SURVEY ON VARIOUS LOAD BALANCING AND SCHEDULING METHODS IN CLOUD COMPUTING

[1]Ankit Shrivastava1, [2]Umesh Lilhore, [3]Nitin Agrawal

[1]M.Tech Research Scholar, [2]Associate Professor & PG Coordinator, [3]Associate Professor
NIIST Bhopal (M.P), India

***ABSTRACT*: - Cloud computing is a new and innovative technology in current digital world, which is basically based on sharing of computing resources over the network at "pay per use basis. Cloud provides various on demand computing services such as PaaS, SaaS and IaaS. Cloud computing reduce cost of ownership and provides optimum utilization of computing resources, which interacts cloud users. Day by day number cloud service providers, computing resources and cloud users are increasing rapidly, this rapid growth creates challenges for cloud service providers to uniformly serve computing resources for all the cloud users. This challenges attract cloud researchers to create efficient scheduling and balancing method for cloud computing. In this survey paper we are presenting study and analysis of various performance improving methods for cloud computing.**

***Keywords*- Cloud computing, Load balancing, Task scheduling, Resource sharing**

I. INTRODUCTION

The word cloud in cloud computing is used as a metaphor for internet so we can define a cloud computing as the internet based computing in which the different services like storage, servers and application are provided to organizations computers and device using internet [2]. So as compared to traditional "own and use" technique if we use cloud computing, the purchasing and maintenance cost of infrastructure is eliminated.

Cloud Computing is a concept of distributed computing in which one has the ability to run an application on many connected systems at the same time. It provides a utility service giving one an access to technology resources that are available on demand. With cloud computing coming into existence, the focus has shifted from provisioning of the infrastructure to business, thus increasing the efficiency of production [5,20]. Cloud computing provides reliable and dependable on-demand infrastructure and services that reduce time and expenses.

1.2. Load Balancing & Scheduling in Cloud Computing

Load balancing in cloud computing systems is really a challenge now. Always a distributed solution is required, because it is not always practically feasible or cost efficient to maintain one or more ideal services just as to fulfill the required demands [11]. Jobs can't be assigned to appropriate servers and clients individually for efficient load balancing as cloud is a very complex structure and components are present throughout a wide spread area.

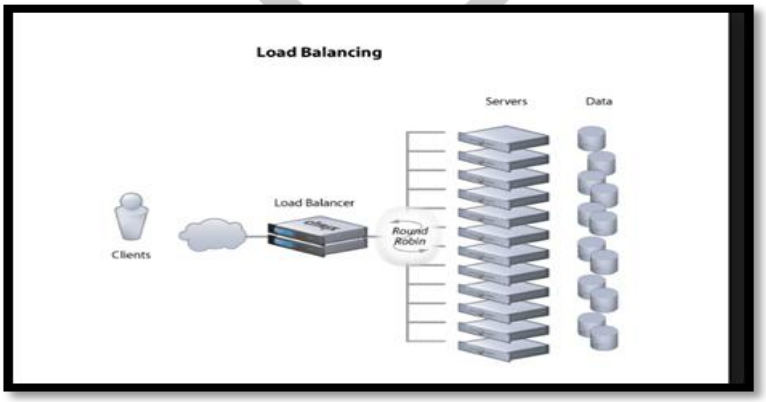

Figure 1.1 Load Balancing in Cloud [21]

"Load balancing "means an even distribution of the total load amongst all serving entities. Load balancing is very essential in distributed computing systems to improve the quality of service by managing customer loads that are changing over time. The request demands of incoming requests are optimally distributed among available system resources to avoid resource bottlenecks as well as to fully utilize available resources [17].

1.3 Why Load Balancing in Cloud?

Load balancing, in general, refers to the method of distribution and allocation of certain tasks amongst the available resources in an efficient way that promotes even and wise utilization [13].

Load balancing methods are important in cloud because it provides-

- **Scalability-** The load balancing algorithm must provide scalability in terms of addition of new resources to address the everyday increasing demand of services with a greatly increasing number of users. This also demands flexibility in accommodating the change [6].
- **Better response time-** The load balancing must be implemented and executed well enough to provide the best possible response to a user.
- **Cost effectiveness-** A good load balancing algorithm must aim for better overall system performance with the cost being quite reasonable.
- **Prioritization-** The tasks must be prioritized so that the critical tasks do not have to face the problem of starvation, or if addressed, the problem of late response.
- **Fair node utilization-** The serving nodes must be utilized efficiently so that no single node is overwhelmed, leaving certain others totally free, or lightly loaded [9].

