A Novel Performance Improvement Model for Cloud Computing

[1]Umesh Lilhore, [2]Dr Santosh Kumar

[1]PhD Scholar, [2]PhD Supervisor,
Department of Computer Science & Engineering
MUIT Lucknow

***ABSTRACT*: Cloud computing is a new innovative technology that provides sharing of computing resources over the network. Cloud users and cloud service providers have a legal agreement for services and users pay as per use. Cloud computing provides resource and cost optimization that attract IT companies to use cloud more. Due to rapid increases in cloud services and cloud users it's more important for cloud service provider to serve better services in fewer casts. Users are demanding for better performance from cloud service providers. Performance of cloud computing depends on various factors such as load balancing, efficient work flow and load distribution among computing devices. In this paper we are presenting a new performance improvement model is introduced "Advance Anticipatory Performance Improvement Model for Cloud Computing (AAP-IMC)", for cloud computing. The proposed "AAP-IMC" performance improvement model presents a set of solutions for load balancing in the Cloud. Proposed model uses combine strategy of another two methods "MFL-APSOM" and "ADRS-DDMC".**

***Keywords*- Cloud Computing, Performance of Cloud computing, Load balancing MFL-APSOM, ADRS-DDMC and AAP-IMC**

1. Introduction

Cloud computing as its name suggests, it is a new innovative computing style where same time various cloud uses the computing resources on the Internet. In the recent era of technology and innovation, this new technology has they want to use[1,4]. Typically Cloud vendors,
chose to implement an efficient automatic load balancing mechanisms into their service delivery system. This approach allows the numbers of resources to be adjusted with changing the demand levels. In well managed Cloud infrastructure, load balancing functions is not an option it's a must. The goal of the load balancers are o promote the availability and service provisioning of the Cloud resources, and the secondary goal is to improve the overall performance.

Benefit	Opportunity
Scalability	Add or remove resources on demands
Elasticity	Allocate and release resources upon usage
Mobility	Accessing the Cloud network is not dependent on time and location
Low infrastructure cost	Helps small businesses to grow sooner
Increased data storage	Access to large storage capacity for data storage and backup plan
Disaster recovery	Using the virtual backups recovery will be 4 times faster than using another system than cloud
Availability	Rapid deployment of the infrastructure to access the resources

2. Related work in performance improvement of Cloud computing-

Cloud Computing offers plenty of benefits to the users. Infinite availability of the resources, beneficial payment emerged as a very strong option for not only large, scale level organizations but also very much useful for small level organizations, those only access the computing resources, whatever

model, and scalability and elasticity specifications of the Cloud networks motivated businesses to change their legacy system with the novel Cloud infrastructure [1, 5]. Scalability and elasticity can be highlighted as the main benefits of the Cloud Computing. Scalability enables users to add or remove resources from their networks at any time [2, 9]. Conversely, elasticity emphasizes on dynamic resource allocations. In fact, elasticity explains the definition of the "pay-as-you-go" model which is allocating and releasing the resources according to the user's requirements argues that, although elasticity introduces many benefits to the users, effective load management is essential to guarantee the availability of the resources [3,5].

Table 2 Cloud benefits & Opportunity

In cloud computing following parameters directly affect cloud performance.

A. Load Balancer and Load Balancing
B. Work flows between nodes
C. Resource consumption such as memory, storage and processors.

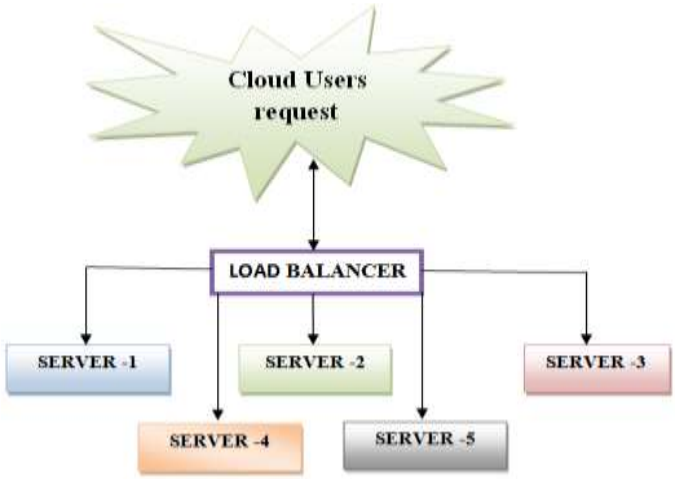

Figure 2- Server utilization model of cloud computing

In Figure 2 presents a hybrid Server utilization model of cloud computing. The load balancer narrows down the server options based on the server address proximity to the location of the requestor. In this case, server 1, 2 and 3 are the closest. Because server 2 is fully utilized, the load balancer passes the request to server 1 or 3 which have not fully utilized their resources [7, 8]. In this scenario applying the closer IP address management will increase the availability and performance of the system.

