

A Study - Load Balancing Techniques in the Cloud Computing

Venkateshwarlu Velde¹, Dr. B. Rama²

Research Scholar, Dept of Computer Science, Kakatiya University, Warangal, Telangana, India¹

Asst. Professor, Dept of Computer Science, Kakatiya University, Warangal, Telangana, India²

ABSTRACT: Nowadays, Cloud computing is a turn-up prototype of the computing world which aims to share data over a single platform. After 2007, cloud computing has turned as buzzwords in ICT industry. 2 players in cloud computing environments, cloud providers and cloud users, seek different goals; providers want to expand revenue by getting high resource utilization, while users want to reduce costs while meeting their performance requirements. However, it is difficult to allocate resources in a jointly optimal way due to the lack of information sharing between them. As an expanded demand of the resources of cloud computing, load balancing is the regular problem to be faced. Various load balancing algorithms have been constructed by various researchers. A close look at those algorithms and the comparison between those algorithms is being done in this review paper.

KEYWORDS: Cloud Computing, Load Balancing, Honey Bee Foraging Algorithm, Ant Colony Optimization Technique, RBAC, RASA

I. INTRODUCTION

Cloud Computing is a new generation which aims to have shared data over one platform. As the technology is booming fast, so does the demands of the clients. This new prototype of cloud computing is appealing vendors which increase its reputation. As the definition of NIST[9] says “Cloud Computing is a model for enabling convenient, an on-demand network connection to a common pool of configurable computing resources (e.g. network, server, storage and applications and services) that can be fast provisioned and discharged with minimal management efforts or service provider interaction. Cloud Computing Services:

According to NIST, Cloud Computing provides 3 services that are described in the table (1) below and these three models club to form cloud computing architecture.

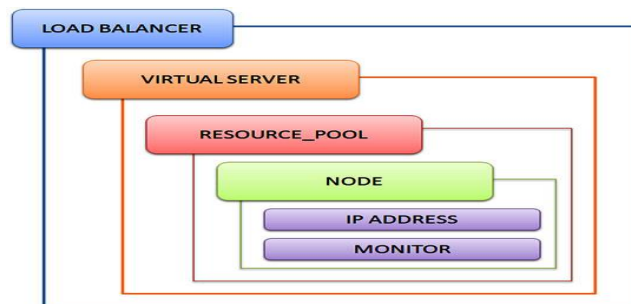


Fig 1: Load Balancer

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 9, September 2016

a) SaaS (Software as a Service)

SaaS provides the third party (Vendor) with the software. Vendor pays for the time of using the software and can use it anywhere there is no need to buy the software. It is referred to as “on-demand” software. SaaS eliminates the need for organizations to install and use applications on their own computers or in their own data centers. This removes the budget of hardware acquisition, provisioning and maintenance, as well as software licensing, installation and support.

Layer: Application layer

E.g: Google Apps, Salesforce.com

b) PaaS (Platform as a Service)

It provides a platform where assets (resources) are available and consumers can themselves create the required application for required application for e.g. web application data and database data. Providers provide network servers and storage services.

Layer: Platform layer

E.g: Google App Engine, Microsoft Azure

c) IaaS (Infrastructure as a Service or Hardware as a Service)

Infrastructure as a Service (IaaS) is a piece of cloud computing that provides virtualized computing resources through the Internet. This model assists users with the hardware on rent for e.g. Server space, network equipment, memory, storage space. In short, consumers buy virtual space and work on it.

