Global Journals LaTeX JournalKaleidoscope™
Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

Survey on Load Balancing in Cloud Computing

R.Gowri Prakash[1], R. Shankar[2] and S. Duraisamy[3]

[1] Bharathiar University

Received: 13 December 2018 Accepted: 1 January 2019 Published: 15 January 2019

Abstract

Cloud computing is a biggest technology. It is different cloud services are SaaS, PaaS, aaS. loud users are able to access software and applications from whenever they need while it is being host by an outside party in cloud. Cloud services are bought on a pay-as-user-go or subscription basis. The cloud users getting into increased day by day. In this paper mainly focused on Load balancing within the cloud computing surroundings has a crucial impact on the performance. Smart load balancing makes cloud computing additional economical and improves user satisfaction. Load balancing may be a pc networking technique to distribute work load across multiple computer or a multiple cluster,

Index terms— cloud computing, load balancing, virtual machine, cloudsim

1 Introduction

Fig. **??**: Cloud computing deployment models loud Computing enables cloud customers to enjoy the on-demand high quality applications and services from a centralized pool of configurable computing resources. This new computing model can relieve the difficult of storage management, allow universal data access with independent geographical locations, and avoid capital expenditure on hardware, software, and personnel maintenances, etc. As cloud computing becomes mature, lots of sensitive data is considered to be centralized into the cloud servers, e.g. Personal health records, secret enterprise data, government documents, etc. Cloud Computing has attracted the enormous companies like Google, Microsoft, and Amazon and regarded as a great influence in today's data Technology industry. Business owners are attracted to cloud computing concept because of several features .Cloud computing refers to applications and services that run on a distributed network using virtualized resources and accessed by common internet protocols and networking standards. It is distinguished by the notion that resources square measure virtual and limitless which details of the physical systems on that C code runs square measure abstracted from the user. cloud computing using deployment models tells you where the cloud is located and for what purpose. Public, private, community, and hybrid clouds square measure readying models. Service models describe the sort of service that the service supplier is providing. The bestknown service models are Software as a Service, Platform as a Service, and Infrastructure as a Servicethe SPI model. The service models build on one another and define what a vendor must manage and client's responsibility is.

2 a) Public cloud

A public cloud is one based on the standard cloud computing model, in which a service provider makes resources, such as virtual machines (VMs), applications or storage, available to the general public over the internet. Public cloud services may be free or offered on a pay-per-usage model **??**1].

3 b) Private cloud

The private cloud infrastructure is operated for the exclusive use of and organization. The cloud may be managed by that organization or a third party. Private clouds maybe either on-or off-premises.[1]

4 c) Hybrid cloud

A hybrid cloud combines multiple clouds (private, community of public) where those clouds retain their unique identities, but are bound together as a unit. A hybrid cloud may offer standardized or proprietary access to data and applications, as well as application portability[3].

5 d) Community cloud

A community cloud in computing is a collaborative effort in which infrastructure is shared between several organizations from a specific community with common concerns (security, compliance, jurisdiction, etc.), whether managed internally or by a third-party and hosted internally or externally **??**2].

6 II.

7 Services of Cloud Computing

As acloud computing has developed different vendors offer cloud that have different services associated with them.

Software as a Service (Saas): this is the top most layer of the cloud computing stack directly consumed by end user .i.e Saas. ? Next generation Saas promises every things as a services over the internet.

? Cloud computing started with similar premises. ? A computing paradigm where them exists a fixable set of computing resources across the internet.

8 [3]

Platform as a Service (PaaS): This provides a service to the user for the layer of package platform. It provides a storage device for the mixed applications and expenditure. User will have associate degree freedom to create their individual applications that gives communications for the user. It presents predefined parts of united OS and also the application server, e.g. LAMP platforms.

9 Infrastructure as a Service (IaaS):

This provides a service to the user for the essential storage and processor infrastructure as a service over the network. For this service user have to be compelled to pay charges, once they use this service over network. During this process the cloud computing give a service over the web, hardware and package in datacenters as a service. The datacenter of hardware and package is termed as Cloud [3].

10 Benefits of Cloud Computing

The ultimate aim of cloud computing benefits are [4]:-Flexibility: There is high rate flexibility.

Speed & Scales: traditional methods to buy and configure hardware and software are time consuming

Easier management of data and information: Sine all data are located on centralized location, data are more or organized making it easy to managed.

Device diversity: we can access our application and anywhere in the world on any system.

Increased storage capacity: increased storage capacity is another benefits cloud computing. as it can store more data as compare to a personal computer.

Easy to learn and understand: since people are quite used to cloud application like Gmail, Google docs so anything related to same is most likely to be understood by user.

Automatic updating: Its saves companies time and effort to update multiple servers.

Risk: Cloud computing services means taking services from remote servers.

Security: security and privacy are the biggest concerns about cloud computing.

Migration Issue: migration problem is also a big concern about cloud computing.

11 IV.

12 Need for Load Balancing

The fundamental point of load adjusting is to appropriate the traffic among the hub similarly in the group to improve things execution of system.

The point of load adjusting is as per the following: 1. To improve the surety of administrations to the purchaser. 2. To improve the client fulfillment. 3. To expand use of asset. 4. To diminish the execution time and holding up time of undertaking originating from various area [8]. 5. To make administration execution better. 6. Keep up bunch steadiness. 7. Assemble a framework that can endure the shortcomings. 8. Accommodate future adjustment. V. Load Balancing in Cloud Computing

Load balancing improve the distribution of workload across multiple computing resource for a cluster, network links, control processing unit, disk drive.

Load balancing is a technique used to distribute workloads uniformly across servers or other compute resources to optimize network efficiency, reliability and capacity. Load balancing is performed by an appliance –either

physical or virtual –that identifies in real time which server in a pool can best meet a given client request, while ensuring heavy network traffic doesn't unduly overwhelm a single server.

In addition to maximizing network capacity and performance, load balancing provides failover. If one server fails, a load balancer immediately redirects its workloads to a backup server, thus mitigating the impact on end users.

Load balancing is usually categorized as supporting either Layer 4 or Layer 7. Layer 4 load balancers distribute traffic based on transport data, such as IP addresses and Transmission Control Protocol (TCP) port numbers. Layer 7 load-balancing devices make routing decisions based on applicationlevel characteristics that include HTTP header information and the actual contents of the message, such as URLs and cookies. Layer 7 load balancers are more common, but Layer 4 load balancers remain popular, particularly in edge deployments.

13 Goals of Load Balancing

? Archive optimal resource utilization ? Maximize throughput ? Minimize response time ? Avoid overload ? Avoid crashing One of the most normally used applications of load balancing is to be produced one web services from multiple server, generally refer to as server farm. Normally load balanced system widespread website, massive web relay chat network, high band with transfer Protocol sites, Network News Transfer Protocol (NNTP) servers, name System (DNS) servers, and databases[2]. By leveling application request across multiple servers, a load balancer reduces individual server load and prevent any one of application server from changing into one purpose of failure therefore rising overall application accessibility and responsiveness. The load balancing algorithms can be categorized mainly into two groups asdiscussed in the following section.

14 a) Static Algorithms

Static algorithms divide the traffic equivalently between Servers. By this approach the traffic on the servers will be disdained easily and it will make the Situation more imperfect. This algorithm, which divides the traffic equally, is announced as round robin algorithm. However, there were lots of problems associated with this algorithm. Therefore, weighted round robin was developing to improve the critical issues of round robin. In weighted round robin algorithm each servers is assigned a weight and according to the highest weight they receive more connections. In a situation, when all the weights become equal, servers will receive balanced traffic [5].

15 b) Dynamic Algorithms

Dynamic algorithms designate proper weights on servers dynamically by searching the whole network. The lightest server is loaded to balance the traffic. However, selecting an appropriate server needs real time communication with the networks, which leads to extra traffic added to the system. Dynamic algorithm predicates on query that are made frequently on servers. However, sometimes prevailed traffic prevents these queries to be answered, and correspondingly more added overhead can be distinguished on network [5].

16 VI. Load Balancing Algorithm in Cloud Computing

There are various load balancing algorithms used in cloud computing. In this study of this paper following different algorithms have been studied. The following are:

Round Robin: Round Robin is simplest algorithm that use the time slicing mechanism. The one of static load balancing algorithm. round robin algorithm is job allocation method. Here time is divided into several sectors and each node is given a specific time quantum interval and quantum node operation [6] in scheduling a time a time quantum plays a very important role, because if the time slice is very large, hen the round robin scheduling algorithm is the same as the FCFS planning **??**6].

The disadvantages of the method is that although the algorithm, is very simple. but it determine the quantum size, it generates an additional load scheduler. it has higher context switches that increase the turn round time, and low throughput.

First Come First Serve: First come, first served (FCFS) is a working framework procedure planning calculation and a system direction-finding the board instrument that naturally executes lined demands and procedures by the request of their entry. With first come, first served, what starts things out is taken care of first; the following solicitation in line will be executed once the one preceding it is finished. FCFS for parallel handling and is going for the Asset with the littlest holding up line time and is chosen for the approaching job. The Cloud Sim toolbox underpins First Come First Serve (FCFS) planning procedure for interior planning of occupations. Assignment of application-explicit VMs to Hosts in a Cloud-based server farm is the obligation of the virtual machine provisioned part. The default approach Serve (FCFS) premise. The Inconveniences of FCFS is that it is non preemptive..Its turnaround and reaction is very low [7].

Global scheduling algorithm: VM resource scheduling in cloud computing environment in this main advantages solution in which on the system after arrangement . in this way is load balancing and reduces or avoids dynamic migration hence resolve the problem of load balancing and high migration cost caused by traditional scheduling algorithm. therefore a monitoring and analyzing mechanism is needed to better solve the problem of load balancing. This is also a further research subject [11].

Min-Min Load balancing algorithm: Min-Min is a static load adjusting calculation, where the parameters related to the activity are perceived ahead of time. Min_Min algorithm, the execution and completion time of the unassigned waiting in queue are identified by the cloud manager. The jobs with minimum execution in time are being assigned first to the processors, so that the task is completed in time. But the tasks with maximum execution need to wait for a specific period of time. As such, all the all the tasks in the processor must be updated and the tasks in the queue must be removed. The task with minimum time execution performs better than the maximum time execution. The main disadvantage of this algorithm is that it leads to starvation. The terminology related to static load balancing for Min-Min is [6]. Opportunistic Load balancing algorithm: it is also one of the static load balancing algorithm. do not consider the present workload in virtual machine. Usually keeps each and every node busy. In this algorithm in with unexecuted task is quickly and in random order to the current node. Each one task is assigned to the node randomly. Disadvantages of the method is that although the algorithm provides a load balancing schedule bus dose not produce in good result. The task in a slow manner. were the current execution time of the node is not calculated [8].

Generalized Priority Algorithm: Client characterize the need as per the client request you need to characterize the parameter of cloudlet like size, memory, data transfer scheduling policy and so on. In the proposed technique, the task are first organized by their size to such a that one having most highest size has highest position.

The key factor for prioritizing task is their size and for Vmis their MIPS. This task performing better than foe FCFS and Round Robin algorithm. The virtual machine is prioritized according to their MIPS values such that one having highest MIPS has highest rank [7].

Weighted Round Robin Algorithm: weighted round robin Algorithm consider the resource capabilities of the VMs and assigns higher number of task to the higher capacity VMs based on the weight age given to each of the VMs. But it failed to consider the length of the task to select the appropriate [9].

17 Weighted Active Monitoring Load Balancer (AMLB):

The active monitoring load balancer maintain information about every VMs and number of requests currently allocated to which virtual machine. When a allocate new VM arrives it identifies the least loaded VM. If there are more then one, the first identified is selected. Active VM Load balancer return the VM id to the data center controller. The data controller sends the request to virtual machine identified id. Data center controller notifies the new allocation and cloudlet is sent to it. The proposed algorithm to implemented in cloud computing using cloudsim toolkit. In java language. these VMs of different processing tasks and request assigned virtual machine to optimized performance parameters such as response time and data processing time giving an load balancing algorithm in cloud environment [10]. process queue. It needs to maintain a separate queue for each and every node. Depending on the priority, the task is taken into concern, by removing the task that is waiting in the overloaded machine. The tasks removed are loaded into lightly loaded machine. Those tasks are known as scout bee for the next step [8]. The behavior of honey bee in load balancing has stimulated to reduce the response time of virtual machine, which also reduces the waiting time. The main disadvantage of this algorithm is, it does not show any improvement in throughput [8].

18 Biased Random Sampling Load Balancing Algorithm:

Biased Random Sampling is a dynamic load balancing algorithm. Here, random sampling method is being used to achieve the load balancing across all the nodes. In this algorithm, all the servers are treated as nodes. This method is represented in the form of virtual graph, constructed with the connectivity which represents the load on each node. Each node is taken as vertex in a directed graph. When a request is received from the client to the load balancer, the load balancer assigns the job to the node that has a minimum of one indegree. Once a job is assigned to the node, the server starts executing the job, indicating the reduction in availability of free resources. After the completion of the job, the node gets incremented by one in-degree, indicating the increase in available resources. The addition and deletion of such processes are completed by the process of random sampling technique. Threshold value is used as a parameter that considers each and every process by representing the maximum walk length. The traversal is from one node to another node until finding a designation is known as a walk. After receiving the request from the load balancer, it compares the current node to the randomly selected node with the threshold value. If the threshold value is equal or greater than the current walk length, the node executes its job, or else it moves to another neighbor node that is randomly selected. The performance decreases as the number of servers increases [13].

19 Ant Colony Optimization Based Load Balancing

Algorithm: This algorithm is designed to seek out the optimal path among the food and colony of ant, based on its actions. The main aim of this approach is to distribute the work load among the nodes in an efficient manner. The regional load balancing node is preferred as head node in Cloud Computing Service Provider. As the request is being sent, the ant starts is first movement from the head node. The ants collect the information from the cloud node and assign the tasks to the particular node. Once the task is assigned to the head node, the ant moves in a forward direction with the overloaded node to the next node checking whether the node is overloaded or not. During the movement, if it finds any loaded node again it moves in a forward direction, else it finds the overloaded

node, it moves in backward direction and replaces were the node found before .Once the job gets successful it is updated, then the result is reported based on the individual result of the ant. After receiving the individual result they are combined together to build the complete report. The solution set is updated automatically, when the ant updates the result for every movement. To prevent backward movement, the ant commits suicide when it reaches the target node [14].

20 Throttled Load balancing Algorithm (TLB)

This algorithm load balancer maintains tables of virtual machine indexes as available are running state (buys). The client server first to make request the data center to find a suitable virtual machine perform the recommended tasks. The data center request a load balancer to distribute virtual machine load balancer scans the index table from available virtual machine index table is scan completed. VMs is found the data center passes request to virtual machine identified in a identifier.

21 Fig. 5: Throttled Algorithm

Data center confirms the load balancing of the new distribution and data center appropriately revises the index table. When processing a client request, if the corresponding virtual machines not found load balancer returns-1 to the data center. The data center request is processed by the data center [14].

22 Ant Colony Optimization Based Load Balancing

Algorithm: This algorithm is designed to seek out the optimal path among the food and colony of ant, based on its actions. The main aim of this approach is to distribute the work load among the nodes in an efficient manner. The regional load balancing node is preferred quantity of available honey. After that, the reapers gather the honey from the sources. Then, again they go for the waggle dance to specify the honey that is left. In load balancing, the servers are combined together as virtual servers, where each and every virtual server has a finds any loaded node again it moves in a forward direction, else it finds the overloaded node, it moves in backward direction and replaces were the node found before **??**6].Once the job gets successful it is updated, then the result is reported based on the individual result of the ant. After receiving the individual result they are combined together to build the complete report. The solution set is updated when the ant updates the result for every movement. To prevent backward movement, the ant commits suicide when it reaches the target node [13].

Active Clustering load balancing Algorithm: 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 [7].While an execution starts, the first node selects the neighbor node. The neighbor node is taken as match make node, which connects the neighbor 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.

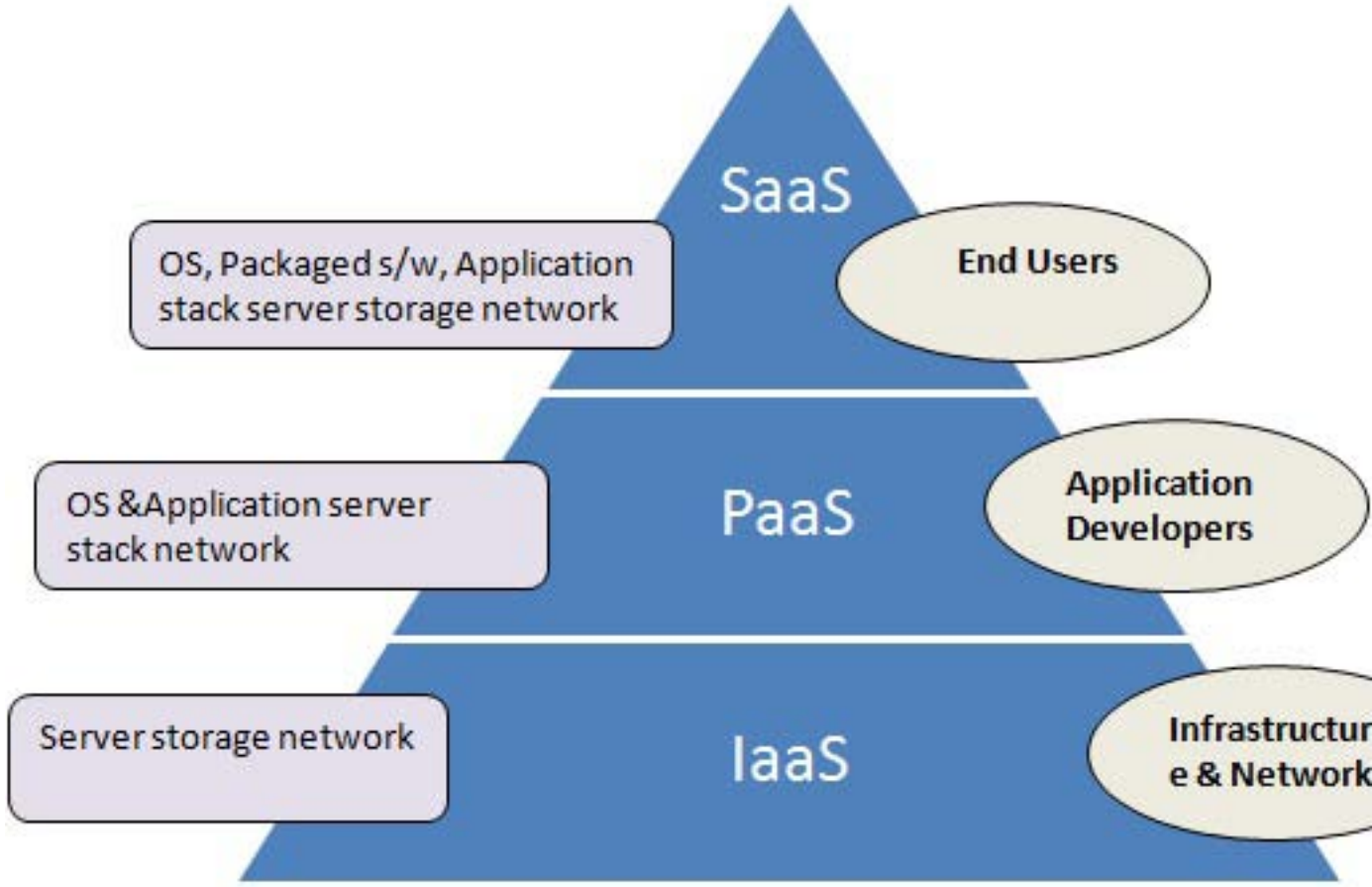
Block-based load-balancing algorithm: Entity Resolution with multiple keys in Map Reduce. Actually, the authors extended the Block Split algorithm presented in **??**olb et al. (2011) by considering more than one blocking key. In their algorithm, the load distribution in the Reduce phase is more precise because an entity pair may exist in a block only when the number of common blocking keys between the pair exceeds a certain threshold (i.e., kc). Since an entity may have more than one kc key, it needs to generate all the combinations of kc keys for potential key comparisons. The proposed algorithm features in the combination based blocking and loadbalanced matching. Experiments using the well-known Cite Seer X digital library showed that the proposed algorithm was both scalable and efficient.

Genetic Algorithm: A genetic algorithm for scheduling and load balancing for static parallel heterogeneous system. Their techniques consider five main factors: Encoding generation of initial populations, fitness function, selection operator and crossover operator.

Each task is considered as a gene. A generation of an initial random population for entry into the first generation was done by the genetic algorithm. Random generator functions of chromosomes are employed. Individual are selected according to their fitness value. Once fitness values have been evaluated for all chromosomes, good chromosomes is selected through rotating roulette wheel strategy. Crossover operator randomly selects two parent chromosomes (chromosomes with higher values have more chance to be selected) and randomly chooses their crossover points, and mates them to produce two child (offspring) chromosomes. However, their approach was able to reduce response time and execution time when compared with LPT, SPT and FIFO algorithms. node, the ant moves in a forward direction with the overloaded node to the next node checking whether the node is overloaded or not. During the movement, if it particular node. Once the task is assigned to the head as head node in Cloud Computing Service Provider. As the request is being sent, the ant starts is first movement from the head node. The ants collect the information from the cloud node and assign the tasks to the Table1: This table find the load balancing techniques

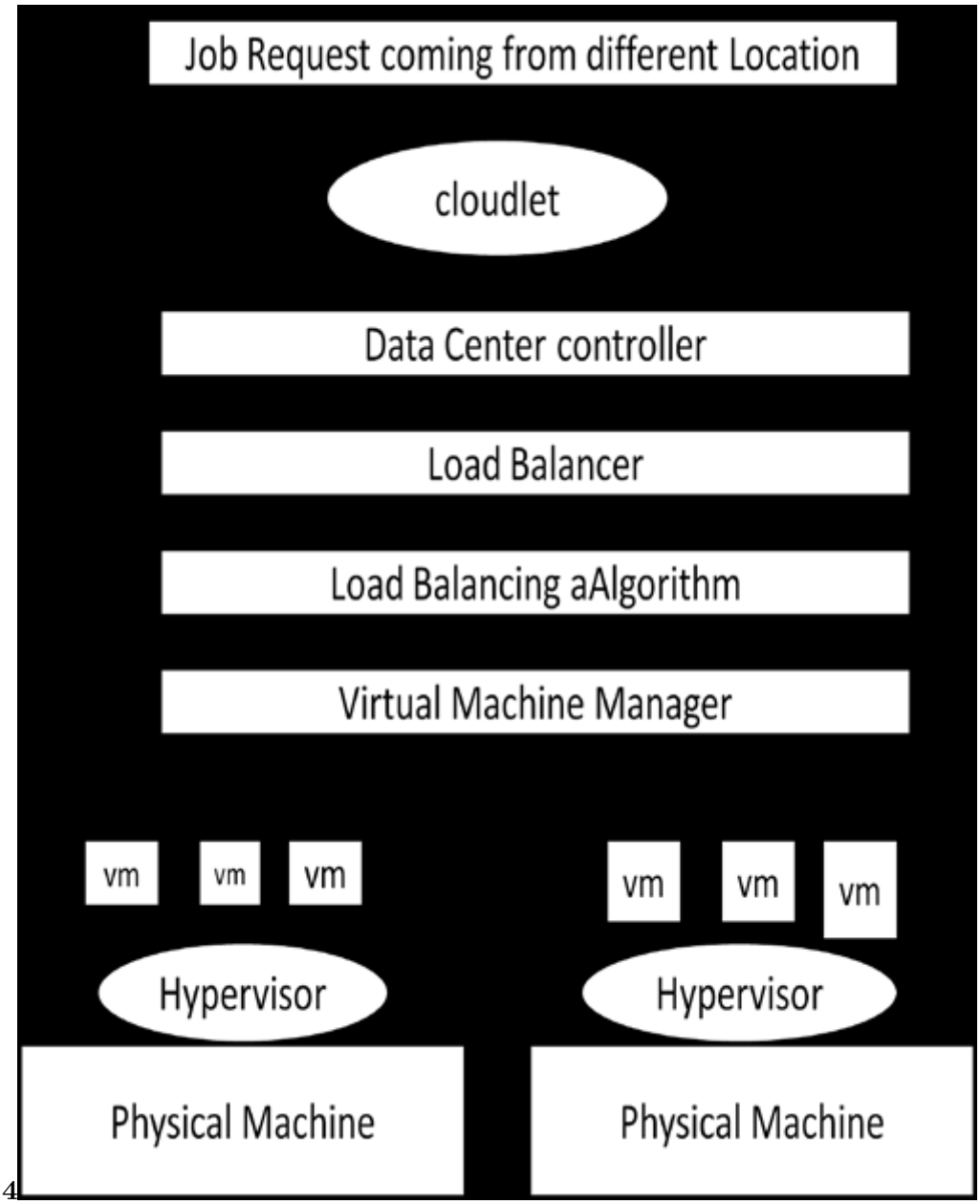
2

Figure 1: Fig. 2 :



3

Figure 2: Fig. 3 :



4

Figure 3: Fig. 4 :

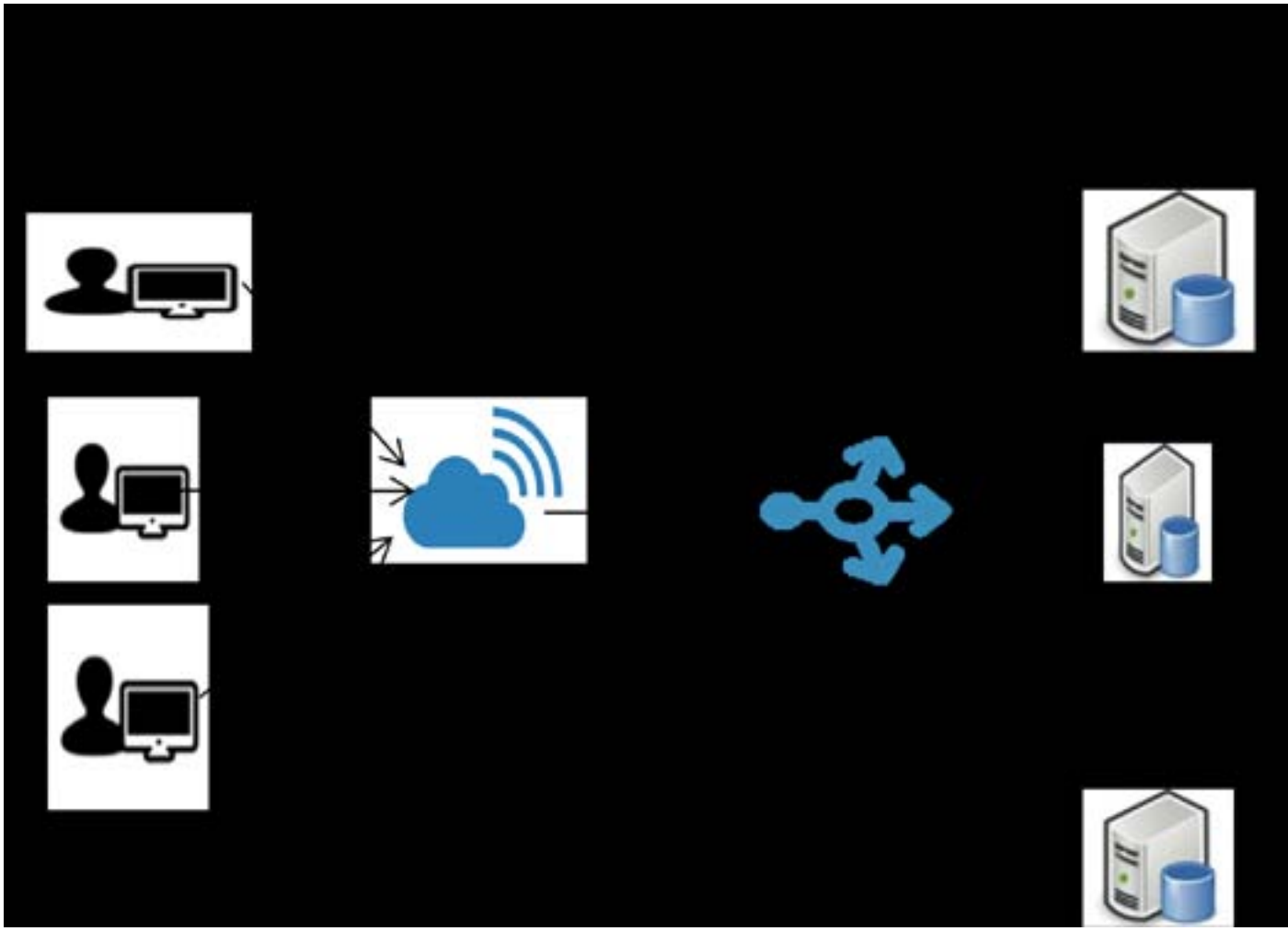


Figure 4:

23 References

[1]

[1] © 2019 Global JournalsSurvey on Load Balancing in Cloud Computing

From table1 and table2 the various nature various load balancing techniques and the timeline are mentioned in a clear manner. This is very helpful to the researchers.

.1 VII.

.2 Conclusion

The performance of the various load balancing algorithms are studied this paper. Load balancing is an important technique for improving distributed system performance by considering the group of hosts in the system to share their workloads. This results in a better utilization of hosts resources, a high system throughput and quick response time of user requests. In load sharing, the incoming client requests are evenly distributed among the participating hosts. Load balancing is a basic errand in cloud computing condition to accomplish greatest usages of assets. In this various proposed algorithm are numerous performance to study parameters in cpu cost, memory cost, configuration time and distance cost.

.3 REFERENCES RÉFÉRENCES REFERENCIAS

[Dharmesh Kashyap (2014)] *A survey of various Load balancing Alogoritham in cloud computing*, Dharmesh Kashyap . nov 2014. p. . (Vol3,issuse11 pp)

