RESEARCH ARTICLE OPEN ACCESS

An Approach on the Impact of Resource Allocation in Cloud Data Centers to Manage Load Balancing

Tanuja Sharma[1], Swati Soni[2]
[1] M.Tech. Scholar, Department of CSE, Takshshila Institute of Engineering and Technology,Jabalpur (M.P) India
Email id : sharmatanuja042@gmail.com
[2]Assistant Professor, Department of CSE, of Engineering and Technology,Jabalpur (M.P) India
Email id : swatisoni@takshshila.org

Abstract:

For delivering the cloud services over the internet, cloud computing has become proficient infrastructural model for hosting cloud services. Server virtualization is key technology that enables cloud computing as a service, which authorizes dynamic sharing of physical resources. Virtualization introduces the problem of virtual machine placement that increases the overheads in load balancing. Existing infrastructure needs the strategy for the VM Placement as it may create poor allocation and load balancing issues. In most of the cases, due to lack of input parameters physical machines are partially loaded that creates issue of fragmentation which leads to in sufficient resources that causes more utilization of physical machines in any infrastructure. We did extensive survey in the said domain and found that, In load balancing approach, when VM place mentis done without measuring its lifetime, that creates fragments on PM. So we propose dynamic priority based spill over technique and add the concept of short life/long life container for solving the fragmentation issue.

KEYWORDS:
VM Placement, Load Balancing; Fragmentation, Data center, Cloud Computing

1. INTRODUCTION

Cloud computing as a novel and entirely internet-based approach provides a highly available, scalable, and flexible computing platform for a variety of applications and has brought about great benefits to both enterprises and individuals [1]. Computing is being changed to a service based model whereby access to these services depend on users' requirements without regard to where the services are hosted or how they are delivered [2]. Such computing model offers many types of services, such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software asa Service (SaaS). With the spread of cloud computing, cloud work flow systems are designed to facilitate the cloud infrastructure to support huge scale distributed collaborative - business and e-science applications [3]

1.1 CLOUD SERVICE MODELS

1. **Cloud Software as Service (SAAS): -** It is also known as ''On demand Software'' and it is a software licensing and it provide the software to consumeron subscription base.
Applications: Business / Multimedia, Web Service
Examples: You tube, Google Apps

2. **Cloud Platform as Service (PAAS):-**In this type of service, the consumer can deploy, the user generated or developed applications which is create by using programming or tools given by provider, on the cloud infrastructure.
Applications: Software Framework (Java/.Net), Data /File Storage
Examples: AWS, Microsoft Azure

3. **Cloud Infrastructure as Service (IAAS): -** This is a capability provided to the consumer by which, it can provision processing, storage, networks and other fundamental computing resources where the consumers can deploy and run the software.
Applications: Hardware Resources (CPU, Memory, Disk)

Examples: Go Grid, Amazon EC2, Data Centers. b. Cloud Deployment Models

- Public Cloud:-This public cloud is available for every organization.
- Private Cloud:-This cloud is available only for particular organization or company.
- Community Cloud: - In this type of cloud deployment model, the infrastructure of the cloud system is commonly used by many of the organizations and supports a specific community with shared concerns.
- Hybrid Cloud:-It is a composition of two or more different clouds that is private or community or public. Element of the hybrid cloud are tightly coupled. Load Balancing algorithms can be of 3 categories are as

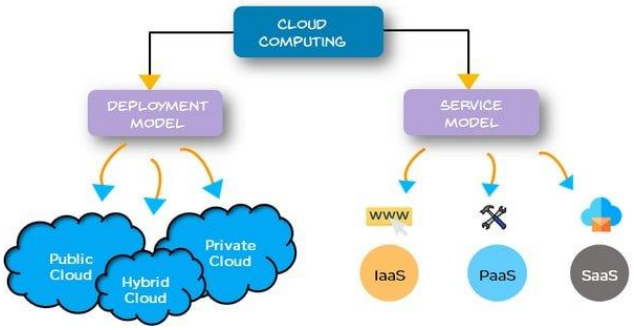

Fig. 1 - Cloud Computing Model

i. **Sender Initiated**: If the load balancing algorithm is initiated by the sender.
ii.**Receiver Initiated**: If the load balancing algorithm is initiated by the receiver.
iii. **Symmetric**: It is the combination of both sender initiated and receiver initiated. Normally, the loads were distributed evenly, uniformly, overloaded, minimally among the nodes to the system.[5][6][7][8].