3. Proposed "AAP-IMC" Model

Cloud load balancing server allocated the load at the time of increasing the several CPUs or memories for their resources to scale up with the increased demands. This service is primarily applied for commercial enterprise demands. In cloud the resource optimization is the most important process. Nowadays, many organizations use cloud storage. Many requests can occur at a time in a cloud which gives more problems in the cloud. The important issues are server crash, failover and outage. Some outages are lengthy and it may take more hours or days to solve the problem.

Cloud computing is a circle of several virtual machines that are sliced into virtual servers and placed at different geographical locations for providing services to customers. The "AAP-IMC" proposed method deals with the following research challenges in the cloud-

a) Efficient load balancing.
b) Optimization of resources.
c) Identification of reliable VMs.
d) Better performance of complete system.

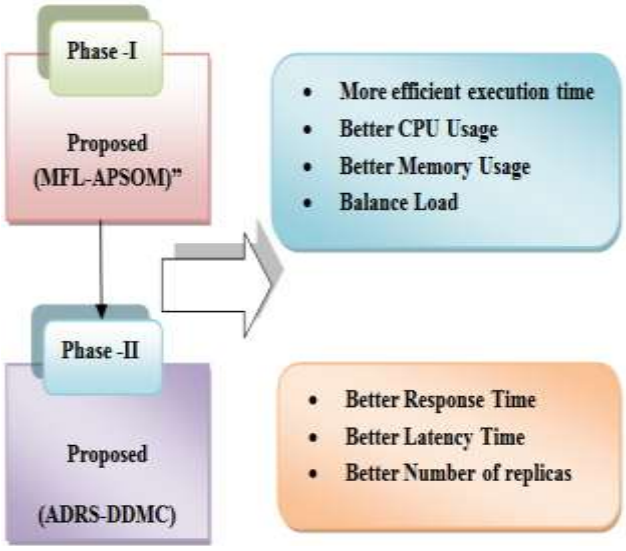

Figure 3- Phases in Proposed Model

The proposed performance improvement model "Advance Anticipatory Performance Improvement Model for Cloud Computing (AAP-IMC)", for cloud computing. The proposed model presents a set of solutions for load balancing in the Cloud. Proposed model uses combine strategy of two proposed methods "MFL-APSOM" and "ADRS-DDMC", in various phases.

3.1 PHASE 1 MFL-APSOM –It is based on "Modified fuzzy Logic and Advance Particle Swarm Optimization Model (MFL-APSOM)", to optimize the total execution time of tasks in the workflow applications. It is based on heuristic algorithm that uses "Modified fuzzy Logic and Advance Particle Swarm Optimization Model (MFL-APSOM)".

The key objective of applying the MFL-APSOM method is to minimize the total tasks execution time by verifying the load fluctuations of the interconnected tasks. The model optimizes the load balancing method by minimizing the total execution time. In the Cloud architecture, in order to manage the allocated tasks, there is an application scheduler designed to balance the workload between available resources.

3.2 Phase 2 "ADRS-DDMC"- It is based on "Anticipatory Data Replication Strategy with Dynamic Distributed Model for Cloud Computing (ADRS-DDMC)". This phase introduces a novel dynamic data replication method that is functioning based on anticipations to create the pre-replicas for future needs of the sites. This method uses anticipatory data replication for increasing the reliability and availability of cloud data. Pre-replication can increase the data availability and robustness of the Cloud systems and hence requested jobs can be completed with minimum execution time and high network usage output. In the second Phase, the reliability and failure VM is identified and in the last phase efficient resource optimization is accomplished.

3.3. Final phase cloud performance improving "AAP-IMC" model-

Final phase of this research work investigates "Advance Anticipatory Performance Improvement Model for Cloud Computing (AAP-IMC)", for cloud computing. The

proposed "AAP-IMC" performance improvement model presents a set of solutions for load balancing in the Cloud. Proposed AAP-IMC model uses combine strategy of two proposed methods "MFL-APSOM" and "ADRS-DDMC", in various phases. Managing accessibility for application performance is a key challenge for QoS.

