ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 3, March 2013

Effective Load Balancing For Dynamic Resource Allocation in Cloud Computing.

K Prasanna Kumar[1], S.Arun Kumar[2], Dr Jagadeeshan[3]

M.Tech(CSE) Student, SRM University,Ramapuram,Chennai,Tamil Nadu,India[1]

Assistant Professor, Department of IT, SRM University,Ramapuram, Chennai, Tamil Nadu,India [2]

Head of the Department, Department of IT, SRM University,Ramapuram, Chennai, Tamil Nadu,India [3]

Abstract — Resources are dynamic in nature so the load of resources varies with change in Configuration of cloud so the Load Balancing of the tasks in a cloud environment can significantly influence cloud's performance. A poor scheduling policy may leave many processors idle while a clever one may consume an unduly large portion of the total CPU cycles. In the existing approach we face overhead issue of distributed dispatching of task to resource. In our proposed system our main goal of load balancing is to provide a distributed, low cost, scheme that balances the load across all the processors. To improve the global throughput of cloud resources, effective and efficient load balancing algorithms are fundamentally important. Various strategies, algorithms and policies have been proposed, implemented and classified for implementing Load balancing in Cloud computing environment. In this paper, we present a combination of algorithm called ACBLA with queue algorithm applied to efficiently schedule computation jobs among processing resources onto the cloud datacenters with less communication overhead.

Keywords: Load Balancing, Communication overhead, Resource Allocation, ACBLA

I. INTRODUCTION

Cloud Computing is a representation for enable the suitable network access to shared pool of resources that can be promptly provisioned and free with minimal supervision effort or service provide relations. The cloud computing service models are Software as a Service (SAAs), Platform as A service (PAAs) and Infrastructure as a Service (IAAs). In a software service model contains a pre made application with required software, Operating system, hardware provided to other cloud service. User will install/develop the own application and software based on need of service. Best Example for cloud computing is Amazon Elastic Compute Cloud (EC2), EC2 cloud computing use IAAs Cloud service. By using the Virtual machines EC2 client creates n number of virtual machines for installing the request cloud software into each machine. Consequently, it is required to apply a special approach that guarantees the work load distribution across the cloud data centres to service n number of user service request per second; it is the job scheduling approach. This approach used to solve the current issue of load balancing with allocation of jobs to cloud data centres and equal Job execution time for all jobs.

Load balancing is emerging technology that facilitates utilization of resource by providing a throughput with minimum response time by sharing the equal load between servers. To achieve load balancing and resource utilization there are few algorithms are used. Best example for load balancing is online shopping cart. Without load balancing, users could experience delays while ordering, transactions and buying. Load balancing solutions usually apply redundant servers which help a better distribution of the communication traffic so that the online purchasing will made easy.

Load balancing is one of the priority issues in cloud computing. Load balancing is used to dispense the workload dynamically evenly across all partition nodes in the cloud data centre to avoid circumstances where some nodes are loaded heavy while others are leisure or doing little work. By improving the performance of system resource utilization, it helps to diminish from the Load balancing issue in cloud computing and also make sure that every computing resource is distributed efficiently and reasonably. When one or more service fail, load balancing helps to provisioning and de-provisioning of instances of applications without fail. The goal of load balancing is improving the performance by balancing the load among these various resources (network links, central processing units, disk drives.) to achieve optimal resource utilization, maximum throughput, maximum response time, and avoiding overload. To distribute load on different systems, different load balancing algorithms are used.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 3, March 2013

II. LITERATURE REVIEW

A study of some the existing algorithm is done and the results are tabulated. The algorithm and techniques existing until the current year have been considered and studied. The algorithm and the complete information of the approaches for dynamic resource allocation. A total number of 10 different research paper and different techniques, algorithm have been studied.

