A Novel Approach for Dynamic Selection of Load Balancing Algorithms in Cloud Computing: Survey

Umang Thakkar, Indrjeet Rajput
Department of Computer Engineering,
Hasmukh Goswami College of Engineering, Vahelal, Gujarat Technological University

***Abstract*- Cloud computing is the most recent advanced emerging technology that provides every kind of service to its users. It generally based on "pay per usage" terminology that provides Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) model that altogether form a cloud. However it also has some issues that is to be resolved. One of the major issue for cloud computing is Load Balancing. Load Balancing is the technique to divide the load among different available resources and equalize with virtual machines to achieve efficiency and increases throughput. This paper first of all shows the theoretical study of load balancing and then shows different types of load balancing algorithms used to balance load. Further it also demonstrates and compares algorithms that gives an idea to merge them and have a choice to select dynamically based on some parameters and situation in which they are better for enhancing efficiency.**

***Index Terms*- Cloud Computing, Deployment Models, Service Models, Challenges, Load Balancing and its algorithms.**

I. INTRODUCTION

Cloud computing is one of the hottest technical topics today, with broad-ranging effects across IT, Information Architecture, Business, Software Engineering, and Data Storage. Cloud Computing is an innovative technology that is revolutionizing the way it does computing. The key concept of cloud computing is that you don't buy the hardware, or even the software, you need anymore, rather you rent some computational power, storage, databases, and any other resource you need by a provider according to a pay-as-you-go model, making your investment smaller and oriented to operations rather than to assets acquisition. It is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (eg networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [1]. There are several different issues of cloud computing that last from years ago, one of that issue is load balancing. Cloud is very high flexible and provides great retrieval of data as per user's requirements and other services, apart from this to handle large amount of data several techniques to optimize load is workload which is total time required by the processor to execute all assigned processes [2]. To solve the issue of load balancing or task scheduling, we classified it into two categories namely static and dynamic. Static algorithms are mostly suitable for homogenous and stable environments and can produce good results. Dynamic algorithms are more flexible and take into consideration different types of attributes in the system both prior to and during run-time [3]. Cloud creates virtual machines on demand of user as per requirements and many times changes its needs rapidly resources like CPU, RAM, storage, bandwidth etc, so to overcome this issues load balancing technique is feasible that distributes workload from overload host to other lightly loaded host to improve overall performance of cloud system. For better efficiency, dynamic compare and balanced algorithms work on threshold values to optimize the cloud data center and balance cloud machines and reduce operational cost of cloud applications [4]. Load balancing is done by load balancers and to make efficient cloud, many load balancing algorithms like ant colony and honey bee are optimized under normal balancer conditions. If machine is idle it performs round-robin algorithm and if machine is normal it performs ant colony or honey bee behavior algorithm whichever suited based on calculation of execution time of task arrived. Central controller that chooses suitable partitions for arriving jobs [5].

II. THEORETICAL REVIEW

The concept of Cloud Computing came into existence in the year 1950 with implementation of mainframe computers, accessible via thin/static clients. Since then,

cloud computing has been evolved from static clients to dynamic ones and from software to services. The term Cloud refers to a Network or Internet. In other words, we can say that Cloud is something, which is present at remote location. Cloud can provide services over public and private networks, i.e., WAN, LAN or VPN. Cloud Computing refers to manipulating, configuring, and accessing the hardware and software resources remotely. It offers online data storage, infrastructure, and application.

Cloud computing offers platform independency, as the software is not required to be installed locally on the PC. Hence, the Cloud Computing is making our business applications mobile and collaborative.

Cloud computing, also known as on-demand computing, is a kind of Internet-based computing, where shared resources, data and information are provided to computers and other devices on-demand. It is a model for enabling ubiquitous, on-demand access to a shared pool of configurable computing resources.