1.2. Load Balancing Algorithms (Cloud Computing)

Min-Min Load Balancing Algorithm This is simple static algorithm and offers excellent performance in task scheduling. The cloud service manager find the completion time of every task. The new task has been waiting in a queue for execution.

This algorithm assigns the task to the resource based on which task has minimum execution time to complete. The pseudo code is following

```
Procedure Minmin(Task Ti)
{
Find execution_Completion_Time of each task
Store the execution_Completion_Time of task Ti
inorderQueue
repeat
{
for each task Ti in orderQueue
{
obtain minimum completionTime from orderQueue;
assign task to vm;
update the execution_Completion_Time;
}
}
Until order queue empty;
}
```

This algorithm works well when the task has minimum execution time however if task has maximum execution time then the task must be wait with undefined time. This will lead the starvation problem. This algorithm is best in the situations where the number of tasks with minimum completion time.

1.3. Max-Min Load Balancing Algorithm

This algorithm is following identical procedure of Min- Min algorithm. This algorithm calculates the execution completion time of all tasks. The maximum completion time is taken and assigned to the corresponding resources.

This algorithm is best in the situations where the amount of tasks with maximum completion time and it take away the starvation. The task minimum completion time has been waiting in ordered queue until the other maximum completion time task must be completed. Here we can understand that this algorithm performs well in a static environment and both the algorithm has their merits and demerits based on the environments.

The performance doesn't depend on the algorithm chosen but indeed the environment taken. Min-Min and Max-Min algorithms are equally performed on the static cloud environment.

2. MEDIUM LEVEL LOAD BALANCING MECHANISM

The new innovative Load Balancing algorithm is to balance the load in medium level. The Server is having 100 rps. The Client A can accept only 50 rps. After reaching the half of the requests from Server automatically redirect the requests to the Client B, if it reaches half load then redirect to Client C and so on. The Medium Level Load Balancing algorithm will give to increase client satisfaction and maximize resource utilization

2.1 OBJECTIVE

Objective of this work is to introduce and evaluate the proposed scheduling and load balancing algorithm by considering the capabilities of each virtual machine (VM), the task length of each requested job, and the interdependency of multiple tasks. Performance of the proposed algorithm is studied by comparing with the existing methods

2.2 METHEDOLOGY:

To carry out this experiment, cloud analyzer simulator will be used. This simulator is build on CloudSim. Results obtained from CloudSim environment are very close to real environment. It is not only used by researcher but also by corporate to perform simulation and to obtain the result for analysis. This Experiment has been carried out considering the following parameters.

3. PROPOSED WORK

We aim to propose a system which handles the problem of resource utilization and solve fragment issue. We use short life/long life container for solving the fragment issue.

Most of the time due to lack of input parameters [short life VMs, long life VMs placement always done on Physical machine which is partially loaded. Which results into so many partially loaded Physical machines and creates issue of fragmentation which leads to insufficient resources that causes more utilization of physical machines in any infrastructure? For solving this problem, we use short life and long life container in our architecture. We define threshold point for calculation of completion time of VM. If the life time of incoming job is smaller than the threshold value then it is defined as a short life job so that jobs store in short life container otherwise it is long life job and store in long life container. Using our architecture resource utilization is maximize and fragment issue seems to be solve.

The architecture of the system is simple, flexible and easy. All the incoming jobs are queued in VM. In VM queue, it stores all the short time and long time jobs. Also, every one of the jobs from VM queue are transferred to DC manager. There is a threshold value which is ascertained in view of VM MIPS in DC manager .In the event that the life time of incoming job is littler than threshold value then it is characterized as a short life VM else it is long life VM. Monitor stores all the data about CPU, stockpiling, and RAM and gives all data to the DC administrator. DC manager take plan of arrangement as indicated by the data of monitor and VM queue.