[Amit Agarwal, saloni Jain Efficient optimal Algorithm of task scheduling in cloud computing environment] *Amit Agarwal, saloni Jain Efficient optimal Algorithm of task scheduling in cloud computing environment*,

[Mishra and Jaiswal ()] 'Ant colony Optimization: A Solution of Load balancing in Cloud'. Ratan Mishra , Anant Jaiswal . *IJWesT* 2012.

[Violettan] *chernenkaya" load balancing in cloud computing*, Violettan . 9788-1-5386-4340- 2/18/&2018IEEE.

[James and Verma ()] 'Efficient VM load balancing algorithm for a cloud computing environment'. Jasmin James , Dr Bhupendra Verma . *IJCSE* 2012.

[Bhoi et al. (ed.) (2013)] *Load Balancing Algorithms in Cloud Computing Environment-A Methodical Comparison*, Upendra Bhoi , N Purvi , Ramanuj . 4, 12. J. Uma, V. Ramasamy, A. Kaleeswaran (ed.) April 2013. Feb 2014. 2 p. . (Enhanced Max-Min Task Scheduling Algorithm in Cloud Computing)

[Gupta and Sanghwan (2015)] 'Load Balancing in Cloud Computing: A Review'. Shikha Gupta , Suman Sanghwan . *International Journal of Science, Engineering and Technology Research* June 2015. 4 (6) .