Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 9, September 2014, pg.502 – 509

RESEARCH ARTICLE

Various Load Balancing Techniques in Cloud Computing

Ashima Narang[1], Dr. Vijay Laxmi[2]

[1]Computer Science and Engineering, Guru Kashi University, India

[2]Computer Science and Engineering, Guru Kashi University, India

[1] ashimanarang04@gmail.com; [2] cse_vijay2003@yahoo.co.in

Abstract— now a day, the resources in the organization uses the power and consumes it by the unutilized Resources that are why the local cloud is becoming very popular. The requirement for cloud environments is not the reduction in consumption of the power only but also the requirement is to decrease the operating cost and improve the reliability of the system. Load balancing provides user satisfaction and also the ratio of resource utilization ratio after ensuring the allocation and efficiency of every resource being computed. This paper includes the various techniques existing for balancing the load in a cloud and their comparison on the basis of various parameters like performance, overhead, scalability etc.

Keywords - Cloud Computing, Virtual machine, Consolidation, Energy-Aware Scheduling, Load Balancing

I. INTRODUCTION

Cloud computing can be classified as a new paradigm for dynamic provisioning computer services supported by data centres that usually employ virtual machine (VM) technology for consolidation[1]. Cloud computing provides infrastructure, platform and software as services that is available to the consumer under the pay as you use model. The customers using a particular cloud, can access the resources provided by a cloud provider, according to the Service Level Agreement (SLA) given by the same cloud provider. In distributed data centres, the technology named virtualization is being used by the clouds to provide the resources to the customer whenever required. Clouds are provided to the customers for giving them three models: Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), and Infrastructure-as-a-Service (IaaS).

Load balancing is one of important issues in cloud computing [2]. It is a technique in which distribution of the dynamic local workload is done equally across the nodes in a cloud in order to avoid the situation where few nodes are overloaded while few are idle. The motive is to get a high user satisfaction and resource utilization ratio to be achieved; therefore it helps in improving the performance and utilization of the resources of the whole system. The popularity

of local cloud implementation is increasing due to the fact that commercialized cloud vendor are not much secure according to many organizations. The organizations using the private cloud can use various implementations of cloud computing while implementing their own private cloud. Some possible solutions for the same are Open Nebula [16] or Nimbus [18] or cloudbus[17]. This architecture achieves the availability and also the ease of scalability. The cloud includes several different types of hardware set ups. A cloud is built by single type of hardware; its nature is to expand various hardware types throughout its lifetime. The main part of power consumption in data centres come from computation processing, disk storage, networks and cooling system [15].

This paper is includes Section II Cloud Architecture. Section III describes energy-aware cloud architectural elements. Section IV describes existing load balancing techniques in cloud computing. Section V describes comparison of existing load balancing technique and Section VI conclusion.

II. CLOUD ARCHITECTURE

Layers:

Cloud computing architecture has the below given abstract layers which begins from bottom and works upwards. Figure 1 can be referred for the five layers that are constituted in cloud computing. The bottom most layer is known as the physical hardware (HaaS). The customers using the cloud for this particular layer are mostly the big corporations whose requirements are extremely large amount of Hardware as a Service. As a result, the cloud-provider runs, oversees, and upgrades its subleased hardware for its customers [4].

The next layer coincides of the cloud's software kernel. The layer acts as a path between the data being processed in the layer of Hardware and software infrastructure layer which is operating the hardware. It is the lowest level of abstraction implemented by the cloud's software and its main job is to manage the server's hardware resources while at the same time allowing other programs to run and utilize these same resources. [5]

The layer above the software kernel is the abstraction layer called the software infrastructure. This layer provides basic network resources to the two layers above it so that it can facilitate a new environment in a cloud that can be delivered to end users as an IT services. The services offered in the software infrastructure layer can be divided into three different categories: Computational resources (IaaS), data storage, and communication [5].

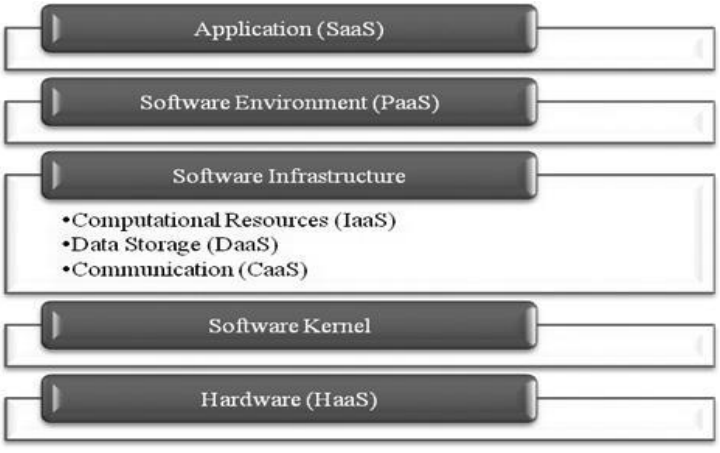

Figure: 1 Cloud Computing Architecture

III.ENERGY-AWARE CLOUD ARCHITECTURAL ELEMENTS

Figure 2 shows the high-level architecture for supporting energy-efficient service allocation in Cloud computing infrastructure [11]. There are four entities included:

