioning of resources. It is a model for broad network access for enabling ubiquitous, convenient approach to a shared pool of computing resources. Cloud computing is an attractive computing model since it allows for the provision of resources on-demand. In the cloud computing domain, the allocation and

reallocation of resources dynamically is the prime focus for accommodating unpredictable demands and, eventually, contribute to high return on investment. Hence, Cloud Computing is making our business application more mobile and collaborative. The consumption of energy associated with the resources allocation should be taken into account. Resource allocation is the key technology of cloud computing domain, which utilizes the computing resources like bandwidth, energy, delay and so on in the network to facilitate the execution of cumbersome tasks that require large-scale computation. A Resource Allocation Strategy (RAS) in Cloud Computing can be understood as any mechanism that aims to assure the application's requirements are attended to precisely by the provider's infrastructure. Cloud providers offer these computing resources as measured services for their clients in a pay-as-you-go fashion. Cloud clients, also called as tenants, request the amount of resources needed to perform certain jobs, to the cloud providers. Upon receiving a client or tenant request, the cloud provider, with the help of virtualization, creates several virtual machine (VMs) on a physical machine (PM) or server and allocates the requested resources to it and thereby reduces the amount of hardware and execution time. The objective of this paper is to focus on various existing resource allocation techniques in cloud computing environment and thereby providing a comparative study.

2. Literature Survey

Shunmei Meng, Wanchun Dou, Xuyun Zhang, Jinjun Chen[1] Live Migration of Virtual Machines:- Migrating operating system instances across distinct physical hosts is a useful tool for administrators of data centers and clusters: It allows a clean separation between hardware and software, and facilitates fault management, load balancing, and low-level system maintenance. By carrying out the majority of migration while OSes continue to run, we achieve impressive performance with minimal service downtimes; we demonstrate the migration of entire OS instances on a commodity cluster, recording service downtimes as low as 60ms. We show that that our performance is sufficient to make live migration a practical tool even for servers running interactive loads.

J. Baliga, R.W.A. Ayre, K. Hinton, and R.S. Tucker. [2]. The availability of the high-speed internet network and IP connections is provide the delivery of the latest network based services. So the network based computing becomes more widespread and rapidly expanding the energy consumption of the network and the network resources are also rapidly grown .This is done when there is increasing more attention to

manage energy consumption across the information and communication technology sectors. while energy uses by data centers having received much attention, but there has been less attention given to the energy consumption of the switching the networks and transmission for connecting users to the cloud. This paper describe the analysis of energy consumption in cloud computing that consider the public and private classes in that we can include the energy consumption in transmission and switching as well as data storage and data processing. This paper also describes the energy consumption is transport and switching having adequate percentage of total energy consumption in the cloud computing. A. Beloglazov,

R. Buyya [3]. This paper describes that in business, scientific and different –applications requires the large computation power so the rapidly increasing the demand of such a resources to the electrical power required by the large scale data centers also increases. This paper also defines for reducing operational costs and also provide quality of services (Qos) using energy efficient resource management system for virtualized data centres with consolidation of VMS are achieved the energy saving according to resource utilization using the live migration results are presented of simulation driven evaluation for VMS dynamic reallocation according to CPU performance equipment . This can results the substantial energy saving and also ensures the reliable Qos.

B. Urgaonkar and C. Chandra. [14]. In this paper, author can proves novel dynamic capacity technique for the multitier internet application that employs the flexible queuing model is used for determining that how much resources allocated to the each tier and predictive and reactive methods combination that used to determine when to provision these resources the experiments demonstrate the techniques for having dynamic workload. This technique doubles the application capacity in five minutes that maintains the response time.

H. N. Van, F. D. Tran, and J.-M. Menaud. [15]. The main aim for data centers in cloud computing is to improve the profit and minimizing the power consumption and maintains SLAs. In this paper, author can describes a framework foe resource management that combines a dynamic virtual machine placement manager and dynamic VM provisioning manager. It can take several experiments that how system can be controlled to make trade-offs between energy consumption and application performance.

S. V. Vrbsky, M. Lei, K. Smith, and J. Byrd. [16]. The energy cost of data centers are rapidly growing now a days, so we use serer consolidation for reduce the energy cost. In this

paper, author analyze the workload of servers by observing potentials for power saving. It also investigates the low risk consolidation. From analysis two new methods are designed that can achieved the power saving.

PROBLEM IDENTIFIED:

On-the-field experiments are mainly focused on the offered QoS, they are based on a black box approach that makes difficult to correlate obtained data to the internal resource management strategies implemented by the system provider. Simulation does not allow to conduct comprehensive analyses of the system performance due to the great number of parameters that have to be investigated. Energy optimization in large-scale data centers

PROPOSED SYSTEM:

Recently, Gartner research reported that the average server utilization in large data-centers is 18% [21], while the utilization of x86 servers is even lower, 12%. These results confirm earlier estimations that the average server utilization is in the 10−30% range [5]. A 2010 survey [6] reports that idle servers contribute 11 million tones of unnecessary CO2 emissions each year and that the total yearly costs for idle servers is $19 billion The alternative to the wasteful resource management policy when the servers are always on, regardless of their load, is to develop energy-aware load balancing policies. Such policies combine dynamic power management with load balancing and attempt to identify servers operating outside their optimal power regime and decide if and when they should be switched to a sleep state or what other actions should be taken to optimize the energy consumption. The term server consolidation is sometimes used to describe the process of switching idle systems to a sleep state.