Cloud computing mainly consists of two types of working models: Deployment models and Service models. Deployment model defines the type of access to the cloud i.e. Public, Private, Hybrid, and Community. Service model categorized into three basic service: Infrastructure as a Service, Platform as a Service, Software as a Service.

Apart from this, cloud computing face different kind of challenges out of which load balancing is the major issue to solve for better efficiency and throughput. Load balancing in cloud computing provides an efficient solution to various issues residing in cloud computing environment set-up and usage. Load balancing must take into account two major tasks, one is the resource provisioning or resource allocation and other is task scheduling in distributed environment. Efficient provisioning of resources and scheduling of resources as well as tasks will ensure: [1]

a. Resources are easily available on demand.
b. Resources are efficiently utilized under condition of high/low load.
c. Energy is saved in case of low load (i.e. when usage of cloud resources is below certain threshold).
d. Cost of using resources is reduced.

III. DEFINITIONS AND TERMINOLOGIES RELATED TO CLOUD COMPUTING AND LOAD BALANCING

3.1 Cloud:- Cloud is a collection of parallel and distributed systems. It is not the combination of cluster and grid but its next generation to clusters and grid. [1]

3.2 Cloud Computing :- Cloud Computing is a scenario where the user can have access to any kind of infrastructure, software or platform in a secure manner- at reduced cost on-demand basis-in an easy to use manner.[1]

3.3 Infrastructure as a Service :- Infrastructure-as-a-Service provides access to fundamental resources such as physical machines, virtual machines, virtual storage, etc. Infrastructure as a service provides companies with computing resources including servers, networking, storage and data center space on a pay-per-use basis.

3.4 Platform as a Service :- Platform-as-a-Service offers the runtime environment for applications. It also offers development and deployment tools required to develop applications. PaaS has a feature of point-and-click tools that enables non-developers to create web application. App Engine of Google and Force.com are examples of PaaS offering vendors.

3.5 Software as a Service :- Software-as–a-Service (SaaS) model allows to provide software application as a service to the end users. It refers to a software that is deployed on a host service and is accessible via Internet.

3.6 Load Balancing :- Load balancing is technique to distribute job or task or load to different nodes to execute the process on shortest time on distributed network. [1]

IV. ALGORITHMS FOR LOAD BALANCING

As Load Balancing is an essential task in cloud computing to achieve maximum utilization of resources, several types of algorithms are developed to reduce response time and increase availability of resources for provisioning and task scheduling. Among all different load balancing algorithms working in dynamic environment, some algorithms are better and improved and results in better performance. Our approach is to combine best algorithms that are fit in different conditions of load balancing.

4.1. A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment. [1]

According to this study paper, Cloud computing is collection of parallel and distributed systems that has multiple domains and hardware is virtualized and provisioned on-demand to end users as pay-per-usage model. Load Balancing is one of the issue of cloud computing. Several working environments under load balancing to make efficient load balancer based on Service Level Agreement (SLA). Load Balancing algorithms are been discussed in different environments.

Advantages:

- Compares Cloud vs Cluster and Grid.
- Proposed two types of load balancing algorithms named Static and Dynamic environment based on cloud.
- Algorithms in Static environment are not flexible, better for homogeneous resource provisioning and requires prior knowledge of nodes.
- Algorithms in Dynamic environment are flexible, better for heterogeneous resource provisioning and does not require prior knowledge of nodes as it adapts runtime changes.
- Comparision shown between different distribution nodes with pros and cons.

Disadvantages:

- Static Load Balancing scheme fails in heterogeneous nature of cloud.
- Distributed nature provides better fault tolerance but requires high degree of replication.

4.2 A Performance Analysis of Load Balancing Algorithms in Cloud Environment. [2]

Load balancing can be achieved by many different kinds of algorithms. Same kind of survey is done by this author. Paper describes different algorithms and compares all with one another using different parameters. Author discussed theory of load balancing that includes issues of load balancing and scheduling and to overcome that issue proposed different types of algorithms either cooperative or non-cooperative mainly divided into two categories i.e Static or Dynamic. Analyzing all the algorithms based on parameters, Honey Bee and Ant Colony found to be better in dynamic nature.

