iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** VIII **Month of publication:** August 2017

DOI: http://doi.org/10.22214/ijraset.2017.8304

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue VIII, August 2017- Available at www.ijraset.com

An Efficient Resource Management Method for Cloud

Ashwini. M[1], Prof. Yogish H. K [2]
[1]*P. G Student, Department of Computer Science and Engineering, SCE, Bangalore, Karnataka*
[2]*H. O. D, Department of Computer Science and Engineering, SCE, Bangalore, Karnataka*

Abstract*: Cloud computing provides on demand access to heterogeneous resources (compute, memory, network, storage etc.). It allows customers to scale up and down their resource usage based on the needs. The resources should be efficiently and reliably managed by the service provider to deliver such services. The resource management in cloud is a challenging task as it needs to satisfy both customers requirement and server performance. In this paper, a resource management method is proposed for efficient resource utilization. The proposed method handles the request in real time and guarantees high resource utilization. The resource management method checks for the jobs running on the virtual machine. If the jobs are idle they are killed and the resources utilized by those are given to other jobs. The scheduling is done based on the resource requirement. The CPU resource parameter is considered in the work.

***Keywords*: Cloud computing, Resource Management, Resource allocation, Virtualization, Scheduling**

I. INTRODUCTION

The National Institute of Standards and Technology (NIST) defines 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". There are numerous advantages of cloud computing, the most basic ones being lower costs, re-provisioning of resources and remote accessibility. Cloud computing lowers cost by avoiding the capital expenditure by the company in renting the physical infrastructure from a third party provider. Due to the flexible nature of cloud computing, more resources can be accessed from cloud providers when there is need to expand the business. The remote accessibility enables users to access the cloud services from anywhere at any time. To gain the maximum degree of the above mentioned benefits, the resources should be managed optimally to the applications running in the cloud. The key challenges in resource management are resource provisioning and scheduling. The first task is to decide whether to accept the jobs and allocate the resources to it or not. It is critical decision, especially when the amount of work exceeds the amount of available resources. The capacity planning and decisions about job admissions are complicated by uncertainties due to the unpredictable arrival and termination of the jobs, fluctuations in demand, and the risk of contentions on different resource levels. Inefficient job admission can lead to underutilization of resources, revenue loss and degradation in the performance. It is therefore important to plan the admission process of the jobs to achieve efficient utilization without sacrificing performance. The partitioning and assigning resources highly impact utilization efficiency. The scheduling of jobs is another important task in resource management. The resources are provisioned based on the requirements for the job execution in cloud environment only if the required resources are available in resource pool. If the required resources are not available then the jobs are processed in the queue. The jobs in the queue are scheduled for execution once the resources are available. The importance of the job, preferences of the job and time required to execute the jobs are considered for efficient resource utilization while scheduling.

II. LITERATURE SURVEY

Abirami S.P[1] In this work a scheduling algorithm named Linear Scheduling of Tasks and Resources (LSTR) is designed to schedule tasks and resources. The scheduling algorithm in LSTR mainly focuses on the distribution of the resources among the users which is able to maximize the chosen Quality of Service (QoS) parameters. The scheduling algorithm designed considering the jobs and the total available virtual machines together. It is named as LSTR scheduling strategy. The scheduling algorithm is carried out based on the prediction that the initial response to the request is made only after assembling the resource for a finite amount of time (1 day or 1hr) but not allocating the resources as they arrive. The resource requests are collected and sorted in different queues based on the threshold value (say 50 for the sample memory request). The algorithm uses the memory request in gigabytes as the input such as Ri= (R1, R2 … Rn). The shortest request in both A[RQi] and B[RQi] is processed first which results in the allocating of resource to more number of requests. K.C Gouda[2] In this work a priority algorithm is used for a better allocation of jobs in the cloud environment. The model considers parameters such as user, time, number of processor request, resource assigned, resource selection

