

# Load Balancing in Public Cloud Using Partitioning Mechanism

Sudhir A Shinde<sup>1</sup>, Kapil N Vhatkar<sup>2</sup>

<sup>1,2</sup>Assistant Professor, Department of CSE, NK Orchid CET, Maharashtra, India

**Abstract:** *Cloud computing is transforming computing paradigm that involves delivering application and services over the internet. Many of the underlying technology that are the foundation of cloud computing have existed for quite some time, Cloud computing involves provisioning of computing, networking and storage resource on demand providing these resource as metered services to the user in "in a pay you go" model. This paper introduces a better load balancing mechanism in a public cloud based on the cloud partitioning technique with a switch mechanism to choose different strategy for different situation .also there is min-min load balancing algorithm use for improve the efficiency in the public cloud.*

**Keywords:** Public cloud, Load balancing model, cloud partition, load balancing algorithms etc

## 1. Introduction

Cloud computing is model for enabling convenient network access to a shared pool of configurable computing resources (e.g networks services , storage, applications and services ).that can be rapidly provisional and released with minimal management effort or service provider interaction .

### 1.1 Characteristics of cloud Computing

#### On-demand self service

Cloud computing resources can be provisioned on demand by the user, without requiring interaction with cloud service provider.

#### Broad network access

Cloud computing resources can be access over the network using standard access mechanisms that provide platform-independent access through the use of heterogeneous client platform such as workstation laptop, tablets and smart phone.

#### Resource pooling

The computing and storage resources providers are pooled to server multiple users using multi-tenancy. Multi-tenant aspects of the cloud allow multiple users to served by the same physical hardware.

#### Performance

Cloud computing provides improved performance for application sine the resources available to the applications can be scaled up or down based on the dynamic application workloads.

#### Reliability

Application deployed in cloud computing environment generally have a higher reliability since underlying IT infrastructure is professionally managed by the cloud service. Cloud service provider specifies and guarantees the reliability and availability levels for their cloud resources in the form of service level agreements (SLAs).

## 1.2 Deployment Model

### Public Cloud

In the public cloud deployment model, cloud services are available to the general public or a large group of companies. the cloud resources are shared among different users(individuals, large originations, small ad medium enterprise and government ).

### Private Cloud

In the private cloud deployment model ,cloud infrastructure is operated for exclusive use of a single organization .private cloud services dedicated for single organization .cloud infrastructure can be setup on premise or off-premise and may be managed internally or by a third party.

### Hybrid Cloud

The hybrid cloud deployment model combines the services of multiple cloud (private cloud or public cloud ). Hybrid cloud are best suited for organization that want to take advantage of secure application and data hosting on a private cloud and at the same time benefit from cost saving by hosting shared application and date in public cloud .

## 2. Load Balancing in Cloud

Load Balancing is a technique in which distribute the dynamic workload equally across the all the nodes .it used for to achieving the good services and better resources utilizations also improve the overall performance of the system.

## 3. Related Work in Load Balancing

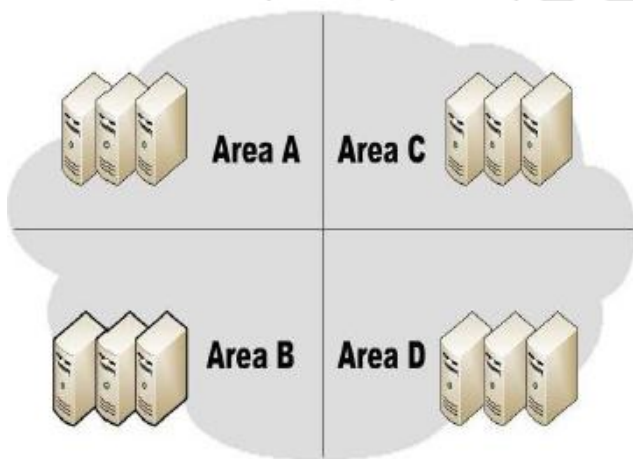
In cloud environment load balancing technique described in a white paper which was written by Adler [7],which describe the different tools and technique which commonly used in load balancing mechanism.

Charko al [8] describe the how to improve the performance and maintain stability in cloud using load balancing

mechanism. There are number of load balancing algorithms used for load balancing in cloud computing environment like Round Robin algorithm , Opportunistic load balancing algorithm , Equally spread current execution algorithm and throttled algorithm[10]. round robin algorithm are static load balancing algorithm in which it select first node randomly . then it allocate the jobs to all another nodes in circular order . so it is not uniform workload distribution so it not so good for cloud environment[13]. Opportunistic load balancing algorithm which not consider the current workload of the system .it keeps all node to be busy .it show load balancing task without good result. In ant colony optimization there is find optimal path from the source of food to colony of ant on the basis of their behavior[12]. In This approach efficient distribution of work load among the node. But each ant build their an own individual result set and it is later on built into a complete solution of system. so ant continuously updates a single result set rather than updating their own result set of the system.

#### 4. System Model

In this system model we are focused on public cloud , public cloud is standard cloud computing model in which number of services provided by service provider .with this [1]. A public cloud it is very large cloud and it include number of nodes and which are present different geographical locations. Cloud partitioning technique is used to manage this large cloud. A cloud partition is a sub part [area] of the public cloud with divided on basis of its geographic locations. The architecture is shown in Fig.1.



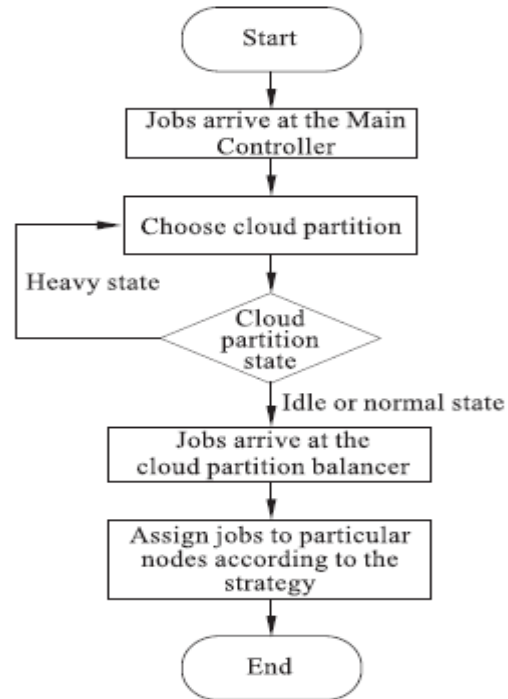
**Fig. 1 Typical cloud partitioning.**

For this system The load balancing strategy is based on the cloud. partitioning mechanism . when creating the cloud partitions, the load balancing will be starts: When there is a any job is arrives on the system, then the main controller decide which the cloud partition should receive and process the job. Then partition load balancer of system decides how to assign the jobs to the nodes. If the load status of a cloud partition is Idle, then this partitioning process the job locally. If the cloud partition load status is not idle or not normal , then this job should be transferred to another partition. The process is shown in below Fig.2.

#### A. Assigning jobs to the nodes in the cloud partition

In this system cloud partition balancer gathers all load information from every node . and this load evaluate the cloud partition status. Status can be divided into three types:

- 1) Load is Idle When Load degree.(N) = 0;
- 2) Load is Normal when  $0 < \text{Load degree (N)} \leq \text{Load degree high}$
- 3) Load is Overload when  $\text{Load degree high} \leq \text{Load degree(N)}$



**Fig. 2 Job assignment strategy.**

#### B. Cloud Partition Load Balancing Strategy

If the cloud partition is idle, then there is many computing resources are available so relatively few jobs are arriving in system. so In this situation, this cloud partition has the capable to process jobs as quickly as possible so a simple load balancing algorithm will be used. There are number of simple load balance algorithm Are used such as the Random algorithm, the Weight Round Robin algorithm , and the Dynamic Round Robin algorithm [12]. For simplicity here we use the Round Robin algorithm.

##### Round Robin algorithm

This very simplest load balancing algorithms[13], in which each new request passes to the next server in the queue. In this algorithm every node has an equal opportunity to be chosen. But in a public cloud, the configuration and the performance of each node of the system will be not the same, so this method may be overload on some nodes.

##### Min-Min Load Balancing Algorithm

In this algorithm the cloud manager identifies the execution time and completion time of the unassigned jobs which are waiting in a queue[14]. then the cloud manager of the system first deals with the jobs having minimum execution time by assigning them to the processors according to the priority of capability to complete the job in specified completion time. If The tasks having maximum execution time then it wait for the unspecific period of time. This algorithm performs better when the

numbers of jobs having small execution time is more than the jobs having large execution time.

### 5. Load Status Table

In this system we build the circular queue and walk through the queue again and again. Jobs will then be assigned to nodes with low degrees of load. Then the node order will be changed when the balancer refreshes the Load status table.

Here two Load Status Tables should be created as: Load Status Table 1 and Load Status Table 2. a flag is also assigned to each table to indicate Read or Write.

When the flag = "Read", it indicates the request is accepted and read it.

When the flag = "Write", it indicates the request are accepted and written.

However, there is read and write table status changed with respective refresh period T.