a) *Consumers/Brokers*: Cloud consumers or their brokers request a service anywhere around the world to the Cloud. An important notice makes the difference between Cloud consumers and users of the deployed cloud services. For example, a company deploying a Web application can be a consumer that represents different workload as per the different number of "users" using it.

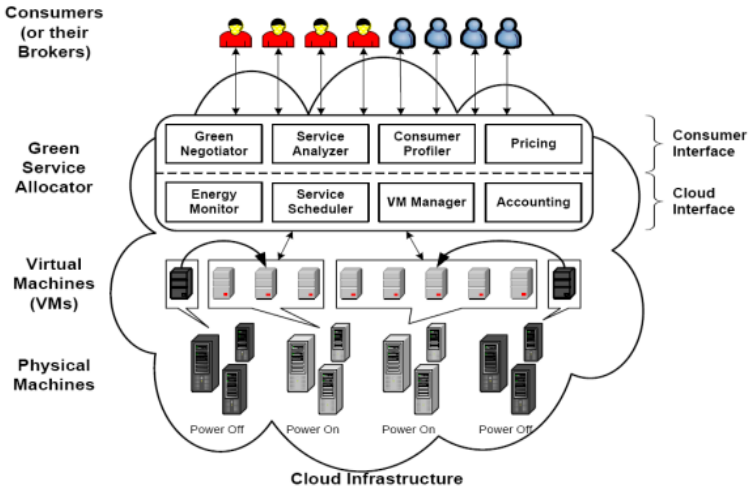

Figure 2: High-level system architectural framework

b) *Green Resource Allocator*: This is an interface between the Cloud infrastructure and consumers. It needs interaction for supporting energy- efficient resource management of the following components:

• *Green Negotiator*: To finalize the SLA, it negotiates with the consumers/brokers with specified prices and penalties (for violations of SLA) between the Cloud provider and consumer depending on the consumer's QoS requirements and energy saving schemes. In case of Web applications, for instance, QoS metric can be 95% of requests being served in less than 3 Seconds.[22]

• *Service Analyser*: Interprets and analyses the service requirements of a submitted request before deciding whether to accept or reject it. Hence, it needs the latest load and energy information from VM Manager and Energy Monitor respectively.

• *Consumer Profiler*: Gathers specific characteristics of consumers so that important consumers can be granted special privileges and prioritized over other consumers.

• Pricing: it decides how service requests are charged to manage the supply and demand of computing resources and facilitate in prioritizing service allocations effectively.

• *Energy Monitor*: it determines that which machine power should be on/off.

• *Service Scheduler*: it Assigns requests to VMs and decides the resources for VMs. It also decides when VMs are to be added or removed to meet demand.

•*VM Manager*: the availability of VMs and Their resource entitlements are traced. Also the migrating VMs across physical machines are kept in record by the VM manager

• *Accounting*: Historical usage information helps in improving the decisions of service allocation.

c) *VMs*: Multiple VMs can be dynamically started and stopped on a single physical machine to meet accepted requests, hence providing maximum flexibility to configure various partitions of resources on the same physical machine to different specific requirements of service requests. Multiple VMs can also concurrently run applications based on different

operating system environments on a single physical machine. In addition, by dynamically migrating VMs across physical machines, workloads can be consolidated and unused resources can be put on a low-power state, turned off or configured to operate at low-performance levels (e.g., using DVFS) in order to save energy.

d) *Physical Machines*: The underlying physical computing servers provide hardware infrastructure for creating virtualized resources to meet service demands.

IV. EXISTING LOAD BALANCING TECHNIQUES IN CLOUD COMPUTING

Following are the various load balancing techniques that are currently being used in cloud computing:

A. Dynamic Round-Robin algorithm

Dynamic Round-Robin [19] method is an extension to the Round-Robin method. It uses two rules that help to consolidate virtual machines. The first rule says that if a virtual machine has finished its assigned job and still other virtual machines are working that are hosted on the same physical machine, then this physical machine is not eligible to accept any new virtual machine. Such physical machines are called "retiring" state physical machines, that means we can only shut down the physical machine after all the other virtual machines finish their execution.