1.4 Load Balancing Methods in Cloud Computing

In complex and large systems, there is a tremendous need for load balancing. For simplifying load balancing globally, techniques are employed that would act at the components of the clouds in such a way that the load of the whole system is distributed [1, 4].Load balancers implement type specific algorithms to make load balancing decisions. Load balancing algorithms can be divided into two categories, Static algorithm and dynamic algorithm.

- ***Static Algorithms-*** A static load balancing algorithm does not take into account the previous state or behaviour of a node while distributing the load. In this approach prior knowledge of system is needed. This has a major impact on the overall system performance due to the unpredictability of load fluctuation of the distributed system. It doesn't depend upon current state of system. Static algorithms are much simpler as compared to dynamic algorithms.

- ***Dynamic Algorithms-***This approach takes into account the current state of the system during load balancing decisions and is more suitable for widely distributed systems such as cloud computing [15].The dynamic load balancing algorithm is applied either as a distributed or non-distributed. The advantage of using dynamic load balancing is that if any node fails, it will not halt the system; it will only affect the system performance. In a dynamic load balanced system, the nodes can interact with each other generating more messages when compared to a non-distributed environment

II RELATED WORK

The evolution of cloud can be dated back to the 1960's from the ideas of pioneers like J.C.R Licklider and John McCarthy. John McCarthy believed that computation cloud organized for public entity. Cloud can be accessed anytime using the internet. So the development of the Internet was one of the major milestones for cloud computing [10].

In 1969, Leonard Klein rock, one of the chief scientists of the original Advanced Research Projects Agency Network (ARPANET) which seeded the Internet, said: "As of now, computer networks are still in their infancy, but as they grow up and become sophisticated, I will probably see the spread of "computer utilities" which, like present electric and telephone utilities, will service individual homes and offices across the country."

Cloud computing is emerging as a new paradigm of large scale distributed computing. It has moved computing and data away from desktop and portable PCs, into large data centers. It has the capability to harness the power of Internet and wide area network (WAN) to use resources that are available remotely, thereby providing cost effective solution to most of the real life requirements. This connection of networks came to be known as Internet.

Following load balancing techniques are currently prevalent in clouds-

- **Round Robin Algorithm-**Round Robin is a very famous load balancing algorithm, in which the processes are divided between all processors. The process allocation order is maintained locally independent of the allocations from remote processors[24].
  In Round Robin, it send the requests to the node with the least number of connections, so at any point of time some node may be heavily loaded and other remain idle [5], this problem is reduced by CLBDM.

- **Central Load Balancing Decision Model (CLBDM)-** CLBDM is a central load balancing decision model, which is suggested by Radojevic and Mario Zagar [9], it's based on session switching at the application layer. The improvement is that, in the cloud it calculated the connection time between the client and the node, and if that connection time exceeds a threshold then connection will be terminated and task will be forwarded to another node using the regular Round Robin rules.

- **Map Reduce-based Entity Resolution** Map Reduce is a computing model and an associated implementation for processing and generating large datasets [10]. Map task and reduce task two main task in this model which written by the user, Map takes an input pair and produces a set of intermediate value pair and Reduce task accepts an intermediate key and a set of values for that key and merges these values to form a smaller set of value. Map task read entities in parallel and process them, this will cause the Reduce task to be overloaded.

- **Ant colony optimization (ACO)-**Kumar Nishant suggested an algorithm [11] of ant colony optimization. In ACO 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 [11] 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.

- **Load balancing of virtual machine resources-**J. Hu et al. [12] proposed a scheduling strategy on load balancing of VM resources that uses historical data and current state of the system. This strategy achieves the best load balancing and reduced dynamic migration by using a genetic algorithm. It helps in resolving the issue of load-imbalance and high cost of migration thus achieving better resource utilization.

- **Index Name Server Algorithm (INS)-**The INS algorithm proposed in [13] the goal is to find an algorithm to minimize the data duplication and redundancy. INS is able to handle the load balancing dynamically .INS have some parameters which help in calculating the optimum selection point like that Hash Code of the block of data to be downloaded, the position of the server, the transition quality, the maximum bandwidth. Another calculation point whether the connection can handle additional nodes or not.

- **Opportunistic Load Balancing (OLB)-**Sang proposed OLB is a static load balancing algorithm that has the goal of keeping each node in cloud busy [14]. However OLB does not calculate the execution time of the node, due to this the tasks to be processed in a slower manner and will cause bottlenecks since requests might be pending waiting for nodes to be free.

- **Honeybee Foraging Behavior-**M. Randles et al. [19] investigated a decentralized honeybee-based load balancing technique that is a natureinspired 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.

