Improved Cloud Computing Using ACO Algorithm with Load Balancing

Mr. Bagul Navin
Cloud Computing Researcher
Silver Oak College of Engineering & Technology
Ahmedabad, India

Mr. Darshil Shah
Assistant Professor
Silver Oak College of Engineering & Technology
Ahmedabad, India

***Abstract*—Cloud computing has opened a new horizon for utilization of resources and their computing power. In cloud computing, is prone to many adversaries such as server crashes, network congestions, low bandwidth etc. Load balancing is one of the main challenges in cloud computing, which is required to distribute the workload equally across all the nodes, detect the fault and remove it and improve the quality of service in Cloud computing which in turn improves the fault tolerance. Here we have proposed an optimized approach of load balancing schemes, using Ant colony optimization algorithm.**

Keywords— Cloud Computing, Load Balancing, ACO

I. INTRODUCTION

Cloud computing is creating as another standard for conveying, arranging, and getting to substantial scale distributed computing applications. With the enormous development of internet-web and its users, Cloud Computing, with its incredible conceivable outcomes in simplicity, Quality of administration and on-intrigue organizations has transformed into an ensuring figuring stage for both business and non-business calculation clients. It is an adoptable innovation as it gives mix of programming and assets which are powerfully adaptable. The dynamic condition of cloud results in different startling deficiencies and disappointments. Distributed computing stores the information and conveyed assets in the open condition. The measure of information stockpiling increments rapidly in open condition. Thus, stack adjusting is a primary test in cloud environment.

II. CLOUD COMPUTING

Distributed computing is show for empowering helpful, on-demand, organize access to shared pool of configurable registering assets (e.g., systems, storage, server, application and administrations) that can quickly provisioned and discharged with insignificant administration exertion or specialist organization cooperation. This cloud model is having five vital qualities, three administration model and four organization models [1].

Cloud computing it's situated in the offer of administrations, we discovered 3 various types of services.

1) Software as a Service (SaaS): These types of services are basically applications provided over internet. Users can access these applications using Web Browser. The end user is completely unaware about the hardware and software that is deployed to provide the service and is simply using interface to access the application. An example of this kind of service is Google Docs.

2) Platform as a Service (PaaS): These services are centered on the idea of deploying the applications or services online along with the hardware and the software pre-requisites containing the solution stack as well. The complete life-cycle of the application or service is covered in this type of service such as design, implementation, testing, deployment, integration with existing architecture etc.

Major characteristics of this type of service are.

- Deployment services
- QA and maintenance
- Scalability
- Collaboration tools.

An example of this type of service is Google App Engine.

3) Infrastructure as a Service (IaaS): These services focus on providing the complete IT needs of an organization including the servers, networks etc along with the necessary software to power it. The end user views the entire infrastructure as a part of the organization itself.

A. Types Of Cloud

1) Public: These are the conventional types of clouds also known as the external cloud, where the service is provided by the internet from a different entity and is visible to everyone. This type of cloud is used by majority of the users. However, the users are not able to access the information of other users on the same cloud.

2) Private: These are more expensive compared to the public cloud platforms. These types of clouds consist of privately owned applications and services. Compared to fully owing the hardware and the software, this type of cloud is much more cost effective and is more scalable.

3) Hybrid: These types of clouds are a mixture of private and public clouds, optimized in such a manner wherein the organization may cut down on investing heavily on the infrastructure and utilizing the public cloud during the peak hours. It is a good option where not much investment is to be done on the infrastructure.

4) Community: This cloud infrastructure is shared by several organizations.

B. Feature of Cloud Computing

1) Flexibility: Cloud computing provides a high rate of flexibility to the client. The client can access their application and data from anywhere on any system in the world.

2) Low Cost: Cloud is offered to the clients at a very low rate, unlike traditional desktop software which results in a decrease in operating cost of IT companies.

3) Better storage capacity: The client can save more data in cloud computing environment than a personal computer. Because of that client need not have to upgrade their computer's memory which reduces the cost for companies and user.

4) Flexible Compatibility: Services of cloud computing can be used using different electronic devices having internet access (ex. Laptop, desktop, mobile phone).

5) Automatic Updating: Cloud computing services enables users to download updates for software, once the server is updated client do not have to do anything which saves company's time and effort.

6) Easier organization of Data and Information: Because of centralized storage of data it is easy to maintain them.

7) Backup and Recovery: Recovery of information is done very efficiently so that taking a backup and restoring data is very easier in the cloud.