The second rule says that if there is a physical machine in the "retiring" state that is being used for long period of time, then instead of waiting for those virtual machines to finish, the physical machine is forced to transfer all the other virtual machines to other physical machines and then shutdown the first physical machine after the migration of the VMs finishes.

The threshold waiting time is represented by the "retirement threshold". A physical machine will be forced to transfer to all the virtual machines and then shut it down as it is in the retiring state but after the retirement threshold, all the other virtual machines could not finish.

These two rules are used by the Dynamic Round-Robin strategy so as to consolidate virtual machines implemented by the Round-Robin method. According to the first rule, adding extra virtual machines to a retiring physical machine is avoided. According to the second rule, the consolidation process become fast and it enables Dynamic Round-Robin to shutdown physical machines, such that the number of physical machine used to run all virtual machines is reduced , hence the power can be saved.

B. A Hybrid algorithm

For the conservation of energy, Ching-Chi Lin[19] proposed the combination of Dynamic Round-Robin and First-Fit to form a Hybrid algorithm. The probability distribution (e.g., a normal distribution) is followed and the number of incoming virtual machines is assumed as a function for time. Hybrid algorithm uses virtual machine's incoming rate for the scheduling of virtual machines. The First- Fit is used by the Hybrid method during rush hours to completely utilize the computing power of physical machines, and then it uses the Dynamic Round-Robin for the consolidation of the virtual machines and thus reduce the consumption of the energy in non-rush hours.

C. *Ant colony optimization (ACO)*:

Kumar Nishant Suggested an algorithm [24] of ant colony optimization. In ACO [24] algorithm when the request in initiated the ant start its movement. Movement of ant is of two ways*:*

Forward Movement : Forward Movement means the ant in continuously moving from one overloaded node to another node and check it is overloaded or under loaded ,if ant find an over loaded node it will continuously moving in the forward direction and check each nodes .*Backward Movement*: If an ant find an over loaded node the ant will use the back ward movement to get to the previous node, in the algorithm [24] if ant finds the target node then ant will commit suicide, this algorithm reduced the unnecessary back ward movement ,overcome heterogeneity, is excellent in fault tolerance.

D. Equally Spread Active Execution. (ESCE) algorithm

The estimation of the job size by the cloud manager and then checking for the for the availability of the virtual machine and also the capacity of the virtual machine. Once the available resource (virtual machine) size and the size of the job matches, then immediately the job scheduler allocates identified virtual machine or resource to the job in a queue. The affect of the ESCE algorithm [3] is that an improvement is seen in the response time and the processing time. The equal distribution of jobs is done, now the complete computing system is load balanced and there are no such virtual machines that are underutilized. Due to this merit, there is a reduction in the cost of virtual machine as well as the costof data transfer.

E. Load balancing mechanism based on ant colony and complex network theory (ACCLB) Algorithm

ACCLB load balancing mechanism [7] based on ant colony and complex network theory from the open cloud computing concepts. The use of the small-world and scale-free characteristics of a complex network is done to achieve efficient load balancing. This technique discourages heterogeneity, is adaptive to dynamic environments, It also encourages fault tolerance and has better scalability that helps in the improvement of the system performance.

F. Load Balancing Min-Min Algorithm (LBMM):

Wang suggested an algorithm called LBMM [23].

LBMM has a three level load balancing framework. In first level LBMM architecture is the request manager which is responsible for receiving the task and assigning it to service manager, when the service manager receives the request; it divides it into subtask and assigns the subtask to a service node based on node availability, remaining memory and the transmission rate which is responsible for execution the task.

G. *Honeybee Foraging Behavior -*

M. Randles et al.[25] investigated a decentralized honeybee-based load balancing technique that is a nature-inspired algorithm for self-organization. It achieves global load balancing through local server actions. Performance of the system is enhanced with increased system diversity but throughput is not increased with an increase in system size. It is best suited for the conditions where the diverse population of service types is required.

H. Resource Aware Scheduling Algorithm (RASA)-

Saeed Parsa and Reza Entezari-Maleki [26] proposed a new task scheduling algorithm RASA. It is composed of two traditional scheduling algorithms; Max-min and Min-min. RASA uses the advantages of Max-min and Min-min algorithms and covers their disadvantages. Though the deadline of each task, arriving rate of the tasks, cost of the task execution on each of the resource, cost of the communication are not considered. The experimental results show that RASA is outperforms the existing scheduling algorithms in large scale distributed systems.

V. COMPARISON OF EXISTING LOAD BALANCING TECHNIQUE

Below table show the comparative study of different load balancing. Difference made on bass of techniques that are used in respective algorithms, advantages and disadvantages.

TABLE 1: COMPARISONS OF DIFFERENT LOAD BALANCING ALGORITHMS

Algorithm	**Description**	**Advantages**
Dynamic Round-Robin[19]	The first rule Avoid adding extra virtual machines to a retiring physical machine. The second rule speeds up the consolidation process and enables Dynamic Round-Robin to shutdown physical machines.	1Power consumption is reduced. 2.Save power 3% more than power-sever implementation
Hybrid[19]	Combination of Dynamic Round Robin and First-Fit algorithms	1.Reduce Power consumption. 2. Easy to implement. 3. Response time is high.
ACO[24]	There are two types for movements: forward and backward. In the both the ant checks the load on the nodes and decides the next move over the node being overloaded or under loaded.	1. Detection of over loaded and under loaded nodes can be done. 2. Path tracing can be done consequently.
ESCE[3]	The random Selection based distributed problem round robin. Selection Depend on least load.	1.Response Time is high. 2. Processing time also high. 3. Simple and easy to implement.
ACCLB[7]	Uses small-world and scale-free characteristics of complex network to achieve better load balancing	1.Overcomes heterogeneity 2. Adaptive to dynamic environment 3.Excellent in fault tolerance 4Good scalability
Min-Min Algorithm	It starts with a set of unassigned tasks. Firstly, minimum time for the completion for all the tasks is found. Then the minimum number of times , the minimum value is selected in which the minimum times among the tasks on the resources. After that according to that minimum time, the scheduling of the task is done on the corresponding machine	1. The same pattern is followed again until all the assigned tasks are on the resources.
Honeybee Foraging Behavior	It achieves global load balancing through local server actions. Performance of the system is enhanced with increased system diversity.	1. Best suited for the conditions where the diverse population. 2. Performance of the system is enhanced.
Resource Aware Scheduling Algorithm (RASA)	Composed of two traditional scheduling algorithms; Max-min and Min-min. RASA uses the advantages of Max-min and Min-min algorithms and covers their disadvantages.	1. Outperforms the existing scheduling algorithms in large scale distributed systems. 2. Used to reduce make span

VI. CONCLUSION AND FUTURE WORK

Cloud Computing has widely been adopted by the industry or organization though there are many existing issues like Load Balancing, Virtual Machine Consolidation, Energy Management, etc. which have not been fully implemented. Central to these issues is the issue of load balancing, that is required to distribute the excess dynamic local workload equally to all the nodes in the whole Cloud to achieve a high user satisfaction.

In this paper, numerous proposed load balancing algorithms have been compared on the

basis of their advantages. But still some work need to be done to improve various factors like response time, throughput and scalability. In the next phase of our work, we will try to develop an algorithm which gives better results for load balancing using these parameters and we will compare their results with the existing algorithms.

REFERENCES

[1]. R. Yamini, *"Power Management In Cloud Computing Using Green Algorithm*", IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM-2012), March 30, 31, 2012.pp-128-133.

[2].R. Yamini, "*Energy Aware Green Task Assignment Algorithm In Clouds"*, International Journal For Research In Science And Advance Technology, Issue-1, Volume-1,pp-23-29.

[3]. Jaspreet kaur *"Comparison of load balancing algorithms in a Cloud"* 2012,International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 3, pp.1169-1173.

[4].Fei Hu, Meikang Qiu, Jiayin li, Travis Grant, Draw Tylor, Seth McCaleb, Lee Butler and Richard Hamner, *"A Review on Cloud Computing:Design Challenges in Architecture and security"*journal of Computing and Information Technology-CIR 19,2011.

[5]. LizheWang,Jie Tao, Marcel Kunze*"Scientific Cloud Computing:Early Definition and Experience*"The 10thIEEE International Confrrence Computing and Communications 2008.

[6]. Anton Beloglazov, Rajkumar Buyya, *"Managing Overload Host For Dynamic Consolidation Of Virtual Machines Cloud Data Centers Under Quality Of Service Constraints"*, IEEE Transaction On Parallel And Distributed Systems, 2012.

[7]. Zehua Zhang, Xuejie Zhang, *"A Load Balancing Mechanism Bassed On Ant Colony And Compel Network Theory In Open Cloud Computing Federation"*, IEEE- International Conference On Automation, May 2010, pp-240-243.

[8].Makhlouf Hadji, Djamal Zeghlache, *"Minimum Cost Maximum Flow Algorithm For Dynamic Resource Allocation In Cloud"*, IEEE-Fifth International Conference In Cloud Computing, Aug-2012, pp-876-882

[9]. Wenhong Tian, Yong Zhao,Minxian Xu, Chen Jing, *"A Dynamic And Integrated Load Balancing Scheduling Algorithm For Cloud Data Center*", Proceeding of IEEE CCIS Feb2011,pp-311-10].

[10]. Zenon Chaczko , Venkatesh Mahadevan , Shahrzad Aslanzadeh and Christopher Mcdermid *"Availability and Load Balancing in Cloud Computing"*, 2011 International Conference on Computer and Software Modeling IPCSIT vol.14 , IACSIT Press, Singapore.

[11]. Rajkumar Buyya, Anton Beloglazov, and Jemal Abawajy,*"Energy-Efficient Management of Data Center Resources for Cloud Computing: A Vision, Architectural Elements, and Open Challenges"*, Proceedings of the 2010
International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2010), Las Vegas, USA, July 12-15, 2010.

[12]. Trieu C. Chieu, Hoi Chan, *"Dynamic Resource Allocation Via Distributed Decisions In Cloud Environment"*, Eight IEEE International Conference on e-Business Engineering, Sept-2011, pp-125-130.

[13].Sivadon Chaisiri, Bu-Sung Lee, *"Optimization of Resource Provisioning Cost in Cloud Computing"*, IEEE transaction on services computing, vol. 5, No. 2 June 2012

[14].Ayman G. Fayoumi, *"Performance Evaluation of a Cloud Based Load Balancer Severing Pareto Traffic"*, Journal of Theoretical and Applied Information Technology, 15th October 2011. Vol. 32 No.1

[15]. Anton Beloglazov and Rajkumar Buyya, Adaptive Threshold-Based Approach for Energy-Efficient Consolidation of Virtual Machines in Cloud Data Centers, Proceedings of the 8th International Workshop on Middleware for Grids, Clouds and e-Science (MGC 2010, ACM Press, New York, USA), In conjunction with ACM/IFIP/USENIX 11th International Middleware Conference 2010, Bangalore, India, November 29 - December 3, 2010.

[16].*OpenNebula http://opennibula.org/*

[17]. *Cloudbus http://www.cloudbus.org/*

[18].*Nimbus*

[19] Ching-Chi Lin, Pangfeng Liu, Jan-Jan Wu. *"Energy- Efficient Virtual Machine Provision Algorithm for Cloud System"*, IEEE 4th International Conference on Cloud Computing,81-88, 09/2011.

[20] Jeffrey M. Galloway, Karl L. Smith, Susan S. Vrbsky. *"Power Aware Load Balancing for Cloud Computing"*, Proceedings of the World Congress on Engineering and Computer Science 2011 Vol. IWCECS 2011, October 19-21, San Francisco, USA,.

[21]. Namarata Swarnkar, Asst. Prof. Atesh Kumar Singh, Dr. R. Shankar, *"A Survey of Load Balancing*

Techniques in Cloud Computing", International Journal of Engineering Research & Technology, Vol. 2Issue8,August2013.
[22]. Karanpreet Kaur, Ashima Narang and Kuldeep Kaur, *" Load Balancing Techniques in Cloud Computing"*, IJMCR,2013.
[23]. Wang, S-C., K-Q. Yan, W-P. Liao and S-S. Wang, *"Towards a load balancing in a three-level cloud computing network,"* in proc. 3rd International Conference on. Computer Science and Information Technology (ICCSIT), IEEE, Vol. 1, pp: 108-113,July 2010.
[24] Nishant, K. P. Sharma, V. Krishna, C. Gupta, KP. Singh, N. Nitin and R. Rastogi, *"Load Balancing of Nodes in Cloud Using Ant Colony Optimization."* In proc. 14th International Conference on Computer Modelling and Simulation (UKSim), IEEE, pp: 3-8, March 2012.
[25] M. Randles, D. Lamb, and A. Taleb-Bendiab, *"A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing"*, Proceedings of 24th IEEE International Conference on Advanced Information Networking and Applications Workshops, Perth, Australia, April 2010, pages 551-556.
[26]. Saeed Parsa and Reza Entezari-Maleki," *RASA: A New Task Scheduling Algorithm in Grid Environment"* in World Applied Sciences Journal 7 (Special Issue of Computer & IT): 152-160, 2009.Berry M. W., Dumais S. T., O'Brien G. W. Using linear algebra for intelligent information retrieval, SIAM Review, 1995, 37, pp. 573-595.