©IJRASET (UGC Approved Journal): All Rights are Reserved

criteria etc. Each jobs submitted by the customer consists of different tasks. For each task different parameters are considered such as time, processor request, importance and price. Based on all the parameters, priority algorithm decides priority among the different task submitted by different customers. The customer's task with higher priority will be given first chance to run. The customer's task with next higher priority will be given next chance and so on. Gunho Lee[3] In the work topological aware resource allocation architecture (TARA) optimizes resource allocation in cloud system. TARA uses a prediction engine and a light weight simulator to estimate the performance of the given resource allocation and a genetic algorithm to find optimized solution in the large search space. TARA reduces the job completion time of the applications by up to 59% when compared to application independent allocation policies. Rashmi K.S [4] In the work users submit their applications to the Cloud service provider. The cloud manager in the data centre distributes the execution load among all the virtual machines by monitoring the virtual machine. Cloud manager maintains a record containing the VM ID; job ID of the jobs that has to be allocated to the corresponding virtual machine. The virtual machine status represents percentage the resource utilization. The cloud manager allocates the resources and distributes the load per the data structure. The cloud manager checks the virtual machine status routinely to distribute the load as per the record. If any virtual machine overloaded then the jobs are migrated to the virtual machine which is underutilized by tracking the record. If there is more than one available virtual machine then the assignment will be made based on the least hop time. After the completion of execution, the cloud manager automatically updates the record. This approach gives less response time in which job rejections will be reduced.

Koti Reddy.S [5] a framework designed for cloud environments called Nephele. It is the first data processing framework to include the possibility of dynamically allocating or de-allocating compute resources from a cloud in its scheduling and during job execution. Nephele follows master-worker pattern. A user must start a virtual machine in cloud which runs the Job Manager. It receives the client's jobs. It is also responsible for scheduling them and coordinates the job execution. Radhika T.V[6] In the work a profit model algorithm is proposed to minimize the infrastructure cost and Service Level Agreement violations. The resource algorithm is used to maintain the resource available in the service provider's centre. The priority algorithm is used for resource allocation jobs. A profit model is used to calculate the maximum profit of cloud administrator by giving service to each user request.

III. METHODOLOGY

In the work a resource management method is used to increase the resource utilization of the virtual machines. The methodology used here is as shown in Fig.1.

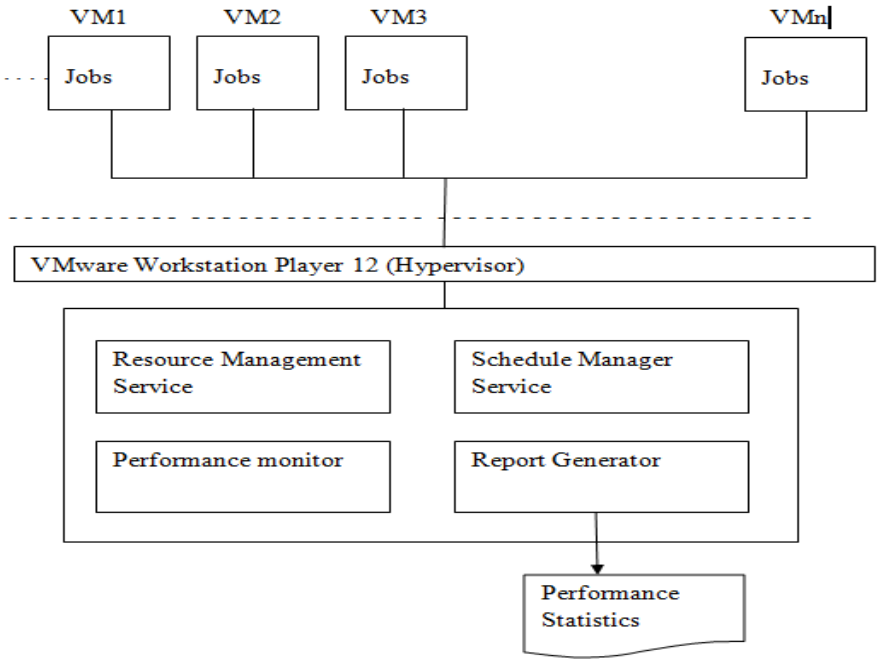


Fig.1 System architecture

A. Resource Management Service

In this module, a resource management service is created. It checks the jobs running on the virtual machine. If there are jobs or tasks idle then those jobs are de-provisioned. These resources are given to other jobs which are waiting for the resources.

