International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 3 | ISSN : 2456-3307

Resource Provisioning and Resource Allocation in Cloud Computing Environment

Ritu Aggarwal
M. Tech Computer Science & Engineering, MMEC Mullana Himalayan Grroup of Professional Institutions, Himachal Pradesh, India

ABSTRACT

Today, Cloud computing become an emerging technology which will has a significant impact on IT Infrastructure. Still, Cloud computing is infancy. In the current cloud computing environment there is numerous of application, consist of millions of module, these application serve from large quantity of users and the user request becomes dynamic. So there must be provision that all resources are dynamically made available to satisfy the needs of requesting users. The resource provisioning was done by considering Service Level Agreements (SLA) and with the help of parallel processing using different types of scheduling heuristic. In this paper we realize such various policies for resource provisioning and issues related to them in current cloud computing environment. Keywords: Cloud computing; Scheduling, Service Level Agreements (SLA), Virtualization, Virtual Machines (VM). - Distributed computing is a developing innovation which gives compelling administrations to the customers. It licenses customers to scale here and there their assets utilization relying on their necessities Because of this, under arrangement and over arrangement issues may happen. To defeat this relocation of administration use Our Paper concentrates on conquering this issue by appropriating the asset to various customers through virtualization innovation to upgrade their profits. By utilizing virtualization, it allots datacenter assets powerfully in view of uses requests and this innovation likewise bolsters green innovation by advancing the number of servers being used. We show another approach called "Skewness", to figure the unevenness in the Multi-level asset usage of a server. By enhancing Skewness, we can join diverse sorts of workloads enough and we can enhance the entire utilization of server assets

Keywords: Cloud computing, overprovision, underprovision, green computing, skewness,Cloud Computing; Cloud Services; Resource Allocation; Infrastructure.

I. INTRODUCTION

A large portion of the associations indicate enthusiasm on cloud, in light of the fact that with minimal effort we can get to assets from cloud in an adaptable and secure way. Cloud shares their asset to various clients. Cost of assets changes essentially contingent upon design for utilizing them. Thus effective administration of assets is of prime enthusiasm to both Cloud Suppliers and Cloud Users. The accomplishment of any cloud administration programming fundamentally relies on the adaptability, scale furthermore, proficiency with which it can use the hidden equipment assets while giving vital execution disconnection[1]. Effective asset administration answer for cloud conditions needs to give a rich arrangement of asset controls for better detachment. Here element asset allotment and load adjusting is the testing undertaking to give viable administration to customers. Because of pinnacle requests for an asset in the server, asset is over used by customers through virtualization. This may corrupt the execution of the server. In under use of asset is exceptionally poor when contrast with over usage, for mapping this we relocate customers handling from VM to other VM. Virtual machine screens (VMMs) give a system to mapping virtual Machines (VM) to physical assets in Physical Machine (PM). Yet, is avoided the cloud[2]. Cloud supplier ought to guarantee that physical

1.1 Advance weights: PM ought to give all the fundamental assets required to process applications on VMs. It fulfills VM needs in light of its ability.

1.2 Green Computing: Advance superfluous utilization of PMs to spare the vitality The work talked about underneath in our Paper makes examinations of how to conquer these two issues in cloud. To start with we need to share the work to servers balancingly contingent on their ability. By sharing server we can play out their undertaking viably to enhance stack on it. Next, we need to upgrade the utilization of asset then no one but we can give adaptable and powerful administration to customers, for this utilization of asset Monitor is essential.[2] By observing, we came to know underutilization and overutilization of assets in PM through VMs. gives the flow for green architecture. So to ascertain the use of asset we present another approach called "Skewness". checks the unevenness of application using skewness algorithms. With the assistance of already utilized asset logs, we need to estimate intermittently for future asset needs. A customer can interest for exceptionally asset arrangement. At the time there might be a chance for inadequate asset, while giving that support of the planned customer, asset and also memory determining is vital. For this we plan "asset guaging calculation"

II. CLOUD COMPUTING

It is one of the most promising technologies in the modern world having a broad array of web-based services aimed at allowing users to obtain a wide range of functional capabilities on a 'pay-as-you-use' basis. Cloud computing is still in its early stages, but the public sector it offers various benefits such as, Cost savings, Highly automated, Flexibility, More Mobility, Increase storage, Business agility, availability of resources etc[6].

2.1Cloud Deployment Models :The entire concept of cloud computing is divided into three forms of cloud. All three have significant characteristics; however their choice depends on the personal requirements of business environment. These include Private Cloud, Public Cloud and Hybrid Cloud which is exceptionally flexible.