S. No	Title of the Paper	Goal	Techniques	Parameter used	Results
1	A Load Balancing Strategy for Virtual Storage	This paper focus on Virtual storage Load balancing	Three layers architecture with two load balancing modules to balance the load	Storage as a Service (SaaS) in Cloud serivce	Load Balancing Virtualization with effective resource utilization
2	Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment	In Cloud green computing achive the dynamically based resources on application demands by optimizing the numbers of servers	System that uses virtualization technology and SKEWNESS ALGORITHM	The concept of "skewness" to measure the unevenness in the multidimensional resource utilization of a server	improve the overall utilization of server resources
3	Research on Dynamic Resource Allocation with Cooperation Strategy in Cloud Computing	To define the Resource multiplexing	heuristic information-based algorithm	Resource Allocation; Cloud Computing; Dynamic; Cooperation Strategy; Parallel Processing	Resource allocation fast and effectively, achieving superior performance as well.
4	Cloud Task scheduling based on Load Balancing Ant Colony Optimization	In this we will reduce the optimization problem in task scheduling	Load Balancing Ant Colony Optimization algorithm	task scheduling; cloud computing; Load Balancing; Ant Colony Optimization; Cloud Sim	our work is to balance the entire system load while trying to minimizing the makespan of a given tasks set.
5	A SURVEY ON SCHEDULING BASED RESOURCE ALLOCATION IN CLOUD COMPUTING	To define dynamic flexible resource allocation for reliable and assured Qos	Topology Aware Resource Allocation (TARA), Linear Scheduling Strategy for Resource Allocation and Dynamic Resource Allocation for Parallel Data Processing.	cloud computing system; reliability; analytical model; resource allocation; quality of service.	system reliability of cloud applications with quality of services
6	A comprehensive survey: artificial bee colony (ABC) algorithm and applications	This paper used to resolve the real world Load balancing real word problems	Artificial Bee Algorithm	Bee swarm intelligence and Artificial Bee Algorithm	To improve the Resource utilization
7	A Survey on Honey Bee Inspired Load Balancing of tasks in Cloud Computing	Main goal of the load balancing technique is to optimize use of resources, maximize the throughput, reduce the response time, and avoid overload of any of the resources.	Bee Inspired Load Balancing of Tasks in Cloud Computing Environment.	Cloud computing, Load balancing, Honey bee foraging behavior	To Avoid resource overload

S. No	Title of the Paper	Goal	Techniques	Parameter used	Results
8	Novel Nature-inspired Algorithm to solve Complex Generalized Assignment Problems	This paper defines a survey on load balancing schemes in cloud environments.	Comparisons of different load balancing dynamic algorithms	Cloud computing, Distributed Virtual Environments, Load Balancing, Peer-to-Peer Systems.	By comparing all algorithm we have come up with different techniques
9	**A Survey on Load Balancing in Cloud Computing Environments**	In Cloud green computing achieve the dynamically based resources on application demands by optimizing the numbers of servers	System that uses virtualization technology and SKEWNESS ALGORITHM	The concept of "skewness" to measure the unevenness in the multidimensional resource utilization of a server	improve the overall utilization of server resources
10	Resource Allocation in Contending Virtualized Environments through Stochastic Virtual Machine Performance Modeling and Feedback	weighted priority based service differentiation strategy to allocate resources in contending conditions to provide performance guarantees as well as load balance and fairness.	Feedback based algorithm is valid	Resource allocation, virtualized environment, performance feedback, scheduling, workload characterization	The results show that the performance feedback based allocation can achieve a higher SLA satisfaction rate as 97.1%, a lower load imbalance index as 18.7%.

III. PROBLEM STATEMENT

Load balancing was notorious as a major worry and it leads to degrade the performance of the resource allocation in dynamic nature. This also makes more difficulty to dispatch the job to resource from the queue and one more difficulty is in provisioning or allocating the Job in cloud data centre. In the existing system more resource overload and energy consumption.

IV. EXISTING SYSTEM

Virtual machine monitors (VMMs) like Xen provide a mechanism for mapping virtual machines (VMs) to physical resources. This mapping is largely hidden from the cloud users. Users with the Amazon EC2 service, for example, do not know where their VM instances run. It is up to the cloud provider to make sure the underlying physical machines

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 3, March 2013

(PMs) have sufficient resources to meet their needs. VM live migration technology makes it possible to change the mapping between VMs and PMs While applications are running. The capacity of PMs can also be heterogeneous because multiple generations of hardware coexist in a data centre.

V. PROPOSED SYSTEM

We have proposed the combination novel based algorithm called to ACBLA with queue algorithm. By using the ACBLA algorithm we going to improve the Load balancing in cloud environment by portioning the nodes into two idle and busy nodes with help of this partioning queue monitoring algorithm, this will dispatch the job based on priority of the task and allocate the resource dynamically. In our approach especially we are reducing the communication overhead and queue waiting by monitoring the target Virtual machine idle time and completion time this can be calculated by using resource start and resource completion time. ACBLA algorithm is used to collect the data about source node and end node. This helps us to allocate the resource dynamically with efficient LB during the run time and map the resource to virtual machine with allotted physical machine. And finally we use a precedence based service separation policy to allocate resources in challenging conditions to provide enactment guarantees as well as load balance and fairness.