©IJRASET (UGC Approved Journal): All Rights are Reserved

B. Schedule Manager Service

In this module, a schedule manager service is created to schedule the jobs. The jobs are executed if the resources are available. The threshold set for CPU is 100%. If the resources to execute the job are not available then it is sent to queue. Once the resources are available the jobs are executed.

C. *Performance monitor*

In this module, a performance monitor is created to display the CPU utilization of the virtual machine. It provides a visual display of the system performance. It is used to query the current CPU load on the virtual machine.

D. Report Generator

In this module, a report is generated which shows the overall system performance. It shows the maximum utilization of the resources by applying the resource management method. The graphical representation of the CPU utilization is generated.

IV. EXPERIMENTAL RESULTS

First the virtual machine is checked for the resource available. If the resources available the jobs are loaded on the virtual machine. Then the performance monitor is checked to see the resource being utilized. If CPU utilization reaches 100% the resource management service searches for the jobs that are idle. The idle jobs are de-provisioned and these resources are given to the jobs which are waiting for the resource. Then the idle job is sent to queue. When the jobs become active the resources to it are provisioned again. It reduces the wastage of resources. The resource utilization of the virtual machine will be increased by allowing more number of active jobs to execute. Thus improves the overall resource utilization of the system by executing more number of jobs and also satisfying more number of customers request with the available resources.

