Research Paper on Generating Load Balancing Algorithm for Cloud Computing

Gagandeep Kaur[1], Dr. Shahista Navaz[2]
[1]PhD Scholar, CSE Department, Shri Rawatpura Sarkar University, Raipur (C.G.),
[2]Asstt. Professor, CSE Department, Shri Rawatpura Sarkar University, Raipur (C.G.),
[1]kgagan619@yahoo.com, [2]shahista.navaz@gmail.com

Abstract:
Cloud load balancing is defined as the method of splitting workloads and computing properties in a cloud computing. It enables enterprise to manage workload demands or application demands by distributing resources among numerous computers, networks or servers. Cloud load balancing includes holding the circulation of workload traffic and demands that exist over the Internet. A many approaches were outlined in literature to enhance performance, job scheduling, storehouse coffers, QoS and cargo distribution. Cargo balancing conception permits data centers to fore stall over-loading or under- lading in virtual machines that as similar is an issue in cloud computing sphere. Accordingly, it bear the experimenters to layout and apply a proper cargo balancer for cloud terrain. The separate study represents a view of problems and pitfalls faced by the current cargo balancing ways and make the experimenters find more effective algorithms. cargo unbalancing problem is a multi-variant, multi-constraint problem that degrades performance and effectiveness of computing coffers. cargo balancing ways feed the result for cargo unbalancing situation for two undesirable angles- overfilling and under- lading. In disdain of the significance of cargo balancing ways to the stylish of our knowledge, there's no comprehensive, expansive, methodical and hierarchical bracket about the being cargo balancing ways. Further, the factors that beget cargo unbalancing problem are neither studied nor considered in the literature. The load balancer distributes data depending upon how busy each server or node is. In the absence of a load balancer, the client must wait while his process gets processed, which might be too tiring and de-motivating for him. Various information like jobs waiting in queue, CPU processing rate, job arrival rate etc. are exchanged between the processors during the load balancing process. Failure in the right application of load balancers can lead to serious consequences, data getting lost being one of them.

Keywords: Cloud computing, Taxonomy, Classification, Cloud service consumer, Cloud service provider, Quality of Service, Load unbalancing, Load balancing.

I. INTRODUCTION

Cloud cargo balancing is defined as the system of splitting workloads and calculating parcels in a cloud computing. It enables enterprise to manage workload demands or operation demands by distributing coffers among multitudinous computers, networks or waiters. cloud cargo balancing includes holding the rotation of workload business and demands that live over the Internet. As the business on the internet growing fleetly, which is about 100 annually of the present business?

The cloud computing can be described as an on- demand service pool which connects colorful waiters to each other for furnishing services to aiming guests. The cloud providers

may contain direct access to these services. thus, the coffers can be used according to the demand. The stoner can prize and modifies the data stored in the shadows. The different services to the stoner are handed on demand using a point called " cloud service provider ". This particularity makes certain that the quantum of services being employed for any number of times can be employed for calculating the expenditure of the stoner to pierce that service. The cloud calculating system provides extremely complicated operations in different surroundings. In addition, some professed concentrated services are handed in each terrain. In cloud computing, common group of coffers is handed to the druggies. Using cloud computing, the druggies can use these coffers according to their need far and wide. The main ideal of this technology is to maintain the minimal cost to pierce the services. It's anatomized that the software and tackle means attained using internet remain present in the virtual system and supports to give the services. The stoner accesses a common group of coffers using cloud computing on the base of demand. The virtualization allows stoner to subscribe and use the services for a certain time period by getting access of the common group of coffers using cloud computing.

Load balancing results can be distributed into two types –

1. Software- grounded cargo balancers Software- grounded cargo balancers run on standard tackle(desktop, PCs) and standard operating systems.
2. tackle- grounded cargo balancer tackle- grounded cargo balancers are devoted boxes which include Application Specific Integrated Circuits(ASICs) acclimated for a particular use. ASICs allows high speed promoting of network business and are constantly used for transport- position cargo balancing because tackle- grounded cargo balancing is briskly in comparison to software result.

II. LOAD BALANCING MODEL BACKGROUND

In cloud computing, cloud waitpersons should always be balanced, to use the resources with their full capacity. sometimes it happens that some waitpersons are heavily loaded while the other waitpersons are under loaded or in idle state. To overcome this problem weight balancing algorithms are used. These algorithms help in allocating every single task by covering weight on each garçon. According to the balancing algorithm is defined as " The weight balancing in murk may be among physical hosts or VMs. This balancing medium distributes the dynamic workload inversely among all the bumps(hosts or VMs). The weight balancing in the cloud is also appertained to as weight balancing as a service(BaaS) ".