C. Load Balancing

Cloud load balancing is the way toward distributing workloads and figuring assets in a cloud computing condition. It is a process of bifurcating workloads and resources in a cloud computing environment. It allows organizations to handle their applications and workload demands amongst its servers and other back-end devices. Cloud load balancing caters to distribution of workload traffic and the surges demands that occurs in the internet. [1].

D. Load Balancing Schemes

1) Ant Colony Optimization (ACO): It is an Improved variant of load adjusting system in light of Ant Colony and Complex Network Theory (ACCLB) in an open distributed computing alliance. Both calculations make utilization of ants' pheromone to gather and update information about the cloud in this manner choosing a particular hub with a specific node in order to assign the task however evenly distributing work among nodes. The ants in proposed calculation persistently start from the Head hub and navigate all around the system making forward and in reverse development to locate the under stacked and over-burden hubs. In ACO two sorts of pheromones are utilized Foraging Pheromone (FP) used to investigate over-burden hub by forward development of ants while Trailing Pheromone (TP) used to find its way back to the under stacked hub. Keeping in mind the end goal to restrict the quantity of ants in the system, they would confer suicide once it finds the objective hub [2].

2) ESWLC (Exponential Smooth Forecast based on Weighted Least Connection): Improved form of Weighted Least-Connection (WLC) alongside its elements, it likewise consider time arrangement and trials. However WLC numbers the associations of every server and reports the fitting server in view of the duplication of a server weight and its tally of associations, ESWLC calculation finishes up doling out a specific errand to a hub simply subsequent to becoming more acquainted with about the hub capacities. ESWLC fabricates the choice in view of the experience of the hub's CPU control, memory, number of associations and the measure of plate space right now being utilized. ESWLC then predicts which hub is to be chosen in view of exponential smoothing [3].

3) Map Reduce: In Map Reduce Fault Tolerance, the ace initially endeavors to allot a guide errand (in the line) whose information is on that machine (information territory) gave that the machine is allowed to handle the demand. If there should be an occurrence of disappointment in the execution, it endeavors to allocate a guide assignment whose information is situated (on a machine) on a similar system switch with that machine (rack area). Accordingly, on event of disappointment finish outline should be re-executed, yet finished diminish errands does not. To guarantee that a fizzled occupation can be recuperated and is being planned with an ensured era, a limit esteem is utilized whereby past it, the fizzled employment will be booked on the following accessible machine regardless of information area.

4) Virtual Machine Mapping (VM Mapping): Depends on multi-dimensional assets to accomplish general load adjust. This calculation embraces the unified control engineering contains planning controller and asset screen as center components of the framework. The planning controller is in charge of VM lifecycle administration and satisfying portion arrangement while the asset screen gathers the data about assets from physical hosts .The procedures required in VM mapping approach experiences taking after four stages: right off the bat; tolerating the demand for virtual machine on FCFS standard, also; getting asset data which thus is keep up by asset screen, thirdly; VM starting position by booking controller, at long last; client can remotely get to the application on cloud.

5) Dual Direction FTP (DDFTP): It is a double heading downloading calculation from FTP servers and its adjusted variant is in , presenting productive adaptation to non-critical failure and load adjusting with insignificant correspondence and coordination overhead while executing administrations in parallel over shared and element heterogeneous dispersed cloud foundation. The primary thought of this calculation is to parts the document into two half and assignment is being executed on two servers, with the end goal that every server begins handling the undertaking in an inverse bearing from the other, one server begins preparing from the earliest starting point in an incremental request while different begins the record downloading from the last square in decrement arrange. The errand is thought to be done when the two servers download two continuous squares meeting at assent. Accordingly, both servers will work autonomously, however will wind up downloading the entire record to the customer in the most ideal time given the execution and properties of both servers. In addition, properties, for example, organize stack, hub stack, arrange speed are naturally thought about, while no run-time observing of the hubs is required, yet it keep up great load adjusting among all partaking server. Also, in the event

that one of the servers flops before finish of errand, second one proceeds with the assignment till it spans to the point where alternate gets fall flat [4].

6) Fault Tolerance Policy on Dynamic Load Balancing (FTDLB): It is an adaptation to internal failure approach that could endure the hub's perpetual disappointments while adjusting heap of constant applications on P2P frameworks. For enhancing the framework unwavering quality, FTDLB copies occupations into various destinations and adaptively alter the heap of continuous applications to accomplish the employment's negligible turnaround time. FTDLB calculation functions as takes after; each site routinely sends "pulse" messages to its neighbor site which incorporates the CPU use, memory utilization, work status, and so forth. When getting of "pulse" messages stops, inside a settled period, it demonstrates disappointment of neighbor site subsequently activating adaptation to non-critical failure arrangement [4].