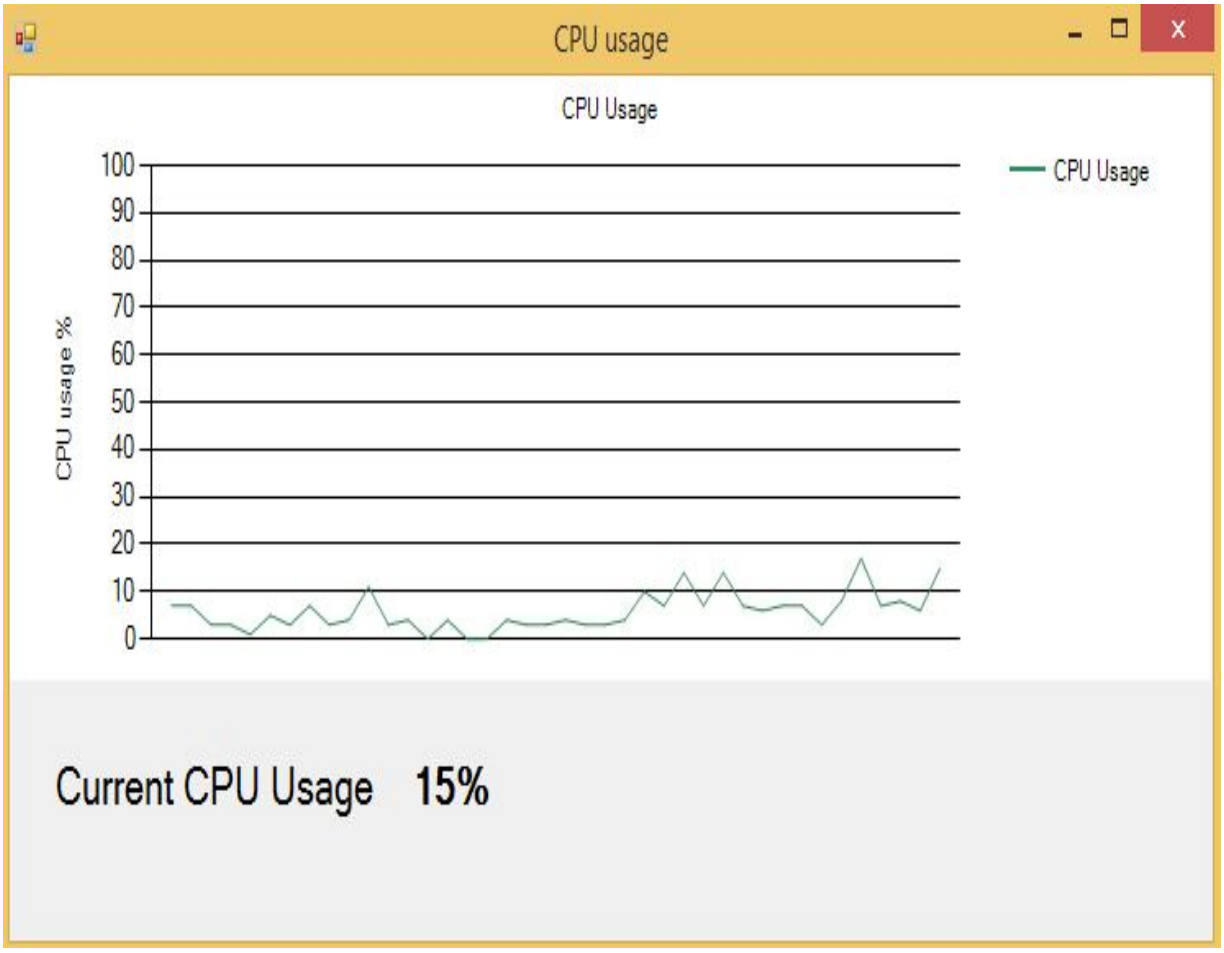


Fig.2 CPU Utilization performance monitor

©IJRASET (UGC Approved Journal): All Rights are Reserved

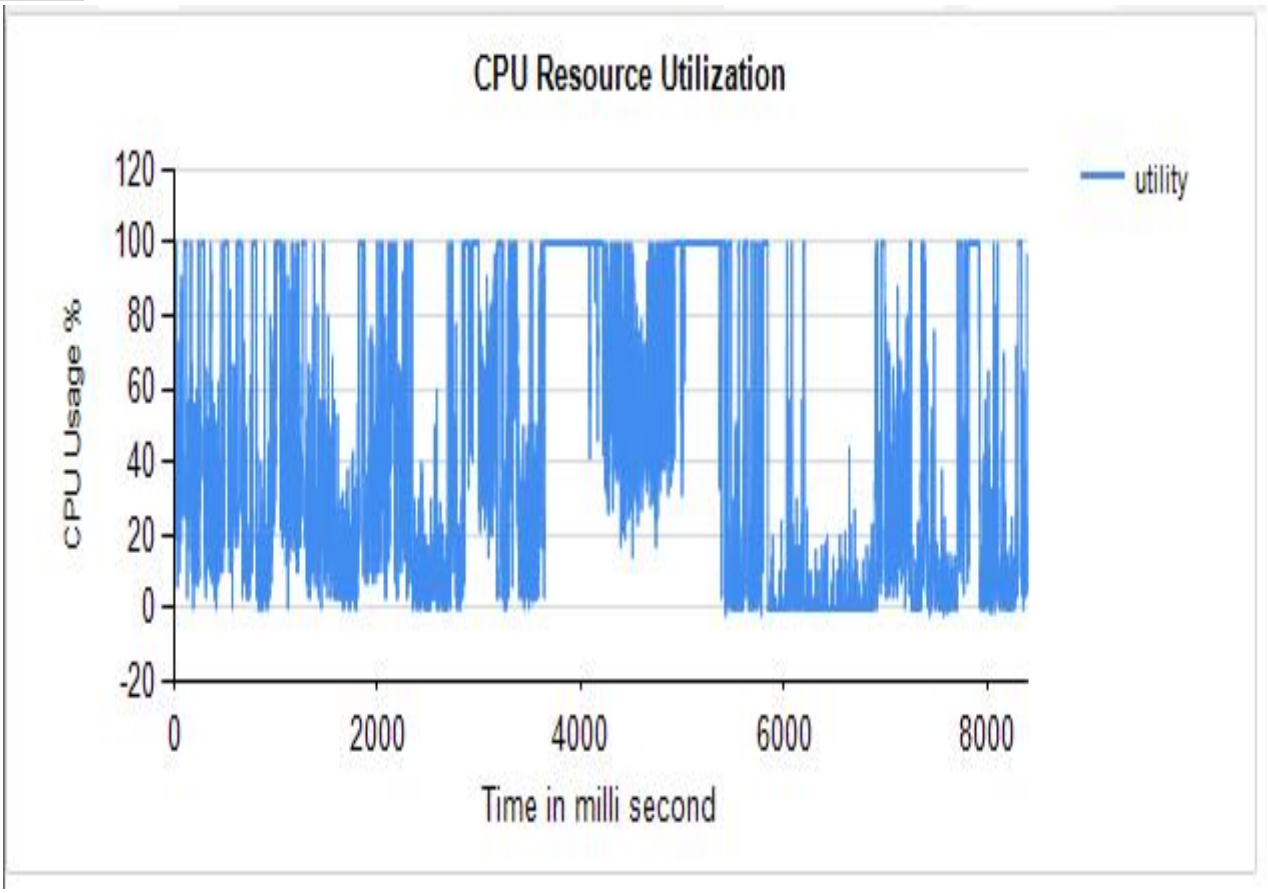


Fig.3 The overall CPU utilization of the system

V. CONCLUSION

In this study a resource management method was created to improve the CPU resource utilization of the virtual machines. The load is applied on the virtual machine by running number of jobs. When the resource reaches the threshold then the idle jobs are de-provisioned to run the jobs waiting for the resources. The system performance is increased by the resource management method. The number of jobs executed increased comparing to traditional scheduling.

REFERENCES

[1] Abirami S. P and ShaliniRamanathan, "Linear scheduling strategy for resource allocation in cloud environment", International journal on Cloud Computing: Services and Architecture (IJCCSA), Vol.2, No.1, 2012, 9-17.

[2] K C Gouda, Radhika T V and Akshatha M, "Priority based resource allocation model for cloud computing", International journal of Science, Engineering and Technology Research (IJSETR), Vol.2. Issue 1, 2013, 215-219.

[3] Gunho Lee, NirajTolia,ParthasarathyRanganathan, and Randy H.Katz, "Topology aware resorce allocation for data-intensive workloads",ACM SIGCOMM Computer Communication Review, 2011, 41(1):120-124.

[4] Rashmi K S, Suma V and Vaidehi M, " Enhancing Load Balancing Approach to Avoid Deadlocks in cloud", International Journal of Computer Applications on Advanced Computing and Communication Technologies for HPC Apllications- ACCTHPCA, June 2012.

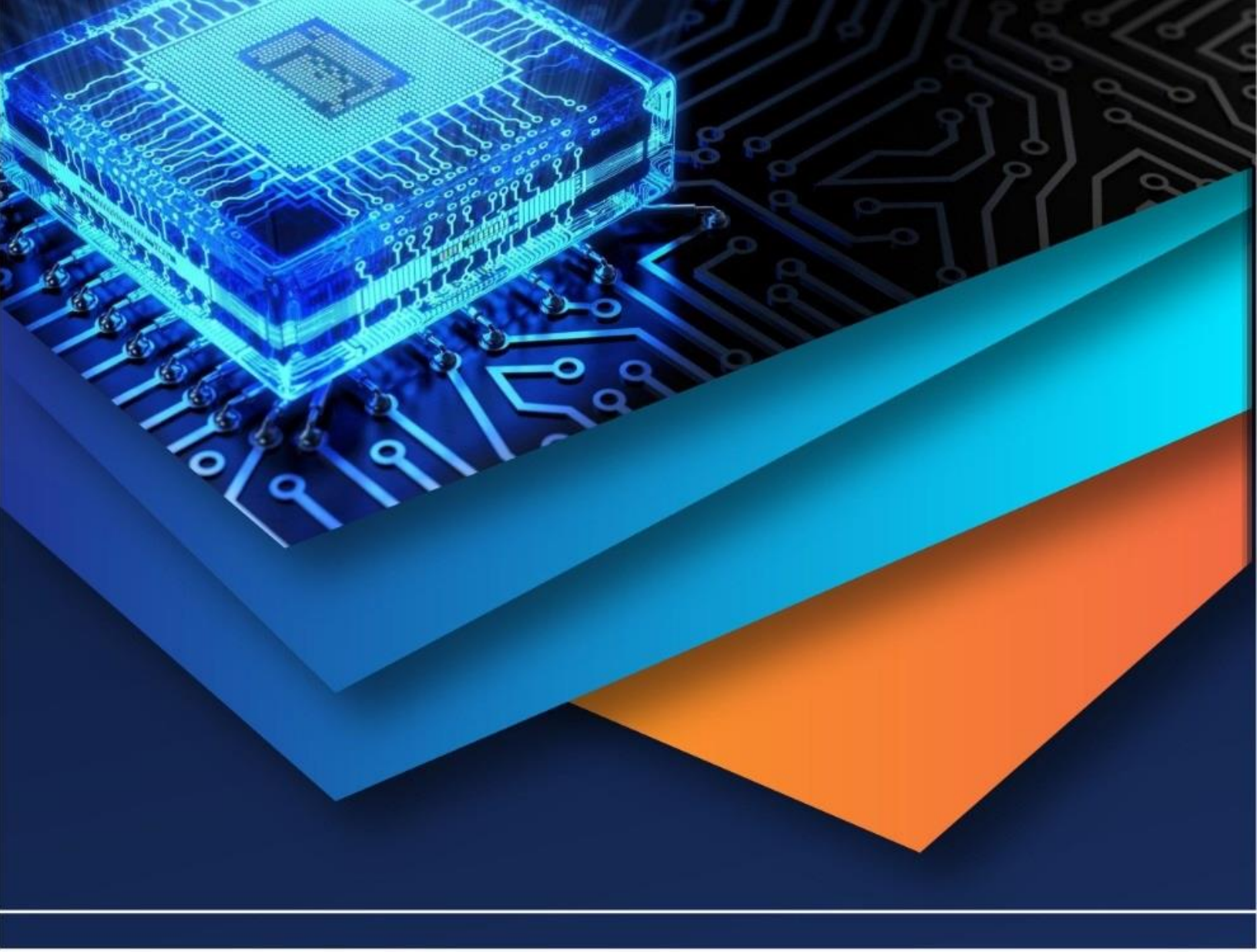
[5] Koti Reddy S, Ch. Subba Rao, "Dynamic Resource Allocation in The Cloud Computing Using Nephele's Architecture", International Journal of Engineering Science And Advanced Technology (IJESAT), Vol.2, Issue 4, Aug 2012, 1144-1151.

[6] Radhika T V, K C Gouda, "Resource Allocation Model for Efficient Management in Cloud Computing", International Journal on Advanced Computer Theory and Engineering (IJACTE), Vol. 2, Issue 4, 2013, 2319-2526.

[7] VMware website. Available: https://www.vmware.com

[8] CPU usage monitoring. Available: http://www.Mentalis.com

©IJRASET (UGC Approved Journal): All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)