ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

# SURVEY: CLOUD PARTITIONING USING LOAD BALANCING APPROACH FOR PUBLIC CLOUD INFRASTRUCTURE

**Rajesh Kumar*, Charanjit Singh**

* Research Scholar, Computer Science and Engineering Department, Maharishi Markandeshwar University, Sadopur, Ambala, Haryana, India

Assistant Professor, Computer Science and Engineering Department, Maharishi Markandeshwar University, Sadopur, Ambala, Haryana, India

## ABSTRACT

The term cloud computing describes distributed computing, virtualization, software, networking and web services having several elements such as clients, data enter and distributed servers etc. This includes fault tolerance, high availability, scalability, flexibility, reduced overhead for users, reduced cost of ownership, on demand services etc. is the process of distributing the load among various nodes of a distributed system to improve both resource utilization and job response time while avoiding a situation where some of the nodes are heavily loaded while other nodes are idle or doing very little work is called as Load balancing. Here in this paper, work on highly robust load balancing or job assignment approach has been discussed for distributed cloud system on the basis of game theory. This paper, is emphasized to incorporate dynamic load based job allocation approaches with the goal to reduce execution time in job scheduling.

**KEYWORDS**: Cloud Computing, Cloud Partitioning, Game Theory, Load Balancing, Public Cloud Infrastructure.

## INTRODUCTION

Since the job arrival pattern in cloud infrastructure is not predictable and the capacities of each node in the cloud differ, for load balancing problem, workload control is crucial to improve system performance and maintain stability. In case of Cloud computing services can be used from diverse and widespread resources, rather than remote servers or local machines. There is no standard definition of Cloud computing. Generally it consists of a bunch of distributed servers known as masters, providing demanded services and resources to different clients known as clients in a network with scalability and reliability of data center. Users get service from a cloud without paying attention to the details [1]. NIST gave a definition of cloud computing as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [2]. More and more people pay attention to cloud computing [4, 5]. Cloud computing is efficient and scalable but maintaining the stability of processing so many jobs in the cloud computing environment is a very complex problem with load balancing receiving much attention for researchers. The load can be CPU load, memory capacity, delay or network load. Load balancing is the process of distributing the load among various nodes of a distributed system to improve both resource utilization and job response time while also avoiding a situation where some of the nodes are heavily loaded while other nodes are idle or doing very little work.

## CLOUD COMPUTING

### *Principles and Cloud Components*

Cloud computing is having the principles as Resource Pooling, Virtualization, Elasticity, Automatic/Easy Resource Deployment, Metered billing. Clients, Datacenter and Distributed Servers comprising to makes a cloud system. These can be shown with the help of diagrams as depicted in Fig 1.1 below. And also how these devices/systems connected with each other. Clouds can be categorized into 3 types named as Public Clouds, Private Clouds and Hybrid Clouds (combination of both private and public clouds).

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

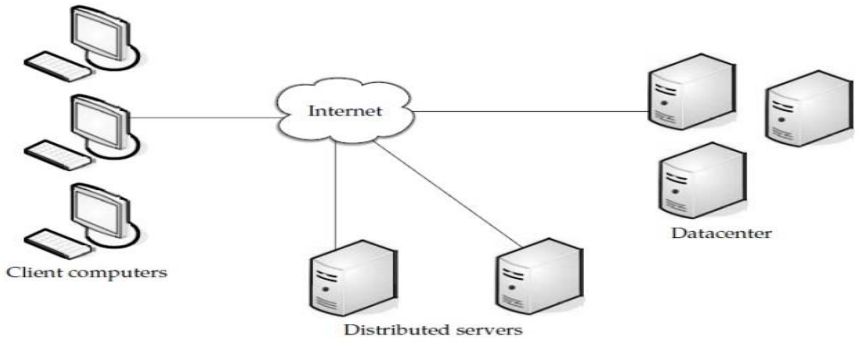


***Figure 1.1: Three components make up a cloud computing solution.***

***Services provided by Cloud computing***

Service means different types of applications provided by different servers across the cloud. It is generally given as "as a service". Services in a cloud are of 3 types as given in [20]:

- Software as a Service (SaaS)
- Platform as a Service (PaaS)
- Hardware as a Service (HaaS) or Infrastructure as a Service (IaaS)

***Load Balancing***

Load Balancing is a process of reassigning the total load to make resource utilization effective and to improve the response time of the job. The important things to consider while developing these algorithm are: estimation of load, comparison of load, stability of different system, performance of system, interaction between the nodes [21].