- **Biased Random Sampling-**M. Randles et al. [4] investigated a distributed and scalable load balancing approach that uses random sampling of the system domain to achieve self-organization thus balancing the load across all nodes of the system. The performance of the system is improved with high and similar population of resources thus resulting in an in-creased throughput by effectively utilizing the increased sys-tem resources. It is degraded with an increase in population diversity.

- **Active Clustering**-M. Randles et al. [11] investigated a self-aggregation load balancing technique that is a self-aggregation algorithm to optimize job assignments by connecting similar services using local re-wiring. The performance of the system is enhanced with high resources thereby in-creasing the throughput by using these resources effectively. It is degraded with an increase in system diversity. ACCLB:

III COMPARISION OF VARIOUS LOAD BALANCING METHODS

Various metrics considered for comparison of above discussed Load balancing techniques in cloud computing are discussed below [11]

- **Nature of an algorithm** defines the nature or behavior of load balancing algorithm i.e. whether static or dynamic pre-planned or no planning.

- **Process Migration parameter** provides when does a system decide to export a process? It decides whether to create it locally or create it on a remote processing element. The algorithm is capable to decide that it should make changes of load distribution during execution of process or not.

- **Resource Utilization** used to check the utilization of available resource given to the cloud. Static load balancing algorithms have lesser resource utilization as static load balancing methods just tries to assign tasks to processors in order to achieve minimize response time ignoring the fact that may be using this task assignment can result into a situation in which some processors finish their work early and sit idle due to lack of work.

- **Stability** can be characterized in term of the delay in the transfer of information between processor and the gain in load balancing algorithm. Static load balancing algorithm considered as stable as no information regarding present workload state is passed processors. However in case of dynamic load balancing such kind of information is exchanged among processors.

- **Predictability** factor is related with the deterministic or nondeterministic factor that is to predict the outcome of the algorithm. Static load balancing algorithm's behaviour is predictable as most of the things like average execution time of processes and workload assignment to processors are fixed at compile-time. Dynamic load balancing algorithm's behaviour is unpredictable, as everything has been done at run time.

- **Reliability** factor is related with the reliability of algorithms in case of some machine failure occurs. Static load balancing algorithms are less reliable because no task/process will be transferred to another host in case a machine fails at run-time. Dynamic load balancing algorithms are more reliable as processes can be transferred to other machine in case of failure occurs.

- **Adaptability** factor is used to check whether the algorithm is adaptive to varying or changing situations i.e. situations which are of dynamic nature. Static load balancing algorithms are not adaptive as this method fails in varying nature problems. Dynamic load balancing algorithms are adaptive towards every situation whether numbers of processes are fixed or varying one.

- **Response Time** is defined as how much time a distributed system using a particular load balancing algorithm is taking to respond? Static load balancing algorithms have shorter response time as one should not forget that in Static load balancing there is lesser overhead so emphasis is totally on executing jobs in shorter time rather than optimally utilizing the available resources. Dynamic load balancing algorithms may have relatively higher response time.

- **Fault Tolerant** enables an algorithm to continue operating properly in the event of some failure. If the performance of algorithm decreases, the decrease is proportional to the seriousness of the failure, even a small failure can cause total failure in load balancing.

✓ **A Comparative analysis of above discussed load balancing algorithms is done below-**

Parameters	Round Robin Algorithm	Throttle based algorithm	FCFS	Honey Bee	ACO
Nature	Static	Dynamic	Dynamic	Dynamic	Dynamic
Process Migration	No	Yes	No	Yes	Yes
Resource Utilization	Less	More	Less	More	More
Stability	Large	Small	Large	Small	Small
Predictability	More	Less	More	Less	Less
Reliability	Less	More	Less	More	More
Adaptability	Less	More	Less	More	More
Response Time	Less	More	Less	More	More
Fault Tolerant	No	No	No	Yes	No

Table 5.1 Comparisons of various load Balancing methods

IV CONCLUSIONS & FUTURE WORKS

This survey paper presents a comprehensive analysis on various load balancing methods. We have studied different load balancing techniques in the cloud environment; we have discussed main issues of these algorithms which must be taken into consideration while designing any load balancing algorithms we have discussed the already proposed algorithms by various researchers in literature, their advantages and disadvantages. A comparison has been done on the basis of different criteria like Throughput, Overhead, Response time, Scalability, Response time, etc.

In future work we will proposed an efficient load balancing and scheduling method for cloud computing to improve the cloud performance and also compare this proposed method with various excising methods to validate the research.

REFERENCES

[1] Ali Al Buhussain, Robson E. De Grande, Azzedine Boukerche, "Performance Analysis of Bio-Inspired Scheduling Algorithms for Cloud Environments ",2016 IEEE International Parallel and Distributed Processing Symposium Workshops 2016 IEEE International Parallel and Distributed Processing Symposium Workshops, 2016, PP 776-786.