Layer: Hardware Layer

E.g: Amazon Ec2, Simple Storage Services (S3)

II. LOAD BALANCING

Load Balancing is a system to distribute the capacity evenly among all the nodes of the network. If any node is heavy i.e. have more capacity than required then its load is given to the node with less load. Hence load balancing helps the over-whelm and under-loaded nodes. Load balancing [5] is a major challenge of cloud computing. Goals of load balancing involve:

1. Optimum resources utilization
2. Maximum throughput
3. Maximum response time
4. Avoiding overload

Load balancing can be static or dynamic

a) Static load balancing: The load does not depend on the current state of the system

But it requires knowledge about the application and resources of the system in this; load is distributed evenly across all the nodes. This algorithm must know about the system resources beforehand. There are four different types of Static load balancing techniques: Round Robin algorithm, Central Manager Algorithm and randomized algorithm.

b) Dynamic load balancing: It depends on the present state of the system. If any node is overloaded then its load is shifted to the under-loaded node. So real-time task is performed here. The Dynamic algorithm is more flexible than

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 9, September 2016

the static algorithm and they don't rely on prior knowledge. In a distributed system, dynamic load balancing has two different ways: distributed and non-distributed. In the distributed one, this algorithm is executed by all nodes present in the system and the task of load balancing is shared among these servers. The communication among nodes to achieve load balancing can take two forms: cooperative also non-cooperative. In the first one, the nodes work side-by-side to achieve a common objective which means is to raise the overall response time, etc. In the second form, each node works independently toward a goal local to it.

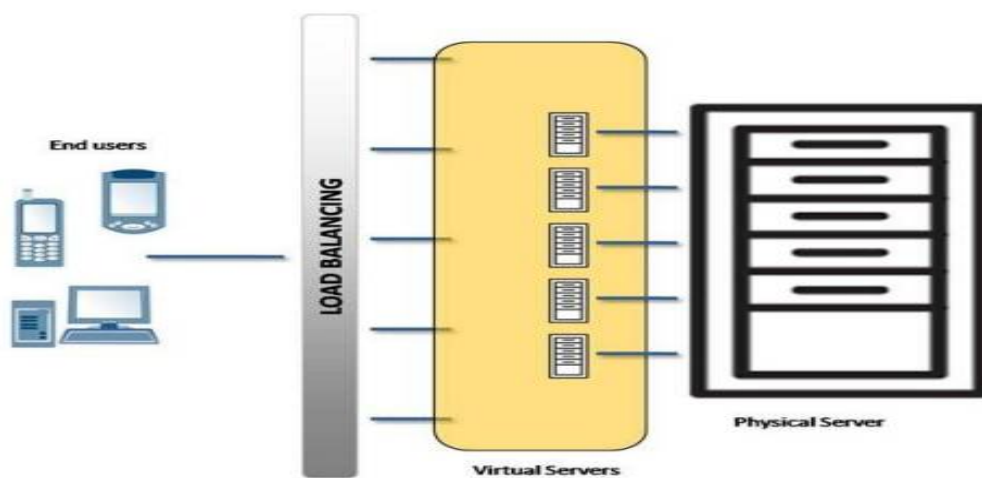


Fig.2: Load Balancing in Cloud Computing

Policies followed in dynamic load balancing are:

Following are the policies of load balancing [5]:

-*Transfer Policy*: The use of the dynamic load balancing algorithm which selects a job for transferring from a local node to a remote node is assigned to as Transfer policy or Transfer strategy.

-*Selection Policy*: It specifies the processors associated with the load transfer (processor matching)

- *Location Policy*: The section of the load balancing algorithm which selects a target node for a transferred task is referred to as location policy or Location strategy.

-*Information Policy*: The sector of the dynamic load balancing algorithm responsible for collecting information about the nodes in the system is referred to as Information policy as an alternative Information strategy.