***Goals of Load balancing***

As given in [21], the goals of load balancing are: To improve the performance substantially, To have a backup plan in case the system fails even partially, To maintain the system stability, To accommodate future medication in the system.

***Types of Load balancing algorithms***

Depending on who initiated the process, load balancing algorithms can be of three categories as given in [21]:

- Sender Initiated: If the load balancing algorithm is initialized by the sender
- Receiver Initiated: If the load balancing algorithm is initiated by the receiver

Symmetric: It is the combination of both senders initiated and receiver initiated Depending on the current state of the system, load balancing algorithms can be divided into 2

- Static: It doesn't depend on the current state of the system. Prior knowledge of the system is needed
- Dynamic: Decisions on load balancing are based on current state of the system. No prior knowledge is needed. So it is better than static approach.

**DYNAMIC LOAD BALANCERS**

***Equally Spread Current Execution***

This algorithm distributes the load randomly by first checking the size of the process and then transferring the load to a Virtual Machine, which is lightly loaded [22].

***Throttled Load Balancer***

In this algorithm, the client first requests the load balancer to find a suitable Virtual machine to perform the required operation [22].

***Honeybee Foraging Algorithm***

The main idea behind the Honeybee Foraging algorithm [23] is derived from the behavior of honeybees. In this case, the servers are grouped as virtual server and each virtual server have a process queue. Each server, after processing a request from its queue, calculates the profit which is analogous to the quality that the bees show in their waggle dance.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

If profit is high, the server stays else, it returns to the forage. This algorithm requires that each node to maintain a separate queue. This computation of profit on each node causes additional overhead.

The disadvantage of this algorithm is that, it does not show any significant improvement in throughput, which is due to the additional queue and the computation overhead.

***Biased Random Sampling***

Biased Random Sampling is a dynamic load balancing algorithm. It uses random sampling of system domain to achieve self-organization thus, balancing the load across all nodes of system. In this algorithm, a virtual graph is constructed with the connectivity of each node representing the load on server. Each node is represented as a vertex in a directed graph and each in-degree represents free resources of that node.

In this algorithm, upon receiving the request by the load balancer, it would select a node randomly and compares the current walk length with the threshold value. If the current walk length is equal to or greater than the threshold value, the job is executed at that node. Else, the walk length of the job is incremented and another neighbor node is selected randomly. The performance is degraded as the number of servers increase due to additional overhead for computing the walk length.

***Active Clustering***

Active Clustering [40, 41] is a clustering based algorithm which introduces the concept of clustering in cloud computing. The process of creating a cluster revolves around the concept of match maker node. In this process, first node selects a neighbor node called the matchmaker node which is of a different type. This matchmaker node makes connection with its neighbor which is of same type as the initial node. Finally the matchmaker node gets detached. This process is followed iteratively. The performance of the system is enhanced with high availability of resources, thereby increasing the throughput. This increase in throughput is due to the efficient utilization of resources.

## ANALYSIS OF LOAD BALANCING TECHNIQUES IN CLOUD COMPUTING

Cloud Computing is an emerging computing paradigm. It aims to share data, calculations, and service transparently over a scalable network of nodes. Since Cloud computing stores the data and disseminated resources in the open environment. So, the amount of data storage increases quickly. In the cloud storage, load balancing is a key issue. It would consume a lot of cost to maintain load information, since the system is too huge to timely disperse load. Load balancing is one of the main challenges in cloud computing which is required to distribute the dynamic workload across multiple nodes to ensure that no single node is overwhelmed. It helps in optimal utilization of resources and hence in enhancing the performance of the system. A few existing scheduling algorithms can maintain load balancing and provide better strategies through efficient job scheduling and resource allocation techniques as well. In order to gain maximum profits with optimized load balancing algorithms, it is necessary to utilize resources efficiently. This paper discusses some of the existing load balancing algorithms in cloud computing and also their challenges.

**A Game-Theoretic Model and Algorithm for Load Balancing in Distributed Systems**

***Distributed Load Balancing Model for Grid Computing***

Most of the existing load balancing strategies was interested in distributed systems which were supposed to have homogeneous resources interconnected with homogeneous and fast networks. For Grid computing, these assumptions are not realistic because of heterogeneity, scalability and dynamicity characteristics. For these environments the load balancing problem is then a new challenge presently for which many research projects are under way. The proposed strategy is naturally distributed with a local decision, which allows the possibility of avoiding use of wide area communication network.