PARAMETERS IN PROPOSED AAP-IMC	
Phase I (MFL-APSOM)	Phase- II (ADRS-DDMC)
Response Time	Mean Job Execution Time
Execution Time	Effective Network Usage
CPU Load	Total Number of Replicas
CPU Time	Soft error rates %
Memory Rate	Resource Availability
Work flow soft error rates	Time Consumption

Table 3.2 Parameters in proposed AAP-IMC

4. Expected outcomes of proposed "AAP-IMC" Model-

Some of the specific advantages of the proposed methodology over the existing load balancing and reliability algorithm are given below-

1) Increase the VM speed
2) Reduce processing time
3) Reduce context switching
4) Reduce waiting time of the request in the queue
5) Increase reliability by identifying failure VMs earlier
6) Efficient resource allocation in VMs

5. Conclusions and Future Work

This research work presents a performance improvement model "AAP-IMC" for cloud computing. Performance of cloud computing can be improve by efficient load balancing, and task scheduling. For validation of our proposed model performance model in future work we will implement the proposed model and existing models and compare various performance parameters.

REFERENCES

[1] Abbadi, I.M. And Anbang Ruan, "Towards Trustworthy Resource Scheduling in Clouds", IEEE-Information Forensics and Security, IEEE Transactions-Pages: 973-984-June 2014.

[2] Alicherry.M. and Lakshman, T.V, "Optimizing data access latencies in cloud systems by intelligent virtual machine placement" , IEEE Communications-Pages: 647- 655-April 2013.

[3] Brototi Mondala, Kousik Dasgupta and Paramartha Duttab,"Load Balancing in Cloud Computing using Stochastic Hill Climbing-A Soft Computing Approach" Science Direct, Pages: 783 -789, May 2012.

[4] Chao-Tung Yang, Wei- Li Chou, Ching-Hsien Hsu and Cuzzocrea, A., "On Improvement of Cloud Virtual Machine Availability with Virtualization Fault Tolerance Mechanism", IEEE Cloud Computing, Pages: 122-129, Dec. 1st 2011.

[5] Dong Oh Son, Hong Jun Choi, Jae Hyung Park and Cheol Hong Kim ," Analysis of Memory Management Policies for Heterogeneous Cloud Computing",IEEE Transactions, Pages:1- 3-June 2013.

[6] Egwutuoha, I.P. ,Shiping Chen , Levy, D.; Selic, B.and Calvo, R.," A Proactive Fault Tolerance Approach to High Performance Computing (HPC) in the Cloud, IEEE Transactions, Pages 268-273, Nov. 2014.

[7] Gupta, P., Goyal, M.K. And Kumar, P. "Trust and reliability based load balancing algorithm for cloud IaaS" IEEE Transactions- Pages 65-69, Feb. 2013.

[8] Habib, S.M. Ries, S. and Muhlhauser M.,"Towards a Trust Management System for Cloud Computing" IEEE-Trust, Security and Privacy in Computing and communications (TrustCom), Pages 033-939, Nov. 2011.

[9] Rohit O.Gupta and Tushar Champaneria, "A Survey of Proposed Job Scheduling Algorithms in Cloud Computing Environment", International Journal of Advanced Research in Computer Science and Software Engineering, PP 782-790, year 2014.

[10] S.Sindhu and Saswati Mukheree, "Efficient Task Scheduling Algorithms for Cloud Computing Environment", ACM Transactions, Pages 79-83, July 2011.

[11] Xu, Gaochao, Pang, Junjie and Fu, Xiaodong," A load balancing model based on cloud partitioning for the public cloud", IEEE Transactions, Pages-34-39, Feb 2013.

[12] Yaojun Han and Xue mei Luo, "Hierarchical Scheduling Mechanisms for Multilingual Information Resources in Cloud Computing", Springer, Pages 268-273, year 2013.

[13] Zhongyuan Lee and Ying Wang, " A Dynamic Priority Scheduling algorithm on service request scheduling in Cloud Computing", published in IEEE Electronic and Mechanical Engineering and Information Technology (EMEIT- 2011), Pages 4665-4669, Aug. 2011.

[14] Zulkar Nine, M.S.Q. Azad, M.A.K., Abdullah, S.and Rahman, R.M. ," Fuzzy logic based dynamic load balancing in virtualized data centers", IEEE, pp 117-121, July 2013

[15] Umesh lilhore and Santosh Kumar, "Analysis of performance factors for cloud computing", International Journal of Information Technology and Management (IJIMTM ignited journal), Vol IX, Issue No. XIV, November 2015, ISSN 2249-4510, PP 305-310.

[16] Umesh lilhore and Santosh Kumar, "Survey on Load balancing in Cloud Computing", International Journal of Information Technology and Management (IJIMTM ignited journal), Vol X, Issue No. XV, May 2016, , ISSN 2249-4510, PP 224-230.