

A Literature survey of various cloud load balancing techniques

¹Ms. Madhuri Kaushal, ²Mr. Mohit Jain

ABSTRACT: Cloud computing is an interesting era of research, where motivation is to find out the best outcome and productive data security and sharing approach. Load balancing in public impair by way of division of cloud just right geographical position. Load balancing is frequently a strategy of controlling the visitors in a cloud atmosphere. Cloud requests hunt for assets for performance. The resources are quite often storage, processing, bandwidth, and many others. Allocation these belongings efficaciously to the entire competing jobs are named as load balancing. This paper will provide a comprehensive survey of cloud load balancing techniques.

1. INTRODUCTION:

Cloud Computing has emerge as probably the most manageable answer for the issues which might be computation intensive. Cloud supplies the approach of sharing assets and offerings to person on demand. Virtualized useful resource and offerings can be utilized without the acquaintance of geographical variances. Cloud computing delivers mechanisms that deliberate on run time request of computing assets like storage, availability, software and so on.

The Cloud platforms distinguish among the many service style, the rate, and the nice of carrier (QoS) as good as performance. This truth brings Cloud patrons the flexibleness of freely settling on target structure from large range of Cloud platforms. However, even as this raises the difficulty of the interoperability among the one-of-a-kind Clouds [1].

Progress of efficient carrier provisioning policies is the principal issues in Cloud study. Today's Clouds exists in an exposed world categorized by way of consistent alterations happening autonomously and unpredictably. On this context, game theoretic ways allow in-depth analytical figuring out of the provider provisioning difficulty [2].The cloud computing almost always offers three kinds of facilities viz: (IaaS) Infrastructure as a service, (PaaS) Platform as a service and Software as a service (SaaS).

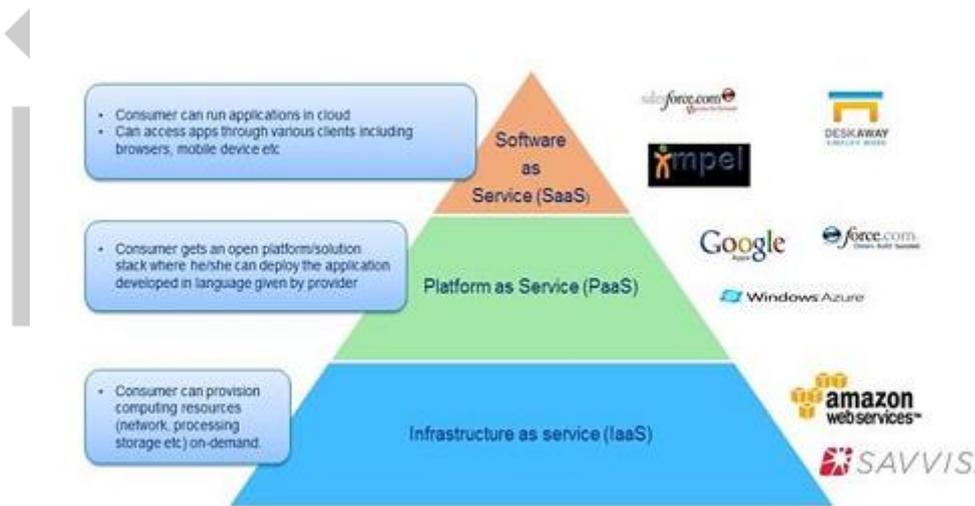


Figure 1: Cloud Service Models [3]

The Cloud computing is a internet based network. Cloud is a collection of services. Cloud provides on demand services. The major services provided through cloud are: hardware service, software service, network service. Cloud computing is a modern field, which revolves around utility computing, service oriented architecture, internet, clients etc.