3.1 System Architecture

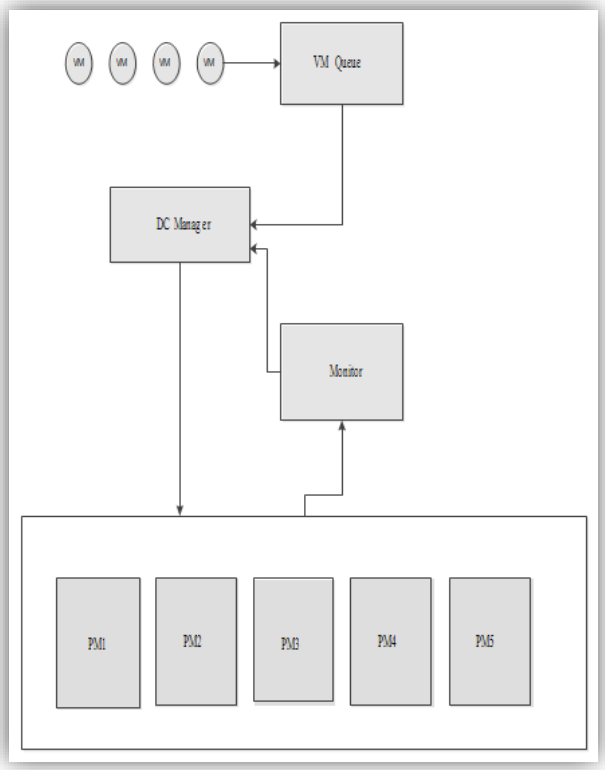

Fig. 2 - Proposed System Architecture

VM Queue : VM queue store all incoming VMs. All short time VMs and long time VMs store in VM queue.

DC manager : DC manager analyze next few jobs(VMs) in VM queue. Using DC manager we characterize incoming job is short time job or long time job using threshold point. DC manager predefine threshold value. If the life time of incoming job is smaller than the threshold value then it is characterized as a short life job else it is long life job. Monitor works with DC manager and provides all information like status of CPU, RAM, storage etc to the DC manager of each individual physical machine. DC manager collect the information about the both VM and PM.

CloudSim is It is basically a Library for Simulation of Cloud Computing Scenarios. It has some features such as it support for modeling and simulation of large scale Cloud Computing infrastructure, including data centers on a single physical computing node. It provides basic classes for describing data centers, virtual machines applications, users, computational resources, and Policies.

Let's indicate the terminology of the emulator (Fig-3)

Region: in Cloud Analyst, the world is divided into 6 regions that coincide with the 6 major continents in the world;

User Base: User Base is considered as a single unit, and is used to generate traffic;

Data Processing Center: brokerage services determine which center should accept and process the request that comes from each user database;

Vm Load Balancer: it is responsible for distributing the load to the available data center. VmLoad Balancer distributes the load in the data center based on the load balancing policy.

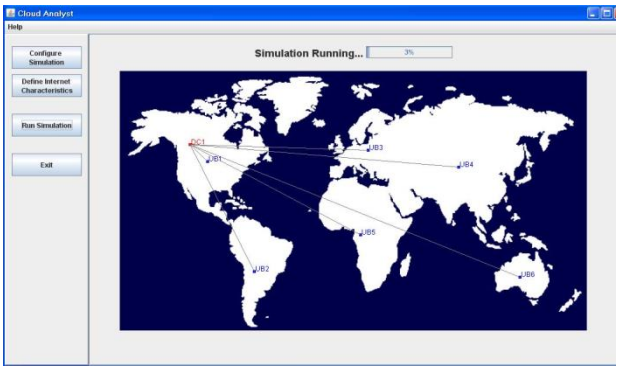

Fig. 3 - Cloud Analyst Simulator

4.2 DATA CENTER NETWORK DESIGN

The Design and simulation for performance analysis will be done by using the OPNET simulation software, as shown in Figure 4 below.