There are different types of weight balancing algorithms which are used for cloud computing; they are categorized in two orders videlicet stationary weight balancing and dynamic weight balancing. stationary weight balancing algorithms allocate tasks to the waitpersons before the florilegium where all the conditions of the resources are known to the algorithm. The allocation of tasks are predicated on those conditions. stationary weight Balanced Algorithm is suitable for small distributed surroundings with high internet speed and ignorable communication detainments It works properly when the systems or bumps have the ignorable differences in the weight, therefore the algorithms which come under the static are generally not suitable for cloud computing. Because in the cloud we have n number of stoners due to which weight largely varies. Dynamic algorithms work in the real time situation, where it takes continuous information about the weight on the garçon. With respect to that it takes the decision of distributing the tasks amongst the waitpersons.

Accordingly we can allocate, reallocate or remove any task from the garçon predicated on the priority. Dynamic weight Balanced Algorithm focuses on reducing communication detainments and execution time for large distributed surroundings. These ways or the algorithms are largely successful for weight balancing the cloud terrain on their bumps among different types of resources. In the formerly numerous times there are multitudinous static and dynamic weight balancing algorithms that have been proposed for the cloud calculating terrain. A detailed comparison is done in. In this section some of the being algorithms proposed by the researchers are mooted. The static weight balancing algorithm factory in small distributed surroundings so they are less complex compared to dynamic which have a largely distributed terrain. Advanced information is demanded in static algorithms analogous as length and number of tasks. The scheduling opinions are taken at runtime by dynamic algorithms and collect time by static. stationary algorithms are not good at balancing weight properly at run time but monitor bumps continuously where dynamic balance loads efficiently and do the monitoring continuously by event base or time interval. stationary algorithms take farther time to break but do not give the optimal result for the complex computational problem, dynamic takes lower time and gives useful results. Traditional types of algorithms come under the stationary bones and the meta- heuristic algorithms come under the dynamic algorithms.

III. RELATED WORK

However, load balancing has been an eagle's eye among researchers because of its essence in cloud computing between the stakeholders' i.e. Cloud Service Provider and Cloud Service Consumer. Based on analysis of existing review literature one of the reasons presented is absence of proper classification among different approaches. A thorough review about the existing work in literature has been presented in this section.

Sovban Nisar, Deepika Arora [1] For handling the problems related to node failure in cloud networks, an algorithm named BFO is used in this research. Several nodes are included in a proposed algorithm. Depending upon the failure rate and minimal execution time, a participant node is chosen among all these nodes. In this scenario, the threshold value is fixed using the master node. There are two parameters included in this threshold value. The master node chooses nodes having equivalent or less failure rate with least execution time as the participant nodes. In comparison to threshold value, the value of node N1 is less.

Mohieddin Harb [2] As a result of the drawback of using throttled load balancing algorithm we proposed the balanced throttled load balancing which work as follow: The index table of all the virtual machines is maintained by Balanced Throttled Load Balancer. This also maintains the state of each virtual machine i.e. whether the virtual machine is busy or available. Initially, at the start of the algorithm, all the virtual machines have been present. Then, Data Center Controller gets the fresh task.

Sambit Kumar Mishra [3] The task allocation algorithms in the cloud are classified based upon the current state of VM. In allocation policy where the current load information of VMs are available before the allocation is said to be a dynamic strategy. Whereas the static strategy acts on VMs without any load information. Load balancing attends in fair allocation of resources to achieve a high user satisfaction and improve the stability of the system. We have proposed a taxonomy for the load balancing algorithms in the cloud environment as

shown in Fig. 3. Resource management plays a major role in the load balancing of cloud resources.

Dr. Sharvani GS [4] With more and more advances being made in cloud computing and its increasing efficiency, companies have started using cloud as their underlying architecture for most of the important operations. The demand for resources is always increasing in these companies and with the help of cloud architecture, all the demand requirements are met easily. Cloud allows them to increase/decrease the load on servers according to their requirements as cloud provides the policy of pay-as-you-go which makes it a good option for the organizations.

Iehab AL Rassan and Noof Alarifi [5] proposed a fine-grained data access control withattribute hidingpolicy for cloud-based IoT. A fine-grained access control policy was also put forward to support an excessive access policy with full attributes hidden for cloud-based IoT. Herein, attribute-based information is fully hidden using arandomizable technique. A fuzzy attribute positioning mechanism is used to locate the attributes of authorized users efficiently. A garbled bloom filter is used for this process. However, the study's use of the garbled bloom filter causes a high number of false positives, which indicate that an attribute is a member of an access policy group when it is really not.