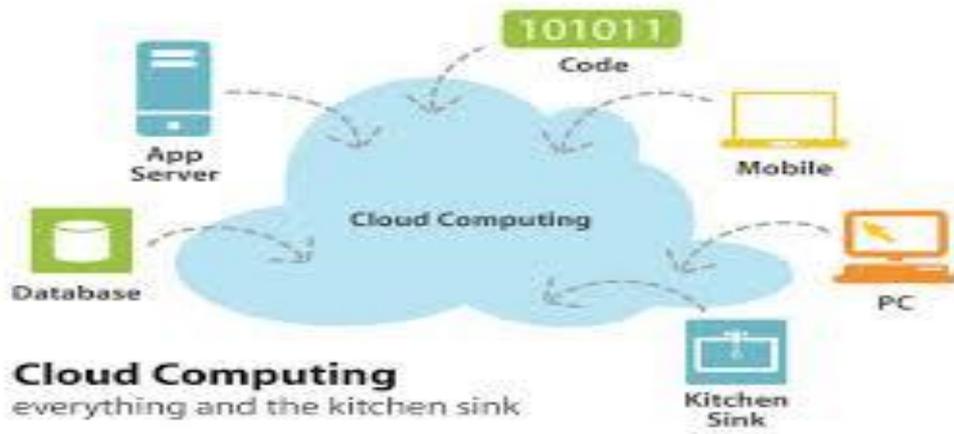


Figure 2: Cloud computing Environment [3]

Cloud Computing still under inside their development stage and also has quite a few issue in addition to challenges out of a several questions in cloud scheduling plays very important role inside determining your current effective execution. Scheduling refers for the set connected with policies to be able to control your order involving function for you to possibly be performed by a computer system. There have been different people associated with scheduling algorithm existing throughout distributed computing system along with job scheduling will be single of them. The main advantage involving job scheduling algorithm will be in order to achieve a good high performance computing and also the simplest process throughput. Scheduling manages availability involving CPU memory and good scheduling policy gives maximum utilization of resource.

2. LOAD BALANCING IN CLOUD COMPUTING

The load balancing is the process of distributing the load among various resources in any system. Therefore load need to be distributed over the resources in cloud-based architecture so that each resources does approximately the equal amount of task at any point of time. The basic need is to provide some techniques to balance requests to provide the solution of the application faster. All cloud vendors are based on automatic load balancing services, it allows clients to increase the number of CPUs or memories for their resources to scale with increased demands. These services are optional and depend on the clients business needs. So the load balancing serves two important needs, firstly to promote availability of Cloud resources and secondarily to promote performance.

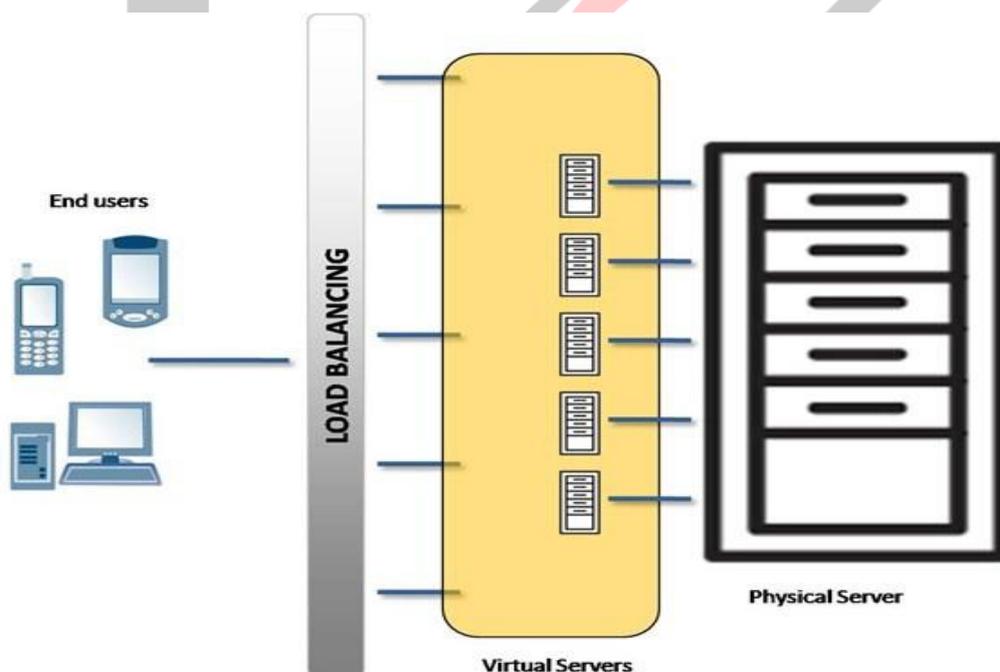


Figure 3: Load Balancing in Cloud Environment [4]