Highlights of proposed system

- ✓ Improve the Load balancing in cloud environment by portioning the nodes into two idle and busy node
- ✓ Communication overhead and queue waiting reduced
- ✓ No longer wait for low priority task , this can be achieved by monitoring the resource execution and calculating the idle time
- ✓ All service request will validated and completed within in mentioned SLA.
- ✓ Overload resource avoidance and energy consumption.

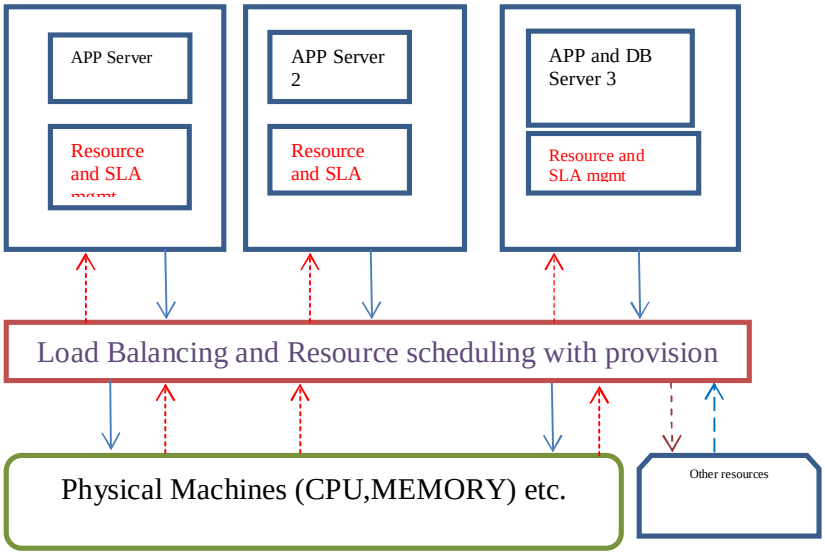

Fig. 1. The framework For resource allocator and Load Balancing.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have suggested a new combination of novel based algorithm called ACBLA with queue monitoring algorithm. This novel combination algorithm used to fragging and phenrome the nodes and provisions the job dynamically without any communication overhead. Queue monitoring algorithm helps to reduce the low priority longer waiting by calculating the idle and execution time of virtual machine.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 3, March 2013

In future, we are planning to improve the load balancing in dynamic based web site and work scheduling in any Operating system. To establish static resource allocation by both static and dynamic load balancing algorithm by mapping the resource to specific virtual machine and then migrating the virtual machine with available physical machine if additional resource required. And also save more power energy consumption with help of green computing by calculating the distance of the resource and execution time.