[2] Chaokun Zhang∗, Yong Cui∗, Rong Zheng†, Jinlong E∗, and Jianping Wu∗, "Multi-Resource Partial-Ordered Task Scheduling in Cloud Computing", Science Direct 2016, PP285-291

[3] Lekha Nema, Avinash Sharma, "Efficient Load Balancing Based on Improved Honey Bee Method in Cloud Computing". IJCTA, 9(22), 2016, pp. 151-161

[4] Khushbu Zalavadiya, Dinesh Vaghela," Honey Bee Behavior Load Balancing of Tasks in Cloud Computing ",International Journal of Computer Applications(0975-8887) Volume 139 – No.1, April 2016.

[5] Mayanka Katyal, Atul Mishra" A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment" International Journal of Distributed and Cloud Computing Volume 1 Issue 2 December 2013J.

[6] Dinesh Babu L D, P.Venkata krishna, "Honey bee behavior inspired load balancing of tasks in cloud computing environments" Applied Soft Computing Volume 13 Issue 5, Elsevier, May 2013, Pages 2292-2303.

[7] Agraj Sharma, Sateesh K. Peddoju "Response Time Based Load Balancing in Cloud Computing" 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT).

[8] J. Li, S. Su, X. Cheng, Q. Huang Z. Zhang "Costconscious scheduling for large graph processing in the cloud" IEEE 2011, 13th International Conference on High Performance Computing and Communications (HPCC), pp. 808–813.

[9] Sen Su, Jian Li,Kai Shuang,Jie WangFröhlich, B."Costefficient task scheduling for executing large programs in the cloud". Parallel Computing, Volume 39, March 2013, pages 177-188

[10] Santanu Dam, Gopa Mandal, Kousik Dasgupta, and Paramartha Dutta "An Ant Colony Based Load Balancing Strategy in Cloud Computing" 2014 Advanced Computing, Networking and InformaticsVolume 2, Springer International Publishing Switzerland.

[11] Shridhar G.Domanal, G.Ram Mohana Reddy "Load Balancing in Cloud Computing using Modified Throttled Algoritham" : Department of Information Technology National Institute of Technology Karnataka Surathkal, Mangalore, India

[12] Sheeja Y S, Jayalekshmi S "Cost Effective Load Balancing Based on honey bee Behaviour in Cloud Environment"2014 First International Conference on Computational Systems and Communications (ICCSC) | 17-18 December 2014 | Trivandrum.

[13] Rodrigo N. Calheiros, Rajiv Ranjan, Anton Beloglazov, C´esar A. F. De Rose and Rajkumar Buyya1, "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms", Published online in Wiley Online Library (wileyonlinelibrary.com), 24 August 2010,

[14] Agarwal A., Jain S., 2014, "Efficient Optimal Algorithm of Task Scheduling in Cloud Computing Environment", International Journal of Computer Trends and Technology (IJCTT), vol. 9 No. 7, pp. 344-349.

[15] Dhinesh Babua L.D., Venkata Krishna P., 2013 "Honey bee behavior inspired load balancing of tasks in cloud computing environments",applied SoftComputing, Vol. 13, Issue 5, pp. 2292-2303.

[16] Fang Y., Wang F., Ge J., 2010, "A task scheduling algorithm based on load balancing in cloud computing", Web Information Systems and Mining,Lecture Notes in Computer Science Vol. 6318, pp. 271-277.

[17] Keshk A. E., El-Sisi A., Tawfeek M. A., Torkey F. A., 2013, "Cloud Task Scheduling Based on Ant Colony Optimization", 8th International Conference on Computer Engineering & Systems (ICCES), , pp. 64-69.

[18] keshk A. E., El-Sisi A., Tawfeek M. A., Torkey F. A., 2013, "Intelligent Strategy of Task Scheduling in Cloud Computing for Load Balancing",International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 2, Issue 6, pp. 12-22.

[19] Kumar P., Verma A., 2012, "Independent task scheduling in cloud computing by improved genetic algorithm", International Journal of Advanced Research in Computer Science and Software Engineering (ijarcsse), Vol. 2, Issue 5, pp. 111-114.

[20] Lombardi F., Di Pietro R., 2011, "Secure virtualization for cloud computing", Journal of Network and Computer Applications, Vol. 34, pp. 1113–1122. Ma L., Lu Y., Zhang F., Sun S., 2013,

[21]"Dynamic Task Scheduling in Cloud Computing Based on Greedy Strategy", Trustworthy Computing and Services Communications in Computer and Information Science Volume 320, pp 156-162.