In order to balance the resources it is important to recognize a few major goals of load balancing algorithms:

- a) Cost effectiveness: first aim is to achieve an overall improvement in system performance at a reasonable cost.
- b) Scalability and flexibility: distributed system in which the algorithm is implemented may change in size or topology So the algorithm must be scalable and flexible enough to allow such changes to be handled easily.
- c) Priority: scheduling of the resources or jobs need to be done on before hand through the algorithm itself for better service to the important or high prioritized jobs in spite of equal service provision for all the jobs regardless of their origin.

3. LITERATURE REVIEW:

This section consists of previous works that had been already proposed by several researchers. Some common approaches are also discussed here that work efficiently with response time, data center processing time and cost.

The work done by [4] proposed a novel load balancing algorithm called VectorDot. This algorithm handles the hierarchical complexity of the datacenter and multidimensionality of resource loads across servers network switches and storage in an agile data center that has integrated server and storage virtualization technologies.

The work done [5] proposed a mechanism CARTON for cloud control that unifies the use of LB and DRL. The LB (Load Balancing) is used to equally distribute the jobs to different servers so that the associated costs can be minimized and DRL (Distributed Rate Limiting) is used to make sure that the resources are distributed in a way to keep a fair resource allocation.

[6] addressed the problem of intra-cloud load balancing amongst physical hosts by adaptive live migration of virtual machines. The load balancing model is designed and implemented to reduce virtual machines migration time by shared storage to balance load amongst servers according to their processor or IO usage.

Work done by [7] presented an event driven load balancing algorithm for real-time Massively Multiplayer Online Games (MMOG). The algorithm after receiving capacity events as input, also analysis its components in context of the resources and the global state of the game session, then generating the game session load balancing actions.

The [8] proposed a scheduling strategy on load balancing of VM resources that uses historical data and current state of the system. Proposed strategy achieves the best load balancing and reduced dynamic migration by using a genetic algorithm.

The [9] proposed a Central Load Balancing Policy for Virtual Machines (CLBVM) that balances the load evenly in a distributed virtual machine/cloud computing environment.

The [10] proposed a load balancing virtual storage strategy (LBVS) that provides a large scale net data storage model and Storage as a Service model based on Cloud Storage. The Storage virtualization is achieved using an architecture that is three-layered and load balancing is achieved using two load balancing modules. It helps in improving the efficiency.

The [11] discussed a two-level task scheduling mechanism based on load balancing to meet dynamic requirements of users and obtain high resource utilization. Algorithm achieves load balancing by first mapping tasks to virtual machines and then virtual machines to host resources thereby improving the task response time, and resource utilization also overall performance of the cloud computing environment.

Author [12] investigated a decentralized honey bee based load balancing technique that is a nature inspired algorithm for self-organization. Algorithm achieves global load balancing through local server actions. Performance of the system is enhanced with increased system diversity but throughput is not increased with an increase in system size. This is best suited for the conditions where the diverse population of service types is required.

4. Proposed Technique:

Updated proposed throttled algorithm is as follows:

INPUT:

- DATA CENTRE REQUESTS R_1, R_2, \dots, R_N
- AVAILABLE VIRTUAL MACHINES VM_1, VM_2, \dots, VM_N

OUTPUT:

- DATA CENTRE REQUESTS R_1, R_2, \dots, R_N ARE ALLOCATED AVAILABLE VIRTUAL MACHINES VM_1, VM_2, \dots, VM_N

STEP1. INITIALIZE ALL OF THE VMS ALLOCATION STATUS TO AVAILABLE IN THE VM STATE LIST.