Muhammad Asim Shahid [6] LB provides a systematic mechanism for the equal distribution of the responsibility to the resources available. The goal is to provide reliable service, including adequate use of the resource, in the event of a disaster of the portion of any service by supplying & de-provisioning the device instance. In addition, LB is aimed at reducing response time for tasks & increasing resource ef_ciency, which increases device ef_ciency at a lower cost.

Asha Sohal, Ramesh Kait [7] Cloud computing is a wide and fastest growing area in terms of computing research and industry these days. It mainly provides services based on IaaS, SaaS, and PaaS. These are the key parameters which decide the role of cloud services to the end users. These services can be offered to the end users through virtualization over the internet. Cloud has many advantages like large scaled computing, flexible infrastructures, pay per use, on demand services and many more. There are some major issues in processing of jobs over cloud computing like security, equal distribution of load, fault tolerance etc. and the biggest challenge over cloud is latency time which means the total time between the data sent by IoT over cloud, processing time and finally reply to the IoT or vice versa.

Sally F. Issawi [8] The request rates received by the data center are not constant all the time. Sometimes large number of requests aggregated in a small period of time creating a burst. This affect the performance of the load balancing algorithm as it increase the processing time and the repose time of the data center. The performance of several load balancing algorithms differs according to the users' requests rate. For example some algorithms work efficiently under low workload while their performance is degraded under high workload and vice versa. To overcome burst problem and benefit from different load balancing algorithms advantages we propose a new load balancing algorithm called Adaptive algorithm.

Sovban Nisar[9] The evolution of Virtualization, Utility computing, Software-as-a-Service (SaaS), Infrastructure-as-a-Service (IaaS) and Platform-as-a-Service (PaaS) all are combined to make a cloud computing and three development models of cloud are public, private and

hybrid. Public cloud services are available for general public over the internet . Private cloud is used for personal use or provides services to single organization. A hybrid cloud is combination of two or more than two public and private cloud which are bounded by service level agreement (SLA). Clients/Users can forward the requests at any time from any geographical location/region for the required services, SLA selects the best resource within user defined deadline and budget. Elastic resource provisioning with quality of service (QoS) parameter (deadline, high availability, priority etc.).

Yelchuri Venkata Sai Harsha[10] Cloud computing is the distribution of diverse offerings along with storage, servers, networking, software programs, intelligence, and analytics, through the internet so as to offer faster innovation, more flexible sources, and economies of scale. Take an illustration of a site open to everyone. A high number of clients can visit a site or online application whenever. A web application's capacity to deal with these client demands without a moment's delay gets intense. It might even cause system failures. The terrible sense of a website being down or not accessible also delivers lost prospective clients for a website owner whose entire career is based on his portal. Load balancing is crucial in this situation.

IV. PROBLEM IDENTIFICATION

Previous methodology uses balanced throttled load balancing algorithm. The algorithm was as follow:

1. The index table of all the virtual machines is maintained by Balanced Throttled Load Balancer. This also maintains the state of each virtual machine i.e. whether the virtual machine is busy or available. Initially, at the start of the algorithm, all the virtual machines have been present.

2. Then, Data Center Controller gets the fresh task.

3. Data Center Controller, then on receiving the call, contacts Balanced Throttled Load Balancer to do the next allocation of the virtual machine.

4. Then balanced throttled Load Balancer construct new map and start to add all available VM.

5. Then balanced throttled Load Balancer deconstructs virtual machine available VM map if this map length is greater than 0 then it gets the first available VM Id from the map, then:

a) VM id is returned by the Balanced Throttled Load Balancer to the Data Center Controller.

b) The Data Center Controller then transfers the call to a respective virtual machine which has been identified by that virtual machine id.

c) Then Data Center Controller gives the notification to the Balanced Throttle load balancer about the allotment about the new virtual machine id, then remove this VM Id from map of available VM.

d) On receiving the call from the Data Center Controller, Balanced Throttled Load Balancer then upgrades the virtual machines available VM map consequently.

When available virtual machine is not found in the available VM Map then:

a) 1 is returned by the Balanced Throttled Load Balancer.

b) Then the request is put into the queue by the Data Center Controller.

c) When all the processing request is completed by the virtual machine and the response has also been received by the Data Center Controller, Balanced Throttled Load Balancer gets a

notification from Data center controller to perform the de-allotment of the respective virtual machine and add this available VM Id to available VM Map.

d) Now when the virtual machine is de-allocated, then the Data Center Controller examines the awaiting call queue. When some waiting calls in the pending queue exist, the processing of the call starts from 3rd step onwards.

The drawback of previous methodology was to take extra time to searching destination server. Every time throttled method has to maintain the availability of server list, this takes so much time. Every time up gradation needed and for that step 3 onwards repeated every time.

V. PROPOSED METHOD

The proposed algorithm will complete in following process:

1. Simulate the raw data for analysis from data warehouse

In first step data collection will be done for further process. The data will co relates various homogeneous data with large amount of data.

2. Implementation of the classification scheme

Classification scheme can be implemented to divide the data in various data clusters.

3. Data Formation

Frame the data according to data model used.

4. Threshold value for master node

The threshold value is fixed using the master node. It searches the node with less failure.

5. Compare the time and failure ratio

Time analysis will be a major part where retrieval methods can compare the minimum time for extracting data as well failure rate must be very less in the time of searching.

6. Find out same cluster server

Cluster of server will help to make sure that data comes from physical section is most accurate and similar.

7. Selection of less loaded server

The method which has responsibility to select most accurate server among cluster for most accurate result.

8. Compare the resultant data

Resultant data can be found using comparative method with most efficient cluster and less time counter.

In this algorithm the cargo balancer responsible for preparing indicator table. It displays the information of the virtual machine state either Available or Busy. When the task is arrived, the cargo balancer assigns the task to suitable virtual machine. Which is available to execute the stoner task. But every time strangled checks the table from first indicator to determine the available virtual machine.

Flow chart

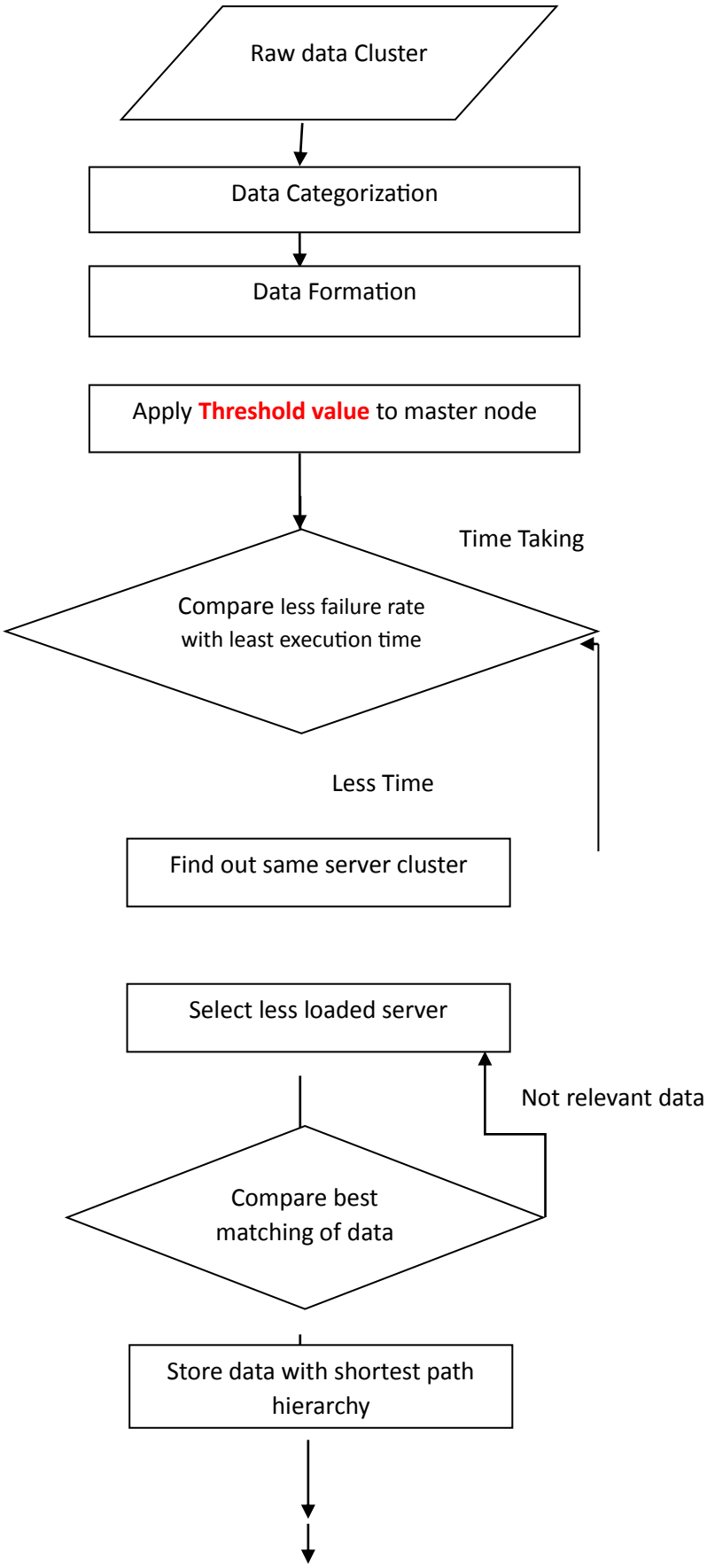

Fig 1: Flowchart for method

VI. Result Implementation

Threshold in case of clouds is an optimal division of loads among a number of master computers, slave computers and their communication links. Our objective is to obtain a minimal partition of the processing load of a cloud connected via different communication links such that the entire load can be distributed and processed in the shortest possible amount of time.