***Game theory and information systems***

The internal consistency and mathematical foundations of game theory make it a prime tool for modeling and designing automated decision-making processes in interactive environments. For example, one might like to have efficient bidding rules for an auction website, or tamper-proof automated negotiations for purchasing communication bandwidth. Research in these applications of game theory is the topic of recent conference and journal papers (see, for example, Binmore and Vulkan, "Applying game theory to automated negotiation," Netnomics Vol. 1, 1999, pages 1–9) but is still in a nascent

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

stage. The automation of strategic choices enhances the need for these choices to be made efficiently, and to be robust against abuse. Game theory addresses these requirements.

As a mathematical tool for the decision-maker the strength of game theory is the methodology it provides for structuring and analyzing problems of strategic choice. The process of formally modeling a situation as a game requires the decision-maker to enumerate explicitly the players and their strategic options, and to consider their preferences and reactions. The discipline involved in constructing such a model already has the potential of providing the decision-maker with a clearer and broader view of the situation. This is a "prescriptive" application of game theory, with the goal of improved strategic decision making. With this perspective in mind, this article explains basic principles of game theory, as an introduction to an interested reader without a background in economics.

## LITERATURE SURVEY

Ranjan Kumar et al. [8] implemented ant colony optimization method and considered the normal performance of ants such as incisive for job, communication by every past and parallel colonial entity etc. This method came out with certain better results but could not deliver the optimum results for an efficient improvement for public clouds along with energetic training scheme. Suriya et al. [9] demonstrated a variety of feature pertaining to province of cloud computing, its development, its general problem, as well as predominantly to problem connected to load balancing. They also described an assortment of approaches advocated earlier. This investigation couldn't lane its inspection for additional improvement and optimization for load assessment for cloud communications.

Yu-lung Lo et al. [10] supported a method for database portability as well as its load evaluation in cloud request and they projected 5 database distribution algorithms for allocated databases for exhibiting data hosting in cloud infrastructure. The equations employed for estimating the variation of database portion. This proposed system was oriented towards presented database portability and access in cloud with scheduled way, but in fact it could not broaden it vision to consider public cloud infrastructure with its functional procedure. So, it cannot be considered optimum for public cloud implementation.

Supriya Kinger et al. [11] talk about numerous factors in cloud infrastructures such as Load Balancing across cloud network, Virtual Machine (VMs) Migration, consolidation of servers in cloud, management of power factor in cloud operation, etc. Main propel has been specified on the learning of load balancing algorithm, pursued by a comparative review of these on top of talk about algorithms in cloud computing with respect to potential cloud factors such as scalability, resource utilization, cloud presentation, variation or response time with computational overhead associated.

Meriem Meddeber et al. [12] advocated a system with dual systems modes where the first modes proposes a distributed kind of load balancing model that ultimately presents any Grid topology in a firm of forest approach or architecture. This was then followed by a load balancing strategy with predominant goal of reduction or minimization in mean execution span in job scheduling and associated cost factor.

Venubabu Kunamneni et al [13] advocated for dynamic load balancing scheme for a cloud infrastructure that facilitates a brief of various topic and system model on load balancing, static load balancing, software as well as hardware oriented load balancing etc. In fact Dynamic load balancing is necessary for a cloud performance enhancement therefore this work cannot be an idle work for future utility.

N. G. Shivaratri et al. [14] emphasized on the issue of fair and optimum load or jobs/task scheduling while exhibiting redistribution of load in the system among its nodes so that the largely performance of the system could be enhanced. Numerous factors such as issues being faced with load distributing in generic cloud infrastructures, encompassing the various motivating factors and numerous design trade-offs for load-distributing algorithms were analyzed and discussed. Additionally a number of load-distributing algorithms were advocated with their respective performances. The discussed approaches were of predominant categories such as sender-initiated algorithms, receiver-initiated algorithms, symmetrically initiated approaches and adaptive kind of scheduling approach.

Chronopoulos et al. [15] presented a game theoretic scheme for solving the issues associated with static load balancing for both the classes; single-class as well as multi user jobs in certain distributed system where the functional entities such as computers are joined together using certain communication media. The objective of this work was to facilitate fairness to all

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

the tasks in case of a single-class system and the various user entities in jobs for multi-user system. For facilitating fairness to all the comprising tasks in the system model they implemented a cooperative game to model the load balancing.

D. Grosu et al. [16] developed the static load balancing issue with single class job distributed framework using a cooperative game theoretic approach amongst various configured computer systems. They even considered Nash Bargaining Solution (NBS) that facilitates a Pareto optimal job scheduling scheme that is functional fairly with all tasks. The Author then developed a cooperative load balancing technique of scheduling for computing NBS.