7) O-Ring (Overlapped Ring): It is a novel design that gives adaptation to internal failure and load adjusting for circulated and element situation. O-Ring utilize the approach of information replication (reflecting) and information dissemination so as to give both adaptation to non-critical failure and load adjusting in efficient way. In the underlying stage information things are reproduced on the neighboring companions on the ring with a specific end goal to accomplish adaptation to internal failure and each associate likewise stores the address of its ancestor and successor. Each ring had a Directory Service which is in charge of steering of solicitations like; information recovery, updates, inclusions and erasures on proper companions. As duplicate of information is as of now being imitated for reinforcement on another associate here and now changes are tended to by moving the limits of duty between companions without the need to move the information itself. In this manner, redistributing the heap in forward and in reverse bearing to adjust the heap speedier, and limiting meddles with simultaneous inquiry preparing. Any sorts of vacillations, that require the development of information, are tended to by moving the reinforcement duplicates of the information out of sight, without exasperating the essential duplicate of the information that is being utilized to deal with solicitations for the information. Alongside more affordable load adjusting of O-Ring likewise accomplishes higher throughput, as it can adjust the heap with lower overheads and can react quickly to load lopsided characteristics. [5].

8) Honey Bee Behavior inspired Load Balancing [HBBLB]: It is procedure which accomplishes even load adjusting crosswise over virtual machine to expand throughput. It considers the need of assignment sitting tight in line for execution in virtual machines. After that work stack on VM computed chooses whether the framework is over-burden, under stacked or adjusted. What's more, in light of this VMs are gathered. New as per load on VM the assignment is booked on VMs. Undertaking which is expelled before. To locate the right low stacked VM for current errand, assignments which are expelled before from over stacked VM are useful. Forager honey bee is utilized as a Scout honey bee in the following strides [5].

E. Cloud Data Center

A data center (usually called a server farm) is a incorporated archive for the capacity, administration, furthermore, dissemination of data and information. Normally, a data center is an office used to house PC frameworks and related segments, for example, media communications furthermore, storage systems. Sometimes there are backup power supplies, redundant data communications connections, environmental controls, and security devices. One benefit to the data center is that physical hard drive storage resources are aggregated into storage pools, from which “logical storage” is created. The heterogeneous way of most storage system permits numerous different vendor's storage equipment to be added to the framework with practically zero detectable impact.

III. BACKGROUNG THEORY

1) Cloud Computing Model And Its Load Balancing Algortihms: In this paper, it describe in depth about cloud computing model, and several features of it. Further it describes cloud computing service models and deployment models of cloud computing. In paper we focus on survey and comparison of different load balancing algorithm : Round Robin algorithm, Honey bee faraging, Min-Min algorithm, Max Min algorithm, OLB, Biased Random Sampling, Ant Colony Optimization, Two phase scheduling load balancing Algorithm. In this paper cloud computing technology is explained which is the most trending technology these days, cloud computing model provides different services to the remote users, and services can be software or any platform or any infrastructure [6].

2) Load Balancing Algorithms in Cloud Computing: A Survey of Modern Techniques: In this paper we provide a comprehensive overview of interactive load balancing algorithms in cloud computing. Each algorithm addresses different problems from different aspects and provides different solutions. Some limitations of existing algorithms are performance issue, larger processing time, starvation and limited to the environment where load variations are few etc. A good load balancing algorithm should avoid the over loading of one node. The aim is to evaluate the performance of the cloud computing load balancing algorithms that have been developed over the period of 2004-2015. we have presented comparison of different load balancing algorithms for cloud computing such as, round robin (RR), Min-Min, Max-Min, Ant colony, Carton, Honey bee etc. We described advantages and limitations for these algorithms showing results in different conditions. The vital part of this paper is comparison of different algorithms considering the characteristics like fairness, throughput, fault tolerance, overhead, performance, and response time and resource utilization [7].

3) Intelligent Cloud Algorithms for Load Balancing problems: A Survey: In this paper, it have survey on the different types algorithm for load balancing. This paper surveys the above intelligent optimization techniques and focus on the Ant Lion Optimizer intelligent technique, also it proposes an implementation of ALO based cloud computing environment as efficient algorithms that expected to supplies

better outcomes in load balancing . Also compare the different methods for load balancing. It gives concise explanation of frequently used algorithms, such as GA, ACO, ABC and PSO [8].

4) Cloud Load balancing based on Ant Colony Optimization Algorithm: In this paper, its approach for improving parameters in Ant colonies functions. In this paper they carry out the new load balancing that depending on the past load balancing , its very helpful in cloud environment. All practical implementation is done on Cloudsim . Result of the simulation shows that response time is reduced and throughput increased compare with the basic ACO [9].

5) A Hybrid Approach of Load Balancing through VMs using ACO, MinMax and Genetic Algorithm: In this paper, create a hybrid algorithm for the VM allocation utilizing the best features of Ant Colony Optimization, followed by Max-Min Ant System and last not least Genetic Algorithm. Paper is split into three sections. The first section creates a random environment in which the user provides the number of Virtual Machine and would like to configure and the Primary Machine count is taken in a contrast. The second algorithm generates an initial allocation probability based on the properties of the host and the VM machines. And third ACO algorithm applied on it and pheromone situation generated. The work mainly focuses on Load Balancing through VM migrations in Cloud Computing. Ant Colony optimization, Max min ant system and Genetic algorithm are combined for balancing the load in data centers [10].

6) Performed Load Balancing Algorithm for Public Cloud Computing Using Ant Colony Optimization: In the presented paper, various research are studied to attain load balancing in cloud computing. Further to this, an improvement is proposed for the load balancing architecture and algorithm of cloud computing which focuses on attaining better response time. It also consists of architecture for load balancing by offering to share the functionalities. The main control node represents the entry point to cloud into two parts. The main controller is assigned the task to bifurcate the user tasks through different regional load balancers. The secondary controller will update the state of the system through various mediums. We can also use a combination of algorithms in order to optimize the bifurcation and mapping of the steps. In conclusion, the proposed architecture is segregated into three levels, each level assigned a unique role and algorithm [11].

IV. PROPOSED METHOD

The proposed method of load balancing depends on Ant Colony Optimization (ACO) idea. ACO is inspired from the subterranean insect states that cooperate in searching conduct. Indeed the genuine ants have motivated numerous scientists for their work and the ants approach has been utilized by numerous specialists for critical thinking in different zones. This approach is approached the name of its motivation ACO. The ants cooperate looking for new source of food and all the while utilize the current sustenance sources to move the food back to the home. The ants leave a pheromone trail after moving starting with one node then onto the next. By taking after the pheromone trails, the ants along these lines spans to the food sources. The force of the pheromone can change on different elements like the quality of food sources, distance of the food source, etc. The ants utilize these pheromone trails to choose the following node. A Data Center server is known as nodes in the proposed framework.

A. *Pheromone Prevalence*

The ant uses two types of pheromones to signify its movement. Depending upon the type of pheromone being updated, ant would tell which type of node it is searching for (i.e. from the under-loaded node to overloaded node and vice-versa). The two types of pheromones updated by ant are mentioned below.

a) Foraging Ant Pheromone type (FP): While ants moving from under loaded node in server to overloaded node, ant will update FP. Equation for updating FP pheromone is.

$$FP(t+1) = (1 - \beta eva)FP(t) + \Delta FP \quad (1)$$

Where,

βeva = Evaporation rate of Pheromone

FP = Foraging pheromone of the edge before the move

$FP(t+1)$ = Foraging pheromone of the edge after the move

ΔFP = Change in FP

b) Trailing Ant Pheromone (TP): While ant moving from overloaded node server to underloaded node, ant will update Trailing Pheromone(TP). Equation for updating TP pheromone is

$$TP(t+1) = (1 - \beta eva)TP(t) + \Delta TP \quad (2)$$

Where,

βeva = Evaporation rate of Pheromone

TP = Trailing pheromone of the edge before the move

$TP(t+1)$ = Trailing pheromone of the edge after the move

ΔTP = Change in TP

B. *Redistribution Policy*

Redistribute the request among nodes is the main task of ant in algorithm. The purpose of this policy is used to calculate that how many request send or given to each server or node. Following example is useful to understand this redistribution policy

Suppose, there are three nodes N1, N2 and N3 with load of 22, 18 and 27 respectively. Now, add the load of N1, N2 and N3, A =22 +18 + 27 = 67. Now, divide A by 3, B = 67/3 = 22.33 i.e. 22. Now N1, N2 and N3 equally gets load of 22.3

V. SIMULATION AND RESULT

Simulation is done using Eclipse. There are 36 simulated nodes in Data Center. Head node with ID 15 is fixed because in the simulation of network of servers it has maximum number of neighbors. Suppose, there are 15000 requests are distributed on nodes and calculated threshold is 417. Threshold value depends on number of requests and availability of resource. Ant Timer is for 1000ms. Number of ants i.e. threads in system is 8.