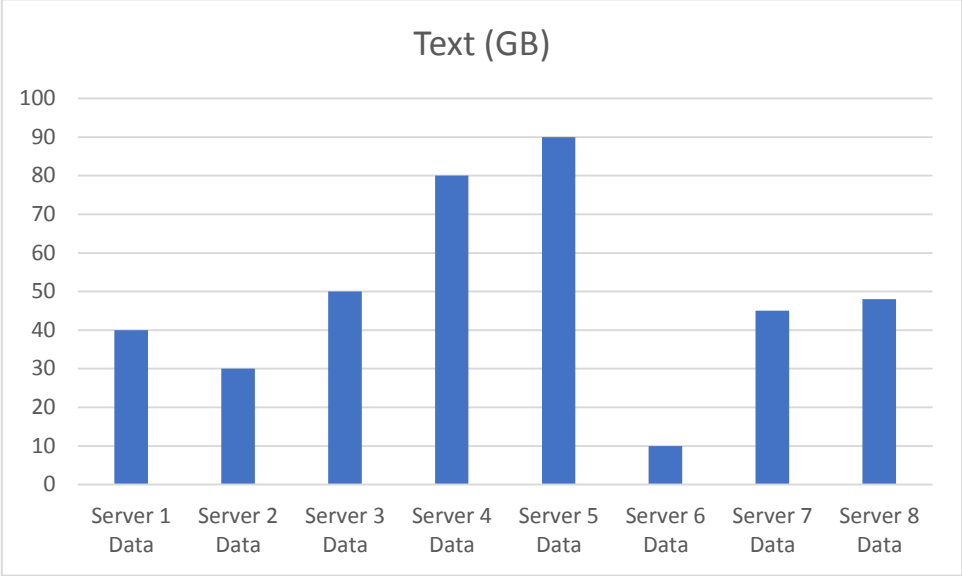

Fig 2: Server Details with text data.