• Private Cloud: A Private cloud not promotes shared environment. This means private cloud is beneficial for those organizations that do not want to share their confidential data with any third party. • Public Cloud: In this type of cloud form, data stored is in cloud server, which is located at a distant place elsewhere. It enables users to share and access data from anywhere and at any point of time. This means public cloud promotes shared environment. Although, a bit risky in terms of data security as business operations are done through Internet, but offers highly scalable environment.

• Hybrid Cloud: A Hybrid cloud is a combine of both and gives users or business entities advantage of both the cloud environments. Suppose, a business enterprise wants to share its services and products with its clients across the globe, but at the same time wants to hide the confidential information from them, Hybrid cloud architecture would suit best for such types of businesses.

2.3 Cloud Service Models :Cloud services are classified into three models, it includes:

• Software-as-a-Service (SaaS) - In this model, providers offers a complete application to the client's for use on demand. EMC Mozy is an example of SaaS.

• Platform-as-a-Service (PaaS) – This model has capability provided to the client's is to deploy applications, supported by the providers. It is also used as an application development environment. Google App Engine and Microsoft windows Azure Platform are examples of PaaS.

• Infrastructure-as-a-Service (IaaS) - This model has capability to provides scalable computing, storage, network and other fundamental computing resources where the client's can able to deploy and run their own software. Amazon Elastic Compute Cloud (Amazon EC2) is an example of IaaS Cloud Service Model Cloud computing has enabled IT organizations and individual to gain benefits, such as automated and rapid resource provisioning, flexibility, high availability and faster time to market at a reduced total cost of ownership. Although there are concerns and challenges, the benefits of cloud computing are compelling enough to adopt it.Challenges of Cloud Following are the major challenges needs to overcome, it includes:

2.3.1 Dynamic Scalability: The most of the application is ready to scale up and scale down the compute nodes dynamically as per the response time of the user's queries. The resource allocation and task scheduling delay is the one of the major factor which leads to the need of dynamic and effective load management system [15].

2.3.2 Interoperability and Portability- When user wants to move their services from one cloud to another is a lack of compatible, time-consuming and labour-intensive because currently there is no standard way exists for interfacing with a different cloud and each provider exposes its own APIs [16].

2.3.3 Cloud Security and Privacy - In cloud computing, Security is one of the most critical issues because people are not worried about the location of data they store in the cloud. Some CSPs may have less transparency than others about their policy of information security. Reliable Service Allocation - Several reliability problems that arise when allocating applications to processing resources in a Cloud computing platform [17].

2.3.4 Compliance Audit Requirements – Now a day's cloud computing services can challenge various compliance audit requirements currently in place. Data locations, cloud computing security policy transparency are all challenging issues in compliance auditing efforts [15].

2.3.5 Automated service provisioning - Resource provisioning decisions must made online for satisfying service level objectives while minimizing operational costs. Scalable Querying and Secure Access: In both grid and cloud environment are two open problems such as scalable provenance querying and secure access of provenance information.

2.3.6 Multi-tenancy: Many CSPs have multi-tenant applications in which multiple independent application are serviced using a single set of resources. When a number of applications executing by the single compute node, then the amount of bandwidth allocated to each application reduces which may lead to performance degradation. Reliability and fault-tolerance: fault tolerance is required which help to improve the reliability of develop system. Cloud computing leads an opportunity to offering testing tools for testing the application against compute failures in clouds [18].

2.3.7 Power: An energy aware resource management offers many type of services to meet the needs of users because it require enormous amount of power for that in cloud computing environment [15].

2.3.8 Data location - cloud computing technology allows cloud servers to reside anywhere, thus the enterprise may not know the physical location of the server used to store and process their data and applications. Although from the technology point of view, location is least relevant, this has become a critical issue for data governance requirements. It is essential to understand that many Cloud Service

Providers (CSPs) can also specifically define where data is to be located.

2.3.9 Disaster recovery - It may be possible that the data may be commingled and scattered around multiple servers and geographical areas for a specific point of time cannot be identified. In the cloud computing model, the primary CSP may outsource capabilities to third parties, who may also outsource the recovery process. This will become more complex when the primary CSP does not ultimately hold the data. One of the important requirements for a Cloud computing environment is providing reliable QoS. It can be defined in terms of Service Level Agreements (SLA) that describes such characteristics as minimal throughput, maximal response time delivered by the deployed system. Although modern virtualization technologies can ensure performance isolation between VMs sharing the same physical computing node, due to aggressive consolidation and variability of the workload some VMs may not get the required amount of resource when requested. This leads to performance loss in terms of increased waiting time, time outs or failures in the worst case. Therefore, Cloud providers have to deal with resources provisioning for user request, while meeting QoS requirements.

III. RELