

A Survey of Load Balancing Algorithms in Cloud Computing

Akansha Makhija¹ & Usvir Kaur²

¹(Student) ²(Assistant professor)

Department of Computer Science,

Sri Guru Granth Sahib World University Fatehgarh sahib, Punjab, India

***Abstract-** Load balancing is a method for reassigning the entire workload to the separate nodes for a group of systems, in order to make the time more efficient and utilize the resources in efficient manner. It is the basic technique for scaling out an application server infrastructure. As request demand increases, new servers can be added to the resource pool, were the load balancer will directly begin sending traffic to the new server. The aim here is to distribute the load amongst the available nodes in such a way that no single node is flooded with requests, while some other node is lightly loaded. In this paper, different kinds of load balancing algorithms are analyzed for cloud computing, to make the network load equally divided in order to provide faster connectivity for all the devices that need the service. Load balancing algorithms play a challenging task in cloud computing.*

Keywords: cloud computing, load balancing

I. INTRODUCTION

Cloud computing is an evolving area that allows users to organize applications with enhanced scalability, availability and fault tolerance. Cloud computing provides internet based platform that is used for computer technology. It describes a diversity of computing concepts. Cloud computing accumulates all the computing resources and manages them automatically. Nowadays world depends on cloud computing to store the public as well as personal information. Cloud computing provides relevant hardware, software and service according to the requirement that users put forward. A cloud computing structure is categorized by its on-need self service, access over internet, pooling of resources, elasticity of service availability and measurement of services utilized by individual users. Cloud computing provides a collective group of resources, including data storage space, networks, computer processing power and specialized corporate and user application. There are four deployment models in

cloud computing. They are Public, Private Community and Hybrid. [1]

Cloud Services Cloud computing provides a number of clouds for providing services. Services can be larger or smaller, and use of a service is measured and customers are billed accordingly. Service Models are the orientation models on which cloud computing is based. These can be categorized into three basic service models as: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS). [1]

II. LOAD BALANCING

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. In computing, load balancing is a networking method that distributes the workload across multiple computing resources such as computers and their cluster, network links, central processing units or disk drives. The aims of load balancing remain to optimize resource use, minimize the response time, maximize throughput and avoid a single resource from being overloaded. Load balancing is often implemented in software, though it can also be performed using hardware or even the combination of software and hardware. A load balancer, as a software program, listens on the port where external clients connect to access services. Requests are forwarded by the load balancer to the backend servers, which respond to the load balancer in return. There is also a privilege of a backup load balancer in case all the servers are busy. In order to prevent a load balancer itself becoming a single point of failure, the implementation is done to provide for higher availability and replication of sessions is done. [2]

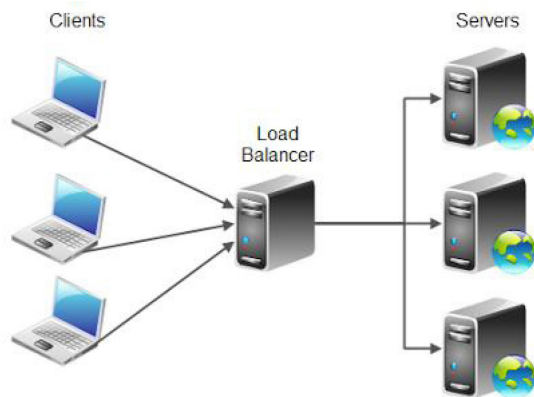


Fig 1: Load Balancing [1]

Benefits of Load Balancing: **Redundancy-**

It describes the process of running two or more, the same servers thus providing a guaranteed event that one server becomes occupied.

Scalability- Even though modest resources requirements are offered, scalability must always be considered for finding the correct host solution.

Resource Optimization- Through load balancing, one can optimize how traffic is circulated to the server cluster so, that it guarantees the best performance.

Security- In security, only one IP is exposed to the web with load balancing, which significantly reduces the amount of break points in case of attack. [1]

III. RELATED STUDY

Amandeep, Vandana Yadav, Faz [3] Mohammad Twenty first century is known as the era of technology. In the technology of computing it is an era of cloud computing. Cloud computing is the most recent technology in today's world of computing and its become more popular day by day due to its great feature (resource pooling, rapid elasticity, scalability, efficiency and on demand service). Cloud computing is built on the base of distributed computing, grid computing and virtualization. Cloud computing is defined as a large scale distributed computing paradigm that is driven by economics of scale in which a pool of abstracted virtualized dynamically.

Soumya Ray, Ajanta De Sarkar [4] The concept of Cloud computing has significantly changed the field of parallel and distributed computing systems today. Cloud computing enables a wide range of users to access distributed, scalable, virtualized hardware and/or software infrastructure over the Internet. Load balancing is a methodology to distribute workload across multiple computers, or other resources over the network links to achieve optimal resource utilization, maximize throughput, minimum response time, and avoid overload. With recent advent of technology, resource control or

load balancing in cloud computing is main challenging issue. A few existing scheduling algorithms can maintain load balancing and provide better strategies through efficient job scheduling and resource allocation techniques as well. In order to gain maximum profits with optimized load balancing algorithms, it is necessary to utilize resources efficiently. This paper presents a review of a few load balancing algorithms or technique in cloud computing. The objective of this paper is to identify qualitative components for simulation in cloud environment and then based on these components, execution analysis of load balancing algorithms are also presented.

Kousik Dasgupta, Brototi Mandal, Paramartha Dutta, Jyotsna Kumar Mondal, Santanu Dam

[5] The next-generation of cloud computing will thrive on how effectively the infrastructures are instantiated and available resources utilized dynamically. Load balancing which is one of the main challenges in Cloud computing, distributes the dynamic workload across multiple nodes to ensure that no single resource is either overwhelmed or underutilized. This can be considered as an optimization problem and a good load balancer should adapt its strategy to the changing environment and the types of tasks. This paper proposes a novel load balancing strategy using Genetic Algorithm. The algorithm thrives to balance the load of the cloud infrastructure while trying minimizing the make span of a given tasks set. The proposed load balancing strategy has been simulated using the Cloud Analyst simulator. Simulation results for a typical sample application shows that the proposed algorithm outperformed the existing approaches like First Come First Serve, Round Robing and a local search algorithm Stochastic Hill Climbing.

Vikas Kumar, Shiva Prakash [6] Cloud Computing refers to the use and access of multiple server based computational resources via a digital network(WAN).Cloud users may access the resources using computer note book, pad computer, smart phone, or other device. In cloud computing applications are provided and managed by the cloud server and data is also stored remotely in cloud configuration. As Cloud Computing is growing rapidly and clients are demanding more services and better results, load balancing for the Cloud has become a very interesting and important research area. Load balancing ensures that all the processor in the system or every node in the network does approximately the equal amount of work at any instant of time. Here in this paper we have discussed many different load balancing techniques used to solve the issue in cloud computing environment. This paper presents

various approaches given by the researchers using the load balancing techniques. **Harmandeep Singh Brar, Vivek Thapar, Kunal Kishor [2]** Cloud computing is overtaking the existing conventional methods of computation and communication over the network. The entire Internet community is often lured by a new paradigm that provides a great level of availability and security with nominal usage charges. Cloud computing, in this perspective, is an important way of disseminating information and providing computational capabilities over the network. The amount of data being stored and the services being provided are increasing at a very fast rate which, in turn, demands greater storage and computing hardware. With a huge number of requests in the form of load to the servers, load balancing becomes an important issue in cloud computing. The aim here is to distribute the load amongst the available nodes in such a way that no single node is flooded with requests, while some other node is lightly loaded. The prevalent scheduling algorithms have been addressing this issue by making use of job scheduling and resource provisioning strategies efficiently. This paper discusses the popular load balancing algorithms, along with the challenges faced. **D. Saranya, L.**

Sankara Maheswari [1] In the current scenario important focus is on the development load balancing algorithms for cloud computing. In this paper, different kinds of load balancing algorithms are analyzed for cloud computing, to make the network load equally divided in order to provide faster connectivity for all the devices that need the service. Among implemented load balancing algorithms, ant colony optimization algorithm has achieved a better performance, when compared to others. Load balancing algorithms play a challenging task in cloud computing.

IV. EXISTING TECHNIQUES

Following are the load balancing algorithms used in cloud computing:

Round Robin Algorithm (RR) [3] In this algorithm all the processes are divided between all processors. In this each process is assigned to the processor in a round robin order. The work load distributions between processors are equal. Different processes have not same job processing time. At many point of time some nodes may be heavily loaded and others remain idle In web servers where http requests are of similar nature and distributed equally then RR algorithm is used .When time quantum is very large then RR Scheduling Algorithm is same as the FCFS Scheduling and when time quantum is too small then Round Robin

Scheduling is known as Processor Sharing Algorithm.

Opportunistic Load Balancing Algorithm [1] It is also one of the static load balancing algorithms, which do not consider the present workload of the VM. It usually keeps each and every node busy. This deals with the unexecuted tasks quickly and in random order to the current node, where each one of task is assigned to the node randomly. This algorithm provides a load balancing schedule but does not produce a good result. The tasks are processed in a slow manner, where the current execution time of the node is not calculated.

Min-Min Algorithm [3] It starts with a set of all unassigned tasks. In this minimum completion time for all tasks is found. Then after that among these minimum times the minimum value is selected. Then task with minimum time schedule on machine. After that the execution time for all other tasks is updated on that machine then again the same procedure is followed until all the tasks are assigned on the resources. The main problem of this algorithm is has a starvation.

Max-Min Algorithm [3] Max-Min algorithm is almost same as the min-min algorithm. The main difference is following: In this algorithm first finding out minimum execution times, then the maximum value is selected which is the maximum time among all the tasks on any resources. After that maximum time finding, the task is assigned on the particular selected machine. Then the execution time for all tasks is updated on that machine, this is done by adding the execution time of the assigned task to the execution times of other tasks on that machine. Then all assigned task is removed from the list that executed by the system.

Active Clustering load balancing Algorithm [1] Active Clustering is an improved method of random sampling. The concept of clustering is used in this algorithm. The main principle of this algorithm is grouping similar nodes together, and working based on those grouped nodes. Grouping of nodes helps the resources to increase the throughput efficiently. In this algorithm, a method called match-maker is introduced. While an execution starts, the first node selects the neighbour node. The neighbour node is taken as match make node, which connects the neighbour node that is same as initial node. At last the match maker node gets disconnected. And this process is done iteratively to balance the load equally. The system performance is improved highly, by increasing the throughput. There is an efficient utilization of resources when there is an increase in throughput.

V. COMPARISON OF LOAD BALANCING ALGORITHMS IN CLOUD COMPUTING

Algorithm	Nature	Environment	Job Allocation	Advantage	Disadvantage
Round Robin	static	decentralized	Selects the First Node in Random Manner, and allocates other nodes in Round Robin Method.	Treats the entire server equally.	Any process is not known in advance.
Opportunistic Load Balancing Algorithm:	static	centralized	Based on the framework of the system.	Keeps every node busy.	The execution time is completed, but the node is still busy.
Min-Min	static	centralized	Identifies and completes the job waiting queue.	Performs better small execution time.	Leads to Starvation.
Max-Min	static	centralized	Finding the minimum execution time and deals with the maximum execution time.	Improves efficiency by increasing concurrent execution.	Execution that takes maximum time need to wait for long time.
Active Clustering load balancing Algorithm	dynamic	Decentralized	Grouping nodes together.	Similar nodes are grouped together.	The performance is poor when there is an increase in variety of nodes.

VI. SUMMARY

Cloud computing is a rising field of information technology (IT). It modifies a wide range of users to access distributed, scalable, virtualized, hardware or software base over the Internet. Load balancing is one of the leading issues of cloud computing. Load Balancing is an essential task in Cloud Computing environment to achieve maximum utilization of resources. Therefore, there is a demand for a regular load-balancing algorithm for efficient utilization of resources. We have given an overview of already proposed algorithms by several researchers in the literature.

VII. REFERENCES

- [1] D. Saranya, L. S. Maheswari. "Load Balancing Algorithms in Cloud Computing: A Review." International Journal of Advanced Research in Computer Science and software Engineering, Vol. 5, No. 7, pp. 1107-1111, July 2015.
- [2] H. S. Brar, V. Thapar, K. Kishor. "A survey of Load Balancing Algorithms in Cloud Computing ." International Journal of Computer Science Trends and Technology(IJCST), Vol. 2, No. 3, pp. 103-106, May-June 2014.
- [3] Amandeep, V. Yadav, F. Mohammad. "Different Strategies for Load Balancing in Cloud Computing Environment: a critical Study."

International Journal of Scientific Research Engineering & Technology, Vol. 3, No. 1, pp. 85-90, April 2014.

[4] S. Ray, A. D. Sarkar. "Execution Analysis of Load Balancing Algorithms in Cloud Computing Environment." International Journal on Cloud Computing Services and Architecture, Vol. 2, No. 5, pp. 1-13, October 2012.

[5] K. Das gupta. "A Genetic algorithm based Load Balancing Strategy for Cloud Computing." International Conference on Computational Intelligence: Modeling Techniques and Applications. 2013.

[6] V. Kumar, S. Prakash "A Load Balancing Based Cloud Computing Techniques and Challenges." International Journal of scientific research and management, 2014.