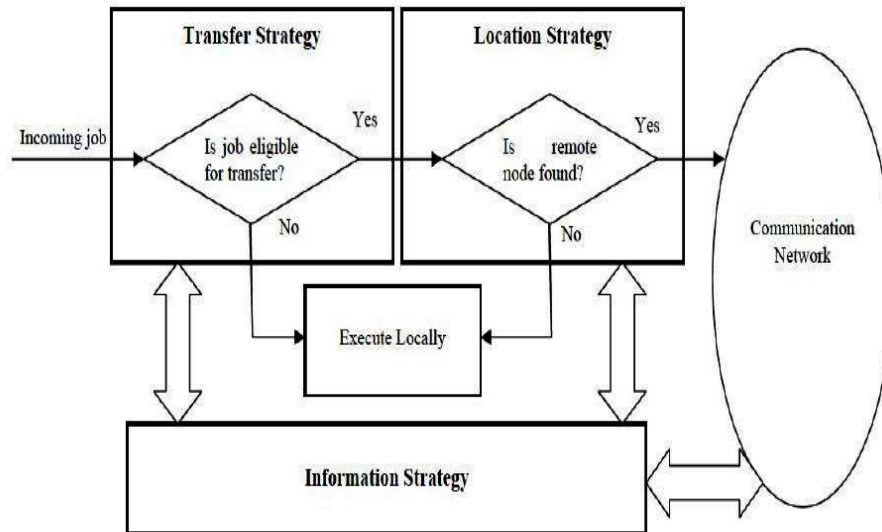


Fig.3 Policies Implementation

III. LOAD BALANCING TECHNIQUES

There are various techniques [1],[3],[4],[6] to balance the load of cloud computing. Some of which are discussed in this paper.

a) Honey Bee Foraging

A colony of honey bee can extend itself over long distances as to find many food sources such as flower patches and then these bees' harvests nectar or pollen from these sources. A small fraction of the colony finds the environment looking for new flower patches. When the food source is encountered the scout bees go in the field surrounding the hive and check for quality beneficial. When they return to the hive, the scouts collect the food harvested. There is an area in the hive called as the "dance floor", where waggle dance is performed by the bees that found a very beneficial food. Through the waggle dance, a scout bee passes the position of its search to an idle spectator, which helps in the using of the flower patch. Here the duration of the dance is according to the scout's rating of the food source, to harvest the best-rated flower patches more foragers get recruited. When dance is done, the scout return to the food source it found to see more food. Till the food is profitable, food sources will be posted by the scouts when they return to their hive. Foragers who are recruited recently may waggle dance as well, which will step-up the recruitment for highly profitable flower patches.

b) Task Scheduling Algorithm based on Load Balancing

Y. Fang et al.[10] discussed a 2-level task scheduling mechanism based on load balancing to meet dynamic requirements of users and obtain high resource usage. It achieves load balancing by first mapping tasks to virtual machines and then virtual machines to host resources.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 9, September 2016

c) Throttled Load Balancing Algorithm

This algorithm makes use of the existence of virtual machines. Client requests the ID of virtual machine. Throttled load balancing algorithm returns that ID to the user. In this algorithm, the load balancer maintains an index table of virtual machines as their states (Available or Busy). The client/server first makes a request to data centre to find a proper virtual machine (VM) to perform the recommended job. The data centre queries the load balancer for allocation of the VM. The load balancer searches the index table from the top until the first available VM is found or the index table is scanned fully. If the VM is found, the load data centre.

d) Ant Colony Optimization Technique

In this approach, a pheromone table was being designed which was updated by ants as per the resource utilization and node selection blueprint. Ants move in the forward direction in search of the overloaded or under loaded node. As the overburden node is traversed, then ants move back to fill the recently encountered under loaded node, so a single table is updated every time.

e) Role-Based Access Control (RBAC)

RBAC is a method used to reduce the load of the cloud. In this, a role is given to each user so that limited applications of the cloud can be accessed by their corresponding number of users. So by this approach, the resources are restricted to the users.

f) Resource Allocation Scheduling Algorithm (RASA)

In this way, virtual nodes are created first. Then the expected response time of each virtual node is found. Then view to the least loaded node criteria, the efficient virtual node is found and ID of that node is returned to the client. In this, Min-Min and Max-Min strategies are pursued. If a number of resources available are odd, then Min-Min strategy is applied else Max-Min strategy is applied.