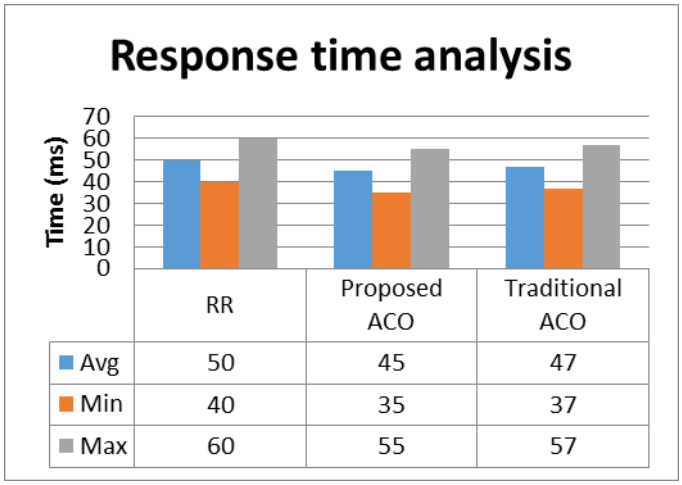

Fig. 1. Simulation results output: Response Time Analysis.

CONCLUSION

In Cloud computing, Main issue is Load balancing with fault tolerance which needs to improve its performance. In this paper, we have given comparative study of all algorithm and parameters of Load Balancing algorithm like round robin (RR), Min-Min, Max-Min, Ant colony, Carton, Honey bee etc.

All algorithms are good in one or in other way but none of them achieve 100% efficiency but none of them gives very less execution time. The propose of load balancing technique gives based on Ant Colony Optimization (ACO) gives optimal resource utilization and cost optimization.

The execution of the framework is upgraded with high accessibility of assets, subsequently expanding the throughput. Nodes of the system work equally and data transferred without effecting on the performance of the system, although it gives optimum throughput.

REFERENCES

[1] Kavyasri M N, Dr. B. Ramesh, "Comparative Study of Scheduling Algorithms to Enhance the Performance of virtual machines in Cloud Computing " 978-1-4673-6725-7/16/$31.00 ©2016 IEEE

[2] Shubham Mittal, Avita Katal, "An Optimized Task Scheduling Algorithm in Cloud Computing " 978-1-4673-8286-1/16 $31.00 © 2016 IEEE

[3] Arnav Wadhonkar and Deepti Thang, "A Survey on different Scheduling Algorithms in Cloud Computing" 978-1-4673-9745-2 ©2016 IEEE.

[4] Ruonam Lin, Qiang Li, "Task Scheduling Algorithm Based on Pre-Allocation Strategy in Cloud Computing" 978-1-5090- 2594-7116/$31.00 ©20 16 IEEE.

[5] Abdul Razaque, Nikhileshwara Reddy Vennapusa, Nisargkumar Soni, Guna Sree Janapati, khilesh Reddy Vangala, "Task Scheduling in Cloud Computing"

[6] Shaowei Lin, Kaijun Ren, Kefeng Deng and Junqiang Song, "A Dynamic Resource Allocation and Task Scheduling Strategy with Uncertain Task Runtime on IaaS Clouds" 978-1-5090-1224-4/16$31.00 ©20 16 IEEE

[7] Bingxu Wang, Dan Tao and Zhaowen Lin, "A load feedback based resource scheduling algorithm for IaaS cloud platform" 978-1-5090-2073-7/16/$31.00©20 16 IEEE

[8] Anita kumari, Sushma jain, "Auction based resource allocation strategy for insrastructure as a service" 978-1-5090-0082-1/16/$31.00©20 16 IEEE

[9] Li-der chou, Hui-fan chen, Fan-hsun Tseng, Han-Chieh Chao, Yao-Jen Chang, "DPRA: Dynamic power-saving resource allocation for cloud datacenter using partocle swarm optimization" 1937-9234 ©20 16 IEEE

[10] Anurag Jain, Rajneesh Kumar, "A comparative analysis of task scheduling approaches for cloud environment" 978-9-3805-4421-2/16/$31.00©20 16 IEEE

[11] Victor Pelaez, Antonio Campos, Daniel F. Garcia, Joaquin Entrialgo, "Automatic scheduling of deadline-constrained bag of tasks in hybrid clouds"

[12] Sowndarya Sundar and Bang Liang, "Communication augmented latest possible scheduling for cloud computing with delay constraint & task dependency" 978-1-4673-9955-5/16/$31.00©20 16 IEEE.