STEP2. MAINTAIN A HASH MAP TABLE OF VM ON THE BASIS OF PROCESSING TIME, RESPONSE TIME, AND AVAILABLE MEMORY. INITIALIZE THE HASHMAP WITH NO ENTRIES.

STEP3. DATACENTERCONTROLLER RECEIVES A NEW REQUEST.

STEP4. DATACENTERCONTROLLER QUERIES NEW LOAD BALANCER FOR NEXT ALLOCATION. HASH MAP IS SORTED IN DESCENDING ORDER OF THE RESPONSE TIME.

STEP5. IF HASHMAP LIST SIZE < VM STATE LIST SIZE, THEN ALLOCATE THE VM. ELSE WAIT FOR THE VM TO GET FREE.

STEP6. WHEN THE VM FINISH PROCESSING THE REQUEST AND THE DATACENTERCONTROLLER RECEIVES THE CLOUDLET RESPONSE, IT NOTICES THE LOAD BALANCER OF THE VM DEALLOCATION.

STEP7. WHEN THE VIRTUAL MACHINE FINISHES THE REQUEST.

- THE DATA CENTRE CONTROLLER SENDS A NOTIFICATION TO UPDATED THROTTLED THAT THE VM ID HAS FINISHED THE REQUEST.
- UPDATED THROTTLED MODIFIES THE HASH MAP TABLE ACCORDINGLY

5. CONCLUSION:

In this sort of paper, we describe load balancing in a public cloud by way of partitioning this cloud into a few sub-clouds. This division of public impair right into a quantity of sub-clouds is finished excellent geographical vicinity. This paper elaborated the notion of cloud load balancing in lucrative manner. It also provided a detailed study of existing and most popular techniques of cloud load balancing. The authors of this paper hope that this research work will be fruitful for future researchers.

References:

- [1] John Harauz, Lorti M. Kaufinan. Bruce Potter, "Data Security in the World of Cloud Computing", IEEE Security & Privacy, Co published by the IEEE Computer and Reliability Societies, July/August 2009.
- [2] National Institute of Standards and Technology- Computer Security Resource Center -www.csrc.nist.gov
- [3] http://en.wikipedia.org/wiki/Cloud_computing.
- [4] Yashpalsinh Jadeja and Kirit Modi, "Cloud Computing - Concepts, Architecture and Challenges", International Conference on Computing, Electronics and Electrical Technologies [ICCEET], IEEE-2012.
- [5] Ram Prassd Pandhy (107CS046), P Goutam Prasad rao (107CS039). "Load balancing in cloud computing system" Department of computer science and engineering National Institute of Technology Rourkela, Rourkela-769008, Orissa, India May-2011.
- [6] J. Sahoo, S. Mohapatra and R. lath "Virtualization: A survey on concepts, taxonomy and associated security issues" computer and network technology (ICCNT), IEEE, pp. 222-226. April 2010.
- [7] Bhaskar. R, Deepu.S. R and Dr.B. S. Shylaja "Dynamic Allocation Method For Efficient Load Balancing In Virtual Machines For Cloud Computing Environment" September 2012.
- [8] R.Shimonski. Windows 2000 & Windows server 2003 clustering and load balancing. Emeryville. McGraw-Hill Professional publishing, CA, USA (2003), p 2, 2003.
- [9] J. Kruskall and M. Liberman."The Symmetric Time Warping Problem: From Continuous to Discrete. In Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison", pp. 125-161, Addison-Wesley Publishing Co., 1983.
- [10] Mr. Nitin S. More, Mrs. Swapnaja R. Hiray and Mrs. Smita Shukla Patel," Load Balancing and Resource Monitoring in Cloud", International Journal of Advances in Computing and Information Researches ISSN: 22774068, Volume 1– No.2, April 2012.
- [11] R. X. T. and X. F. Z,"A Load Balancing Strategy Based on the Combination of Static and Dynamic, in Database Technology and Applications (DBTA)",2nd International Workshop,2010.
- [12] M Randles, D. Lamb, and A. Taleb-Bendiab, "A comparative study into distributed load balancing algorithms for cloud computing," 2010 IEEE 24th international conference on advanced information networking and application workshops,2010, pp. 551-556.