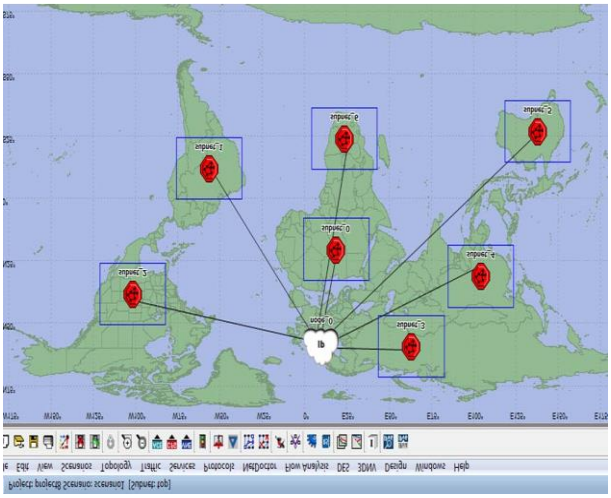

Fig. 4 - Performance analysis using OPNET Simulator

4.3 OPNET

Fig. 4 also illustrate the design of data-centers subnets, it show that each subnet include storage devices, server, routers and work stations. There are five subnets each of them are about data-center and they connected together by wire line. Each region contains one data-center by using OPNET Data Center Configuration Parameter: Total 05 data center will be considered for the simulation environment. Architecture of each data centre is given in Table-1

Table 1 - Initial conditions

VM	Core units	Hours
VM	2	3
VM	1	1
VM	6	6
VM	3	2
VM	3	5
VM	2	2

Here we define VM, VM core units and time. Total capacity of physical machine is 9 units. Show the matrix, it is initial stage,

In next stage and several stages we have final stage, PM1 is full so newly coming VM place in PM2. Show the final condition of example in Table-2

Table 2 - Final Stage of VM Core Units an time

	PM1(Total unit=9)	PM2(Total unit=9)
Short life	VM1 is finish	VM4 (2hours) VM6 (2hours)
Long life	VM3(4hours) VM5(5hours)	
Remaining unit	0	4

4.4 SYSTEM PERFORMANCE

In first graph, we define the placement process of VM. In existing work all short life and long life VM placement is done on same physical machine and in our proposed work, we arrange all short life VMs on same container and long life VM on same container. So our resource utilization is maximizing, show in graph.

In second graph, we define utilization of resource or PM. According to the Optimization function if the division of VM (mips) and PM (mips) is near to the 1 than fragmentation is less, so utilization of PM is more. And the value is the nearest to the 0 than the fragment is more, so utilization is less. Base on that optimized value of 1 and 0 we create graph.

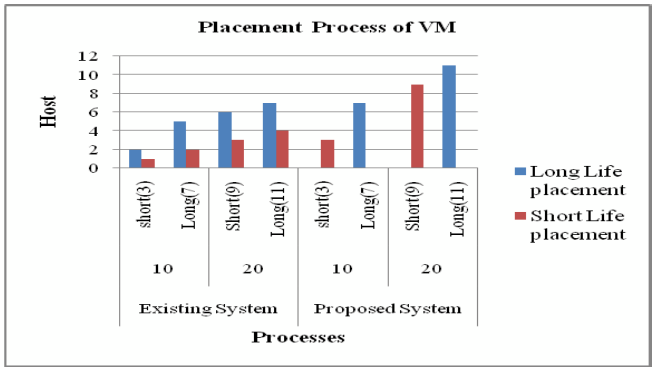

Fig. 5 - Placement Process of VM

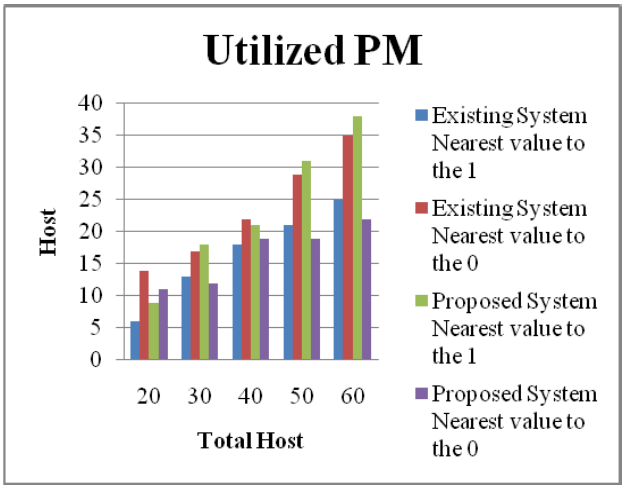

Fig. 6 - Utilization of PM

5. CONCLUSION AND FUTURE WORK

After doing rigorous survey on various issues in resource utilization, we found that load balancing, VM placement and fragmentation are the greatest issue in cloud computing. So we propose dynamic priority based spill over technique and add the concept of short life/long life container for solving the fragmentation issue. Our architecture maximizes the resource utilization also we minimize fragmentation issue with using short life/long life container at physical machine in our architecture.