REFERENCES

1. E. Anderson, D. Patterson, E. Brewer, "The Magicrouter, an application of fast packet interpo.sing", University of California, Berkeley,May 1996.
2. D. Andresen, T. Yang, V. Holmedahl, O.H. Ibarra, \SWEB: Toward a scalable World Wide Web-server on multicomputers", Proc. of 10th IEEE Int'l. Symp. on Parallel Processing, Honolulu,April 1996.
3. M.F. Arlitt, C.L. Williamson, \Web-server workload characterization: The search for invariants",IEEE/ACM Trans. on Networking,vol. 5, no. 5, pp. 631,645, Oct. 1997.
4. M. Baentsch, L. Baum, G. Molter, "Enhancing the Web's infrastructure: From caching to replication",IEEE Internet Computing,vol. 1, no. 2, pp. 18 &27, Mar.-Apr. 1997.
5. P. Barford, M. Crovella, "Generating representative Web workloads for network and server performance evaluation", Proc. of ACM Sigmetrics '98, Madison, WI, pp. 151-160, June 1998.
6. P.Boyle, \Web site traÆc cops: "Load balancers can provide the busiest Web sites with nonstop per-formance", PC magazine,Feb. 1997.
7. CARDELLINI, M. COLAJANNI, P.S. YU, "DNS DISPATCHING ALGORITHMS WITH STATE ESTIMATORS FOR SCALABLE WEB-SERVER CLUSTERS", WORLD WIDE WEB JOURNAL, BALTZER SCIENCE PUBL., VOL. 2, JULY-AUG. 1999.
8. [10] V. Cardellini, M. Colajanni, P.S. Yu, "Redirection algorithms for load sharing in distributed Web-server systems", Proc. of 19th IEEE Int'l. Conf. on Distributed Computing Systems (ICDCS'99), Austin, TX,, June 1999.
9. M. Colajanni, P.S. Yu, D.M. Dias, "Analysis of task assignment policies in scalable distributed Web-server systems", IEEE Trans. on Parallel and Distributed Systems, vol. 9, June 1998.
10. Chun-Tian, C. and L. Zhi-Jie,. "Parallel algorithm for grid resource allocation based on Nash equilibrium" PP.53-66 , 2006.
11. Chao, K.M., R. Anane, J. H. Chen and R. Gatward,. "Negotiating agents in a market-oriented grid". Proceeding. of the 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid, 2002
12. Colorni, A., M. Dorigo and V. Maniezzo, "Ant colony system for job-shop scheduling." 1999.
13. Oper. Res. St. Comput. Sci., Cui, T., T. Xue and K. Nahrstedt." Optimal Resource Allocation in Overlay Multicast". ICNP, 137-172, 2003
14. Czajkowski, K., I. Foster, C. Kesselman, V. Sander and S. Tuecke, "BSNAP: A protocol for negotiating service level agreements and coordinating resource management in distributed systems", 2002.
15. N. Bobroff, A. Kochut, and K. Beaty, "Dynamic Placement of Machines for Managing SLA Violations," Proc. IFIP/IEEE Int'l Symp.Integrated Network Management (IM '07), "TPC-W: Transaction Processing Performance Council," 2007 http:// www.tpc.org/tpcw/,2012,
16. J.S. Chase, D.C. Anderson, P.N. Thakar, A.M. Vahdat, and R.P. Doyle, "Managing Energy and Server Resources in Hosting Centers," Proc. ACM Symp. Operating System Principles (SOSP '01), Oct. 2001.
17. C. Tang, M. Steinder, M. Spreitzer, and G. Pacifici, "A Scalable Application Placement Controller for Enterprise Data Centers," Proc. Int'l World Wide Web Conf. (WWW '07), May 2007.
18. M. Zaharia, A. Konwinski, A.D. Joseph, R.H. Katz, and I. Stoica, "Improving MapReduce Performance in Heterogeneous Environments,"
19. Proc. Symp. Operating Systems Design and Implementation (OSDI '08), 2008.
20. M. Isard, V. Prabhakaran, J. Currey, U. Wieder, K. Talwar, and A. Goldberg, "Quincy: Fair Scheduling for Distributed Computing Clusters," Proc. ACM Symp. Operating System Principles (SOSP '09), Oct. 2009.
21. Li, L., Yi, Y., Lian, L., and Wanbin, S.,"Using Ant Colony Optimization for SuperScheduling in Computational Grid" in 2006 IEEE Asia-Pacific Conference on Service Computing, ISBN: 07695-2751-5, 2006.
22. Liang, B., Yanli, H., Songyang, L., Weiming, Z., "Task Scheduling with Load Balancing using Multiple Ant Colonies Optimization in Grid Computing" in 2010 Sixth International Conference on Natural Computation (ICNC 2010), DIO: 10.1109/ICNC.2010.5582599, pp.2715-2719, 2010.
23. Patricia Takako Endo, Andre Vitor de Almeida Palhares, Nadilma Nunes Pereira, 2011. Resource Allocation for Distributed Cloud: Concepts and Research Challenges, IEEE, july 2011.
24. M.Asad Arfeen, Krzysztof Pawlikowski, Andreas Willig .2011, A Framework for Resource Allocation Strategies in Cloud Computing Environment, 2011 35th IEEE Annual Computer Software and Applications Conference Workshops.
25. Atsuo Inomata, Taiki Morikawa, Minoru Ikebe. 2011, Proposal and Evaluation of a Dynamic Resource Allocation Method based on the Load of VMs on IaaS, IEEE 2011.
26. Ikki Fujiwara, Isao ono, Kento Aida, Applying Double-sided Combinational Auctions to Resource Allocation in Cloud Computing, 10th Annual International Symposium on Applications and the Internet, 2010

BIOGRAPHY

K.Prasanna Kumar is a M.Tech (CSE) Student in the Department of Computer Science and Engineering, SRM University, Ramapuram, Chennai, Tamil Nadu, India. He received Bachelor of Technology in Information Technology degree in 2006 from Anna University, Guindy, Tamil Nadu, and India. He research interests are Cloud computing and Load balancing.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 3, March 2013

S.ArunKumar is an Assistant Professor in the Department of Computer Science and Engineering, SRM University, Ramapuram, Chennai, Tamil Nadu, India. He received M.Tech (CSE) degree in2010 from SRM University, Chennai, Tamil Nadu, India .He has been professional teaching experience in 6 years. He research interests are Cloud Computing in Network Security