Table 1: Comparison of the above discussed techniques

Load Balancing Methods	Parameters	Metrics	Demerits
Honeybee Foraging Technique	Execution Time, Overheads, Throughput	Maximize throughput, Low overheads	Low Priority Load
Task Scheduling Algorithm based On Load Balancing Algorithm	Response Utilization, Performance, Response to Request Ratio	Maximize Response Utilization, Increased performance	Doesn't improve response to request ratio
Throttled Load Balancing Algorithm	Load Movement Factor, Communication Cost, Network Delay	High Load Movement Factor	High Communication Cost, High Network Delay

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 9, September 2016

Ant Colony Optimization	Performance , Resource Utilization, Fault Tolerance, Scalability	Performance increased, Resource Utilization high, Fault Tolerance excellent, Scalability Good	Complex Network
RBAC	Performance, Resource Utilization, Energy, CPU Burst Time	High Performance, High Resource Utilization, CPU Burst time decrease	Response Time increases
RASA	Performance, Execution Time	Performance increases, Execution time Decreases	Less Fault Tolerance

IV. OBJECTIVE

In this paper we surveyed various load balancing techniques for cloud computing. The main purpose of load balancing is to satisfy the customer requirement by allocating load dynamically among the nodes and to make maximum resource utilization by re- assigning the total load to the individual node. This ensures that every resource is distributed efficiently, and evenly. So the performance of the system is increased.

V. CONCLUSION

As the Cloud Computing is the charming concept in the present and upcoming time, the researchers had developed various methods to cope up with the load balancing problem being faced while working with cloud computing. In this paper a light is being thrown on the recent methods being used. And comparison is done between them based on various parameters. Future work can be done to balance the load by distributed policy so as to reduce the workload of the executer.

REFERENCES

- [1] R.G.Rajan ,V.Jeyakrishnan, "A Survey on Load Balancing in Cloud Computing Environments", InternationalJournal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 12, December 2013.
- [2] P. Salot, "Asurvey of various scheduling algorithm in cloud computing environment", IJRET, Volume: 2 Issue: 2, Feb 2013.
- [3] A.K. Sidhu, S. Kingar, "Analysis of Load Balancing Techniques in Cloud Computing", International Journal of Computers & Technology, Volume 4 No. 2, ISSN 2277-3061, March-April, 2013.
- [4] S. MohanaPriya, B. Subramani," A new approach for load balancing in cloud computing", International Journal of Engineering and Computer Science, ISSN: 2319-7242 Volume 2 Issue 5 May, 2013 Page No. 1636-1640.
- [5] Suresh M., ShafiUllah Z., Santhosh Kumar B.," An Analysis of Load Balancing in Cloud Computing", International Journal of Engineering Research & Technology (IJERT),Vol. 2 Issue 10, October – 2013, ISSN:2278-0181
- [6] A. Kaur, N. Bansal, "Cross-Breed Job Scheduling for Reducing the Server Load Using RBAC at Cloud", International Journal of Advanced Research in Computer Science and Software Engineering", Volume 3, Issue 5, ISSN: 2277 128X, May 2013
- [7] N. J. Kansal, I. Chana," Cloud Load Balancing Techniques: A Step towards Green Computing", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 1, January 2012, ISSN (Online): 1694-0814
- [8] N. Sran, N. Kaur, "Comparative Analysis of Existing Load Balancing Techniques in Cloud Computing",International Journal of Engineering Science Invention ,Volume 2 Issue 1, PP.60-63,ISSN (Online): 2319 – 6734,ISSN (Print): 2319 – 6726, January. 2013.
- [9] P. Mell,T.Grance, "The NIST Definition of Cloud Computing (Draft)"
- [10] Y. Fang, F. Wang, and J. Ge, "A Task Scheduling Algorithm Based on Load Balancing in Cloud Computing", Web Information Systems and Mining, Lecture Notes in Computer Science, Vol. 6318, pages 271-277,2010.
- [11] Geethu Gopinath P P, Shriram K Vasudevan, "An in-depth analysis and study of Load balancing techniques in the cloud computing environment" Procedia Computer Science 50 (2015) 427 – 432.
- [12] Tran Cong Hung, Nguyen Xuan Phi, "Study the Effect of Parameters to Load Balancing in Cloud Computing", International Journal of Computer Networks & Communications (IJCNC) Vol.8, No.3, May 2016