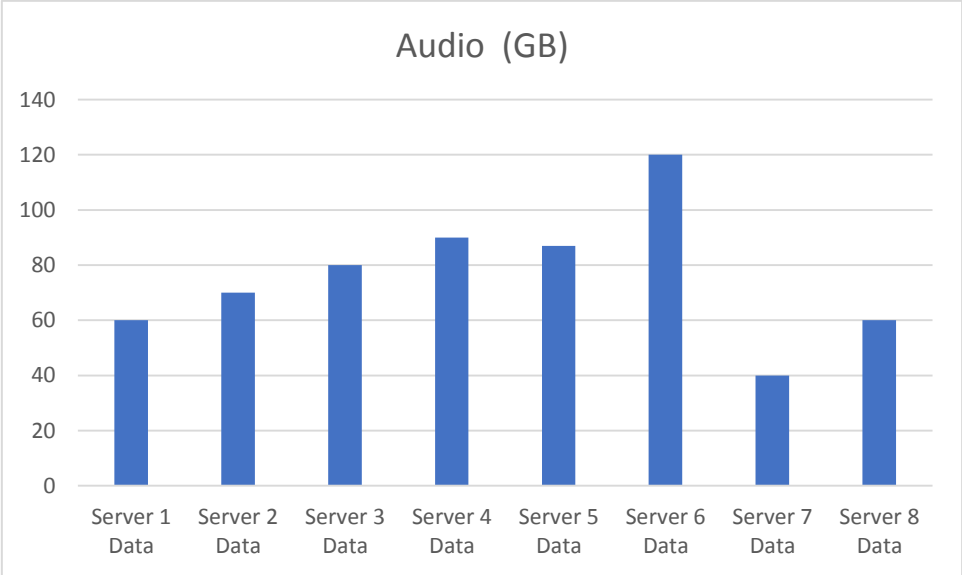

Fig 3: Server Details with audio data

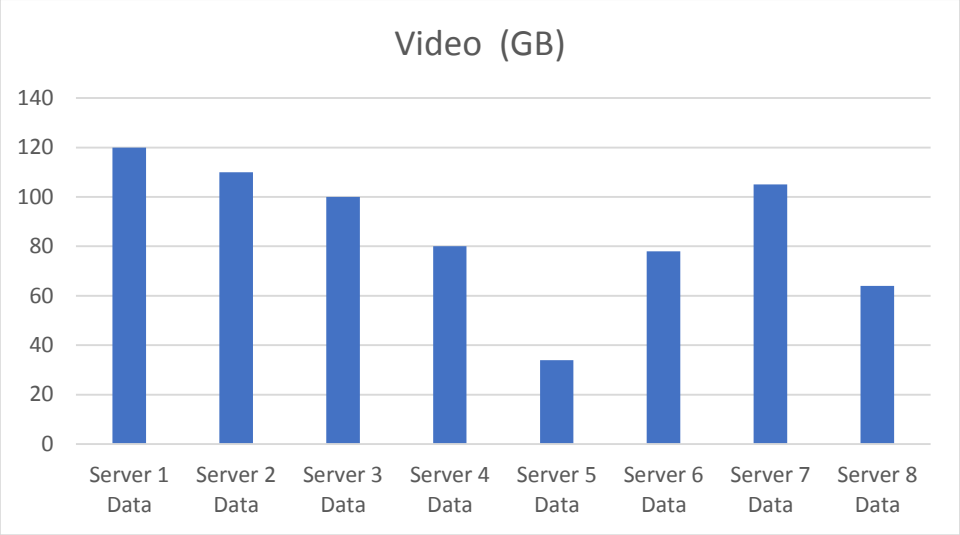

Fig 4: Server details with video data

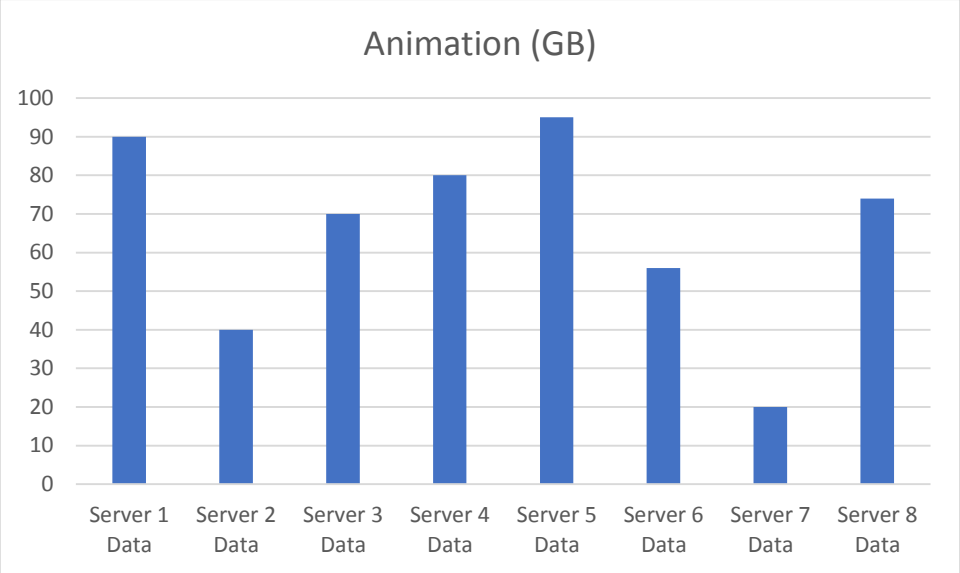

Fig 5: Server details with video data

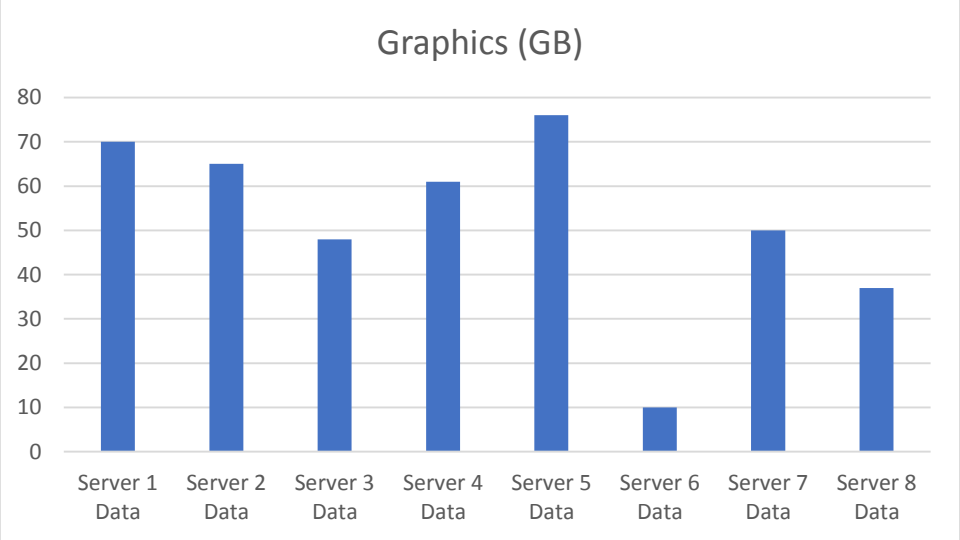

Fig 6: Server details with Graphics data

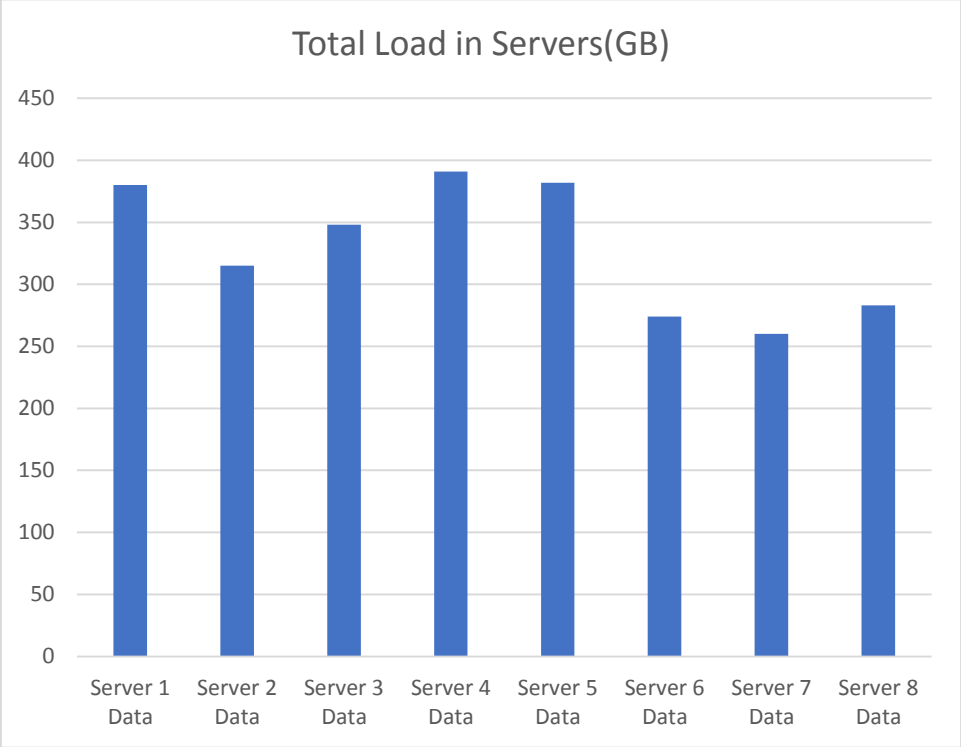

Fig 7: Server details with Total Load data

VII Conclusion

Measurement/reporting time of both the approaches for the same no. of slave computers corresponding to the same master. Herethe inverse link speed b is taken as 1 and the inverse measurement speed a is 0.5 for both the cases. Number of master computers is taken to be constant equal to 50. The plot shows thatthe measurement/reporting time is smaller in case of simultaneous reporting as compared tosequential reporting. It is because in case of sequential reporting, some of the slaves recieve almost zero load from its master. Number of effective slaves in this case is less as comparedto the simultaneous reporting case. Hence with increase in no. of slaves with respect to a master, the finishing time remains almost same in case of sequential reporting whereas in case of simultaneous reporting, the finishing time decreases for the increase in no. of slaves corresponding to a single master. The graph shows that the finishing time can be improvedby increasing the number of slaves under a master computer in a cloud only to some extent before saturation in case of sequential measurement and sequential reporting strategy. But finishing time can be decreased significantly in case of simultaneous measurement start andsimultaneous reporting termination by increasing the no. of slaves under a single master computer.

Cloud Computing is a vast concept and load balancing plays a very important role in case of Clouds. There is a huge scope of improvement in this area. We have discussed only two divisible load scheduling algorithms that can be applied to clouds, but there are still other approaches that can be applied to balance the load in clouds. The performance of the given algorithms can also be increased by varying different parameters.

References

[1] Mohamed, Shereen Yousef, et al. "A proposed load balancing algorithm over cloud computing (balanced throttled)." International Journal of Recent Technology and Engineering 10.2 (2021): 28-33.

[2] Sha, Kewei, et al. "A survey of edge computing-based designs for IoT security." Digital Communications and Networks 6.2 (2020): 195-202.

[3] Mishra, Sambit Kumar, Bibhudatta Sahoo, and Priti Paramita Parida. "Load balancing in cloud computing: a big picture." Journal of King Saud University-Computer and Information Sciences 32.2 (2020): 149-158.

[4] Sharvani, G. S. "An auto-scaling approach to load balance dynamic workloads for cloud systems." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12.11 (2021): 515-531.