Advantages:

- Dynamic algorithms are always better than static one based on comparison of parameters shown.
- Parameters that are listed to compare load balancing algorithms: Nature, Response Time, Resource Utilization, Fault Tolerant, Waiting Time, Reliability, Throughput, Turnaround Time, Adaptability, Stability, Process Migration, Predictability, Cooperative and Overhead Rejection.

Disadvantages:

- Author suggested ant colony optimization algorithm to modify on capacity of memory and overhead in communication parameters in heterogenous cloud.

4.3 Methodical Analysis of various balancer conditions on public cloud division. [3]

In this research paper, author made a creativity of using static and dynamic algorithms combining together and thought of making load balancing easier on certain assumed conditions i.e Idle and Normal Condition. Load Balancing is performed by balancer on specific node name Central Controller (CC). When job is arrived, balancer checks the condition of node based on parameters like free space, capacity of node and makespan time which is overall task completion time. When node found idle performs Round Robin algorithm which is static and when node found normal performs Ant Colony or Honey Bee behavior whichever is best for load balancing behavior. In overload condition, load balancing not possible by balancer and thus selects other balancer.

Advantages:

- Performed algorithm for data storage in database server.
- Time required by Ant Colony and Honey Bee to assign the node is calculated.
- Time required by Ant Colony algorithm is less than Honey Bee algorithm.
- Suggests that division techniques gives better result for load balancing as compared to balancing without division.

Disadvantages:

- Result is obtained by considering small number of jobs. Large number of jobs must be considered for precise result with varying load and type.

4.4 A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms. [4]

In this survey paper, author discussed theory of load balancing and determines challenges and algorithms of different types to resolve the issue of load balancing by giving better performance. Challenges faced during load balancing according to author are spatial distribution of cloud nodes, storage or replication, algorithm complexity and point of failure.

Advantages:

- Discussed Pros and Cons of Load Balancing Algorithms.
- DDFTP (Dual Direction downloading algorithm from FTP servers) found to be efficient load balancing algorithm.
- Load Balancing Min-Min (LBMM) found to be more reliable task assignments to nodes.
- Ant Colony is fault tolerant and has no implementation complexity.

Disadvantages:

- DDFTP algorithms requires full replication of data files and implementation somewhat complex.

4.5 A New Clustered Load Balancing Approach for Distributed Systems. [5]

Author shows a new way for load balancing approach by estimating four different parameters like CPU Utilization, Memory Utilization, CPU Queue length and Response Time of system to decide workload of each node. Based on cluster load balancing policy (CLB), author calculated workload of each node in format of 3 tuple $<N_{AVL}$, ART, $MQL_I>$. This provides information of each node to cluster manager which then decides to balance the load on different type of networks like Intra-cluster load balancing, Intra-cluster task transfer and Inter-cluster load balancing.

Advantages:

- Reduces communication over global network by using 3 tuple load metric.
- Reduces average response time of jobs over WAN.
- New way of estimating system load which are CPU Utilization, Memory Utilization and CPU Queue length and Response Time.

Disadvantages:

- During simulation, jobs that are selected are with no precedence and no deadline.

4.6 Load Balancing of Virtual Machine Using Honey Bee Galvanizing Algorithm in Cloud. [6]

In this paper, load balancing in cloud is performed using new approach with Honey Bee Galvanizing algorithm. Load Balancing is been compared with behavior of bees to solve the issue of it. Honey bee forage technique is employed for task allocation and cargo balancing. Once task is been allotted to VMs, current load is calculated and if VM becomes overload, task is transferred to neighborhood VM whose load is below threshold which is same as behavior of scout bees that patches flower exploitation regularly and improves algorithm reducing makespan and also reduces response time. Experiments are shown by work done in cloud simulator CloudSim-3.0.1 Tools.