Read Table			Write Table		
Load Balancer	Node Name	Status	Load Balancer	Node Name	Status
LB0	Node0	Overload	LB0	Node0	Overload
LB0	Node1	Normal	LB0	Node1	Overload
LB1	Node0	Normal	LB1	Node0	Overload
LB1	Node1	Idle	LB1	Node1	Normal
LB2	Node0	Idle	LB2	Node0	Normal
LB2	Node1	Idle	LB2	Node1	Normal
LB3	Node0	Normal	LB3	Node0	Overload
LB3	Node1	Normal	LB3	Node1	Overload

### 6. Simulation Result

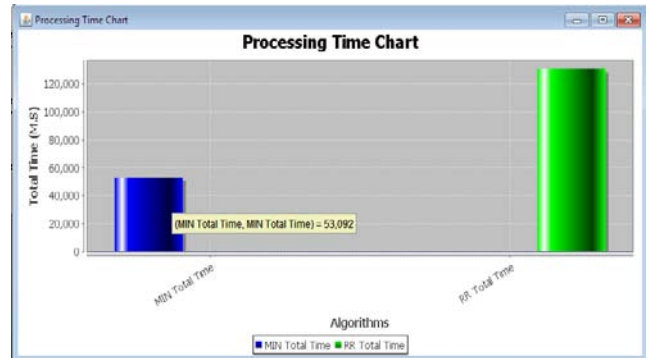
In cloud computing it is required to manage and balance the workload on this system and it decides the quality of the system.

#### Performance Measure

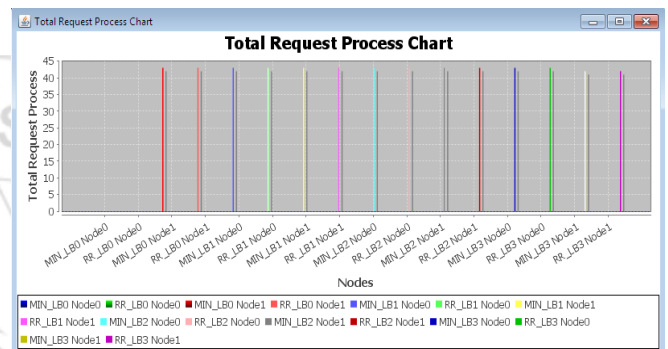
Number of Request	Response Time (in ms)	
	Round Robin Load Balancing Algorithm	Min-Min Load Balancing Algorithm
100	93818	26404
200	111827	53092
300	129775	54330

To attain better performance and good efficiency for balancing the load is very essential. Using Round robin and min-min load balancing algorithms the work load is evenly distributed

among all the nodes of cloud which resulted in optimal resource utilization so total response time also improved.



Thus our system resulted in avoiding the excessive overloading of individual nodes.



### 7. Conclusion

Cloud computing environment provides everything to the user as a service over the internet, load balancing is the major issue of cloud computing because the overloading of the system may lead to poor performance and which can make the dissatisfaction of the customer. For efficient utilization of resources it is required to use efficient load balancing algorithms.

In this paper we have compared the Round Robin load balancing algorithm and min-min load balancing algorithm. Then experimentally it is shown that the min-min load balancing algorithm gives better performance than the Round Robin load balancing algorithm.

### References

- [1] Hunter, The why of cloud [http://www.gartner.com/DisplayDocument?doc\\_cd=226469&ref=g\\_noreg](http://www.gartner.com/DisplayDocument?doc_cd=226469&ref=g_noreg), 2012.
- [2] Gaochao Xu, Junjie Pang, and Xiaodong Fu A Load Balancing Model Based on Cloud Partitioning cloud for the public cloud.
- [3] Microsoft Academic Research, Cloud computing, <http://libra.msra.cn/Keyword/6051/cloudcomputing?query=199%20cloud%20computing>, 2012.
- [4] N. G. Shivaratri, P. Krueger, and M. Singhal, Load distributing for locally distributed systems, Computer, vol. 25, no. 12, pp. 33-44, Dec. 1992.
- [5] Adler, Load balancing in the cloud: Tools, tips and techniques, <http://www.rightscale.com/whitepapers/Load-Balancing-in-the-Cloud.pdf>, 2012.

- [6] K. Nishant, P. Sharma, V. Krishna, C. Gupta, K. P. Singh, N. Nitin, and R. Rastogi, Load balancing of nodes in cloud using ant colony optimization, in Proc. 14th International Conference on Computer Modelling and Simulation (UKSim), Cambridgeshire, United Kingdom.
- [7] Maria Spinola, —An Essential Guide to Possibilities and Risks of Cloud Computing: a Pragmatic Effective and Hype Free Approach for Strategic Enterprise Decision Making!. (white paper) 2009.
- [8] Ratan Mishra and Anant Jaiswal, —Ant Colony Optimization: A solution of Load Balancing in Cloud!, International Journal of Web & Semantic Technology (IJWesT), April 2012
- [9] Pooja Samal, Pranati Mishra, |Analysis of variants in Round Robin Algorithms for load balancing in Cloud Computing! (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 4 (3) , 2013, 416-419.
- [10] D T. Kokilavani, Dr. D. I. George Amalarethinam —Load Balanced Min-Min Algorithm for Static Meta Task Scheduling in Grid computing International Journal of Computer Applications Vol-20 No.2, 2011.