M. Randles et al. [17] developed a model for cloud infrastructure developed on well-established Web Services, networks, application estimation and its virtualization, with a goal to come up with certain advantages such as minimum computational costs, system flexibility and resource accessibility for service users. The retrieved advantages are expected for further impelled demand in cloud applications, rising both the Cloud's customer base as well as the scale of Cloud installations. They emphasized on the establishment of certain robust load balancing techniques. In their work they investigated few solutions for distributed systems while taking into account of load balancing. They employed three system solutions using Honeybee Foraging characteristics, biased random sampling approach and then ultimately an active clustering approach was advocated.

Xuejie Zhang et al. [18] developed a load balancing approach on the basis of certain bio-inspired optimization approach called ant colony optimization with a noble complex network theory for public cloud computing requirements. The implementation of this approach with ant colony emerged out with numerous enhancements such as fair load scheduling and reduced execution time. Their developed scheme took benefits of the behavioural advantages of Complex Network. Ultimately, they exhibited their system performance with factors such as qualitative analysis and quantities analysis in terms execution times, data processing etc.

Considering the above mentioned manuscript it can be found that a number pf works have been done for cloud computing and especially for load balancing and system optimization.

## EXISTING SYSTEM AND ITS LIMITATAIONS

Since the job arrival pattern is not predictable and the capacities of each node in the cloud differ, for load balancing problem, workload control is crucial to improve system performance and maintain stability. Load balancing schemes depending on whether the system dynamics are important can be either static or dynamic. However, load balancing in the cloud is still a new problem that needs new architectures adapt to many changes. Chaczko et al [5] described the role that load balancing plays in improving the performance and maintaining stability. There are many load balancing algorithms, suchas Round Robin, Equally Spread Current Execution Algorithm, and Ant Colony algorithm. Nishant et al [6] used the ant colony optimization method in nodes load balancing. Randles et al [7] gave a compared analysis of some algorithms in cloud computing by checking the performance time and cost. They concluded that the ESCE algorithm and throttled algorithm are better than the Round Robin algorithm. Some of the classical load balancing methods is similar to the allocation method in the operating system, for example, the Round Robin algorithm and the First Come First Served (FCFS) rules

Limitations: Few of the limitations are as follows: Cloud computing environment is a very complex problem with load balancing receiving. The job arrival pattern is not predictable and the capacities of each node in the cloud differ, for load balancing problem, workload control is crucial to improve system performance and maintain stability. The Existing system is static so the load balancing technique is not supporting for dynamic Operations among‘N’ number of Cloud Partions in the same cloud or different cloud. There is no cloud partition for load balancing in the existing system. Static schemes will take more system configurations and huge process time. If the system is idle or busy then it will take more time to come to its original position (Normal Position). The existing system uses the Round Robin algorithm which will take more time and cost to allocate the cloud and to maintain the cloud storage.

Considering the present day scenario of public cloud utilities it is very significant to develop an efficient cloud partition based load balancing approach employing dynamic scheduling. The enhancement by security features can be a vital achievement for public cloud infrastructures.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

## CONCLUSION

In this paper, reviews from various authors have given in regards with cloud computing for good load balancing. Various techniques as well as methodologies are given their with their studies and results from their experiments. Good load balance will improve the performance of the entire cloud. However, there is no common method that can adapt to all possible different situations. Various methods have been developed in improving existing solutions to resolve new problems. Each particular method has advantage in a particular area but not in all situations. Therefore, the current study integrates several methods and switches between the load balance methods based on the system status. A relatively simple method can be used for the partition idle state with a more complex method for the normal state. The load balancers then switch methods\as the status changes.