Advantages:

- New concept of algorithm that reduces makespan and response time based on behavior of bees.
- To transfer on other neighborhood VM, we first need to calculate fitness and workload of each node.

Disadvantages:

- QoS factors of tasks are not considered.

4.7 User-Priority Guided Min-Min Scheduling Algorithm For Load Balancing in Cloud Computing. [7]

In this research paper (base paper), author discussed Min-Min scheduling algorithm and improved with two different strategies. One to improve load balanced algorithm in order to reduce makespan and increase resource utilization by introducing new modified algorithm named Load Balancing Improved Min-Min (LBIMM) scheduling algorithm. Second to provide different levels of quality of services i.e. user-priority, author introduced User-Priority Awarded Load Balance

Improved Min-Min (PA-LBIMM) to provide priority for demand of tasks submitted by VIP users to serve them first than to ordinary users.

Advantages:

- Two new scheduling algorithms are been proposed i.e LBIMM and PA-LBIMM to improve load balance of resources and to provide priority to each user so as to execute VIP users task earlier than to ordinary task respectively.
- LBIMM are capable of decreasing completion time of task. Overall 20% improved on average resources utilization ratio better performance than Min-Min algorithm.
- PA-LBIMM out-performs both Min-Min and LBIMM decreasing over 20% of average completion time of VIP tasks.

Disadvantages:

- Did not considered deadline of each task, high heterogeneity of interconnection networks, geography location of tasks and resources and other QOS requirements.
- Tasks are independent.

V. CONCLUSION

By studying many research and survey papers, different types of algorithms for load balancing came to know which are dynamic and decided to merge them to get better efficiency than to perform individually. Load balancing algorithms are good in distributing load equally to all virtual machines and perform task by increasing response time. For enhancing such performance, we came up with novel approach to select algorithms dynamically based on some condition and situation in which they are better, for which we need to calculate some parameters.

REFERENCES

[1] Mayanka Katyal, Atul Mishra "A Comparative Study of Load Balancing Algorithms in Cloud Computing Environment"IJDCC, 2013

[2] A Performance Analysis of Laod Balancing Algorithms in Cloud Environment, 2015 International Conference on Computer Communication and Informatics(ICCCI-2015),Jan,08-10,2015,Coimatore,INDIA

[3] Methodical analysis of various balancer conditions on public cloud division,@2015 IEEE,Anisaara Nadaph, Vikas Maral

[4] Klaithem Al Nuaimi, Nader Mohamed, Mariam Al Nuaimi, Jameela Al-Jaroodi"A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms" IEEE, 2012

[5] Moumita Chatterjee, S.K.Setua "A New Clustered Load Balancing Approach for Distributed Systems" IEEE, 2015.

[6] Monika Rathore, Sarvesh Rai, Navdeep Saluja "Load Balancing of Virtual Machine Using Honey Bee Galvanizing Algorithm in Cloud" IJCSIT, 2015.

[7] Huankai Chen, Professor Frank Wang, Dr Na Helian, Gbola Akanmu "User-Priority Guided Min-Min Scheduling Algorithm For Load Balancing in Cloud Computing" IEEE, 2013

[8] Cloud Computing Features, Issues and Challenges: A Big Picture ,2015 International Conference on Computational Intelligence & Networks. Deepak Puthal DOI:10.1109/CINE.2015.31

[8] Shagufta Khan, Niresh Sharma “Effective Scheduling Algorithm for Load Balancing (SALB) using Ant Colony Optimization in Cloud Computing” in IJARCSSE, 2014.

[9]Ruhi Gupta,"Review on Existing Load Balancing Techniques of Cloud Computing" in International Journal of Advanced Research in Computer Science and Software Engineering,Volume 4, Issue 2, February 2014

[10] Randles, M., 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 Workshops (WAINA), Perth, Australia, April 2010.