REFERENCES

[1] Dynamic Recourse Allocation Based on Priority in Various Data Centers Custom-Based waiting queue technique in Cloud Computing Sebastian Perez-Salazar, Ishai Menache , Mohit Singh , Alejandro Toriello 2021

[2] Dynamic Resource Allocation Method Based on Symbiotic Organism Search Algorithm in Cloud Computing July-Sept. 2022, pp. 1714-1725, vol. 10 10.1109/TCC.2020.3002205 Ali Belgacem, École Militaire Polytechnique (EMP), Alger, Algeria

[3] Aarti Singha, Dimple Junejab, Manisha Malhotraa, "Autonomous Agent Based Load Balancing Algorithm in Cloud Computing"- International Conference on Advanced Computing Technologies and Applications (ICACTA- 2015)

[4] B. Subramani, ‘‘A New Approach For Load Balancing In Cloud Computing’’, IEEE ,vol.2, pp.1636-16405, May 2013.

[5] Buyya, Rajkumar. "Market-oriented cloud computing: Vision, hype, and reality of delivering computing as the 5th utility". In Proceedings of the 2009 9th IEEE/ACM.

[6] Geethu Gopinath P, Shriram K Vasudevan, "An in- depth analysis and study of Load balancing techniques in the cloud computing environment"-.2nd International Symposium on Big Data and Cloud Computing (ISBCC’15).

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

[8] Jinhua Hu," A Scheduling Strategy on Load Balancing of Virtual Machine Resources in Cloud Computing Environment", 3rd International Symposium on Parallel Architectures, Algorithms and Programming 978-0-7695- 4312-3/10 © 2010IEEE (published)

[9] Jayant Baliga, Robert W. A. Ayre, Kerry Hinton, and Rodney S.1Tucker, Fellow IEEE, “Green Cloud Computing: Balancing Energy in Processing, Storage, and Transport”.

[10] S. Kapoor and C. Dabas, Cluster based load balancing in cloud computing, Proc. Eighth International Conference in Contemporary Computing (IC3), 2015, 76-81.

[11] Klaithem Al Nuaimi," A Survey of Load Balancing in Cloud Computing: Challenges and Algorithms", 2012 IEEE Second Symposium on Network Cloud Computing and Applications 978-0-7695-4943-9/12

[12] Mohit and Jitender Kumar, International Journal of Advanced Research in Computer Science and Software Engineering , "A Survey of Existing Load Balancing Algorithms in Cloud Computing".

[13] S. Mohinder, R, Ramesh, D. Powar, ‘‘Analysis of Load Balancers in Cloud Computing’’, International Academy of Science, Engineering & Technology, vol.2, May 2013.

[14] Poulami Dalapati1, G. Sahoo, "Green Solution for Cloud Computing with Load Balancing and Power Consumption Management"- International Journal of Emerging Technology and Advanced Engineering,Vol3:2013

[15] Saurabh Kumar Garg and Rajkumar Buyya, “ Green Cloud computing and Environmental Sustainability” .(references)

[16] Sidhu A, S. Kinger, “Analysis of Load Balancing techniques in Cloud Computing”, Council for innovative research international

Journal of Computer & Technology, vol.4, March-April 2013.

[17] Sumalatha M.R, C. Selvakumar, T. Priya, R.T. Azariah, and P. M. Manohar, "CLBC-Cost effective load balanced resource allocation for partitioned cloud system", Proc. International Conference on Recent Trends in Information Technology (ICRTIT),2014, 1-5.

[18] The Amazon Elastic Compute Cloud (Amazon EC2), http://aws.amazon.com/ec2/

[19] Yilin Lu," A Hybrid Dynamic Load Balancing Approach for Cloud Storage", 2012 International Conference on Industrial Control and Electronics Engineering 978-0-7695-4792-3/12 © 2012 IEEE.

[20] Yuvapriya Ponnusamy, S Sasikumar, "Application of Green Cloud Computing for Efficient Resource Energy Management in Data Centres", International Journal of Computer Science and Information Technologies, Vol 3:2012.