[5] Chen, Shang-Liang, Yun-Yao Chen, and Suang-Hong Kuo. "CLB: A novel load balancing architecture and algorithm for cloud services." Computers & Electrical Engineering 58 (2017): 154-160.

[6] Shahid, Muhammad Asim, et al. "A comprehensive study of load balancing approaches in the cloud computing environment and a novel fault tolerance approach." IEEE Access 8 (2020): 130500-130526.

[7] Khan, Atif Ishaq, et al. "Intelligent cloud-based load balancing system empowered with fuzzy logic." Computers, Materials and Continua 67.1 (2021): 519-528.

[8] Kumar, Mohit, and Subhash C. Sharma. "Deadline constrained based dynamic load balancing algorithm with elasticity in cloud environment." Computers & Electrical Engineering 69 (2018): 395-411.

[9] Tamilvizhi, T., and B. Parvathavarthini. "A novel method for adaptive fault tolerance during load balancing in cloud computing." Cluster Computing 22 (2019): 10425-10438.

[10] Mishra, Kaushik, and Santosh Majhi. "A state-of-art on cloud load balancing algorithms." International Journal of computing and digital systems 9.2 (2020): 201-220.

[11] Ghomi, Einollah Jafarnejad, Amir Masoud Rahmani, and Nooruldeen Nasih Qader. "Load-balancing algorithms in cloud computing: A survey." Journal of Network and Computer Applications 88 (2017): 50-71.

[12] Shaw, Subhadra Bose, and A. K. Singh. "A survey on scheduling and load balancing techniques in cloud computing environment." 2014 international conference on computer and communication technology (ICCCT). IEEE, 2014.

[13] Ray, Soumya, and Ajanta De Sarkar. "Execution analysis of load balancing algorithms in cloud computing environment." International Journal on Cloud Computing: Services and Architecture (IJCCSA) 2.5 (2012): 1-13.

[14] Issawi, Sally F., Alaa Al Halees, and Mohammed Radi. "An efficient adaptive load balancing algorithm for cloud computing under bursty workloads." Engineering, Technology & Applied Science Research 5.3 (2015): 795-800.

[15] Jairam Naik, K. "A dynamic ACO-based elastic load balancer for cloud computing (D-ACOELB)." Data Engineering and Communication Technology: Proceedings of 3rd ICDECT-2K19. Springer Singapore, 2020.

[16] Kumar, Mohit, and Subhash C. Sharma. "Deadline constrained based dynamic load balancing algorithm with elasticity in cloud environment." Computers & Electrical Engineering 69 (2018): 395-411.

[17] Saif, Mufeed Ahmed Naji, et al. "CSO-ILB: chicken swarm optimized inter-cloud load balancer for elastic containerized multi-cloud environment." The Journal of Supercomputing 79.1 (2023): 1111-1155.

[18] Mishra, Sambit Kumar, Bibhudatta Sahoo, and Priti Paramita Parida. "Load balancing in cloud computing: a big picture." Journal of King Saud University-Computer and Information Sciences 32.2 (2020): 149-158.

[19] Al Nuaimi, Klaithem, et al. "A survey of load balancing in cloud computing: Challenges and algorithms." 2012 second symposium on network cloud computing and applications. IEEE, 2012.

[20] Kaur, Rajwinder, and Pawan Luthra. "Load balancing in cloud computing." Proceedings of international conference on recent trends in information, telecommunication and computing, ITC. Citeseer, 2012.

[21] Mesbahi, Mohammadreza, and Amir Masoud Rahmani. "Load balancing in cloud computing: a state of the art survey." Int. J. Mod. Educ. Comput. Sci 8.3 (2016): 64.

[22] Joshi, Shalini, and Uma Kumari. "Load balancing in cloud computing: Challenges & issues." 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I). IEEE, 2016.

[23] Khan, Rafiqul Zaman, and Mohammad Oqail Ahmad. "Load balancing challenges in cloud computing: a survey." Proceedings of the International Conference on Signal, Networks, Computing, and Systems: ICSNCS 2016, Volume 2. Springer India, 2016.

[24] Afzal, Shahbaz, and G. Kavitha. "Load balancing in cloud computing–A hierarchical taxonomical classification." Journal of Cloud Computing 8.1 (2019): 22.

[25] Aditya, Abhijit, Uddalak Chatterjee, and Snehasis Gupta. "A comparative study of different static and dynamic load balancing algorithm in cloud computing with special emphasis on time factor." Int. J. Curr. Eng. Technol 3.5 (2015): 64-78.

[26] Katyal, Mayanka, and Atul Mishra. "A comparative study of load balancing algorithms in cloud computing environment." arXiv preprint arXiv:1403.6918 (2014).