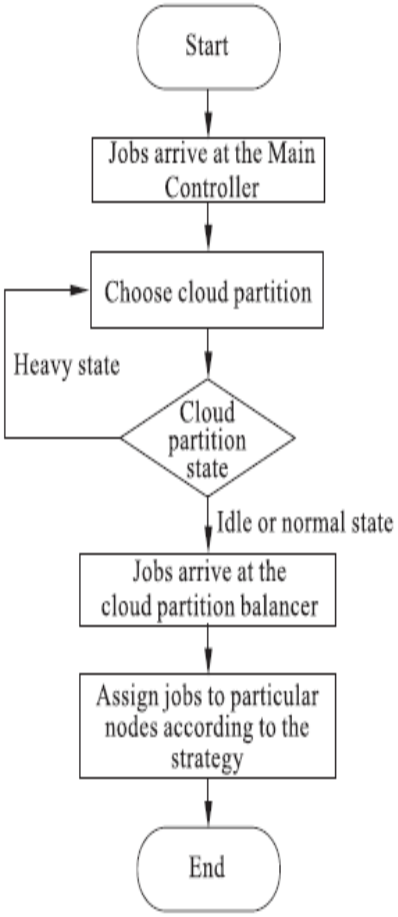
Challenges and metrics for energy-aware load balancing. Some of the questions posed by energy-aware load balancing are: 1. Under what conditions should a server be switched to a sleep state? 2. What sleep state should the server be switched to? 3. How much energy is necessary to switch a server to a sleep state and then switch it back to an active state? 4. How much time it takes to switch a server in a sleep state to a running state? 5. How much energy is necessary to migrate a VM running on a server to another one? 6. How much energy is necessary to start a VM on the target server? 7. How to choose the target for the migration of a VM? 8. How much time it takes to migrate a VM? Two basic metrics ultimately determine the quality of an energy-aware load balancing policy: (1) the amount of energy saved; and (2) the number of violations it causes. In practice, the metrics depend on the system load and other resource management policies, e.g., the

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 12
August 2016

admission control policy and the QoS guarantees offered. The load can be slow- or fastvarying, have spikes or be smooth, can be predicted or is totally unpredictable; the admission control can restrict the acceptance of additional load when the available capacity of the servers is low. What we can measure in practice is the average energy used and the average server setup time. The setup time varies depending on the hardware and the operating system and can be as large as 260 seconds [9]; the energy consumption during the setup phase is close the maximal one for the server. The time to switch the servers to a running state is critical when the load is fast varying, the load variations are very steep, and the spikes are unpredictable. The decisions when to switch servers to a sleep state and back to a running state are less critical when a strict admission control policy is in place; then new service requests for large amounts of resources can be delayed until the system is able to turn on a number of sleeping servers to satisfy the additional demand

IMPLEMENTATION:

ASSIGNING JOBS TO THE CLOUD:



When a job arrives to the cloud it will send to the controller, but the server act as in three ways

1. Idle: if cloud does not handling any request then the cloud is idle, in this state the cloud is ready to receive the request.
2. Normal: when cloud handling request then it is normal mode but the request's or not exceeded more than the limit.

3. Overload: if the cloud handling more than the request then it is overloaded.

CONCLUSION:

Load balancing is one of the main issues of cloud computing and balancing the load energy efficiently is more major task to do. In this paper, some energy aware load balancing algorithms are discussed. These techniques are aimed to allocate the resources to the vm requests in a way to reduce the energy consumption. Each of these have some merits and demerits. In future, we will try to design an algorithm that is able to overcome some of these demrits and can improve the resource utilization energy efficiently while considering other performance factors also.

REFERECES:

[1]. Kansal, Nidhi Jain, and Inderveer Chana. "Cloud load balancing techniques: A step towards green computing." IJCSI International Journal of Computer Science Issues 9, no. 1 (2012): 238-246.

[2]. G. Pallis, ―Cloud Computing: The New Frontier of Internet Computing‖, IEEE Journal of Internet Computing, Vol. 14, No. 5, September/October 2010, pages 70-73.

[3]. R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility, Future Generation Computer Systems, 25:599_616, 2009.

[4]. P. Mell and T. Grance, The NIST Definition of Cloud Computing, National Institute of Standards and Technology, Information Technology Laboratory, Technical Report Version 15, 2009

[5]. Jing, Si-Yuan, Shahzad Ali, Kun She, and Yi Zhong. "State-of-theart research study for green cloud computing." The Journal of Supercomputing 65, no. 1 (2013): 445-468.

[6]. Zhang, Qi, Lu Cheng, and Raouf Boutaba. "Cloud computing: state-of-the-art and research challenges." Journal of internet services and applications 1, no. 1 (2010): 7-18.

[7]. A. Khiyaita, M. Zbakh, H. El Bakkali, and D. El Kettani, ―Load balancing cloud computing: state of art,‖ in Network Security and Systems (JNS2), 2012 National Days of, pp. 106–109, IEEE, 2012

[8]. Marston, Sean, Zhi Li, Subhajyoti Bandyopadhyay, Juheng Zhang, and Anand Ghalsasi. "Cloud computing—The business perspective." Decision Support Systems 51, no. 1 (2011): 176-189

SANKARA BABU VEMULAPALLI: is pursuing M.Tech degree in, Computer Science and Engineering from Jagruti Institute of Engineering and Technology, Telangana State, India.

Mrs.N.SUJATHA is presently working as Associate Professor in, Department of computer science and engineering, Telangana State,India.She has published several research papers in both International and National conferences and Journals.