## REFERENCES

[1] M. D. Dikaiakos, D. Katsaros, P. Mehra, G. Pallis, and A. Vakali, "Cloud computing: Distributed internet computing for IT and scientific research", Internet Computing, vol.13, no.5, pp.10-13, Sept.-Oct. 2009.
[2] P. Mell and T. Grance, "The NIST definition of cloud computing" http://csrc.nist.gov/ publications/nistpubs/800- 145/SP800-145.pdf, 2012.
[3] Microsoft Academic Research, Cloud computing, http:// libra.msra.cn/Keyword/6051/cloud-computing? query=cloud%20computing, 2012.
[4] Google Trends, Cloud computing, http://www.google.com/trends/explore#q=cloud%20computing, 2012.
[5] M. Randles, D. Lamb, and A. Taleb-Bendiab, A comparative study into distributed load balancing algorithms for cloud computing, in Proc. IEEE 24th International Conference on Advanced Information Networking and Applications, Perth, Australia, 2010, pp. 551-556.
[6] Wuhib, F.; Stadler, R.; Lindgren, H., "Dynamic resource allocation with management objectives—Implementation for an OpenStack cloud," Network and service management (cnsm), 2012 8th international conference and 2012 workshop on systems virtualiztion management (svm) , vol., no., pp.309,315, 22-26 Oct. 2012
[7] Hui Zhang; Guofei Jiang; Yoshihira, K.; Haifeng Chen; Saxena, A., "Intelligent Workload Factoring for a Hybrid Cloud Computing Model," Services - I, 2009 World Conference on , vol., no., pp.701,708, 6-10 July 2009
[8] Ranjan Kumar and G Sahoo "Load Balancing Using Ant Colony in Cloud Computing" International Journal of Information Technology Convergence and Services (IJITCS) Vol.3, No.5, October 2013
[9] Suriya Begum, Dr. Prashanth C.S.R "Review of Load Balancing in Cloud Computing" IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 2, January 2013 ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814
[10] Yu-lung Lo and Min-Shan Lai "The Load Balancing of Database Allocation in the Cloud" Proceedings of the International Multi Conference of Engineers and Computer Scientists 2013 Vol I, IMECS 2013, March 13 - 15, 2013, Hong Kong
[11] Amandeep Kaur Sidhu, Supriya Kinger "Analysis of Load Balancing Techniques in Cloud Computing" International Journal of Computers & Technology Volume 4 No. 2, March-April, 2013, ISSN 2277-3061
[12] Belabbas Yagoubi, Meriem Meddeber "Distributed Load Balancing Model for Grid Computing" Revue ARIMA, vol. 12 (2010), pp. 43-60
[13] Venubabu Kunamneni "Dynamic Load Balancing for the Cloud" International Journal of Computer Science and Electrical Engineering (IJCSEE) ISSN No. 2315-4209, Vol-1 Iss-1, 2012
[14] N. G. Shivaratri, P. Krueger, and M. Singhal, Load distributing for locally distributed systems, Computer, vol. 25, no. 12, pp. 33-44, Dec. 1992.
[15] S. Penmatsa and A. T. Chronopoulos, Game-theoretic static load balancing for distributed systems, Journal of Parallel and Distributed Computing, vol. 71, no. 4, pp. 537-555, Apr. 2011.
[16] D. Grosu, A. T. Chronopoulos, and M. Y. Leung, "Load balancing in distributed systems: An approach using cooperative games", in Proc. 16th IEEE Intl. Parallel and Distributed Processing Symp., Florida, USA, Apr. 2002, pp. 52-61.
[17] M. Randles, D. Lamb, and A. Taleb-Bendiab, "A comparative study into distributed load balancing algorithms for cloud computing", in Proc. IEEE 24th International Conference on Advanced Information Networking and Applications, Perth, Australia, 2010, pp. 551-556.
[18] Zehua Zhang, Xuejie Zhang "A load balancing mechanism based on ant colony and complex network theory in open cloud computing federation" Industrial Mechatronics and Automation (ICIMA), 2010 2nd

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

International Conference on (Volume:2) Page(s): 240 – 243
[19] Martin Randles, David Lamb, A. Taleb-Bendiab, A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing, 2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops.
[20] Anthony T.Velte, Toby J.Velte, Robert Elsenpeter, Cloud Computing A Practical Approach, TATA McGRAW-HILL Edition 2010.
[21] Ali M. Alakeel, A Guide to Dynamic Load Balancing in Distributed Computer Systems, IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.6, June 2010.
[22] Nitika, Shaveta, Gaurav Raj, International Journal of advanced research in computer engineering and technology Vol-1 issue-3 May-2012
[23] Yatendra sahu, M. K. Pateriya "Cloud Computing Overview and load balancing algorithms", Internal Journal of Computer Application Vol-65 No.24, 2013.
[24] Nayandeep Sran, Navdeep kaur "Comparative Analysis of Existing Load balancing techniques in cloud computing", International Journal of Engineering Science Invention, Vol -2 Issue-1 2013
[25] Nidhi Jain Kansal, Inderveer Chana "Existing Load balancing Techniques in cloud computing: A systematic review" Journal of Information system and communication Vol-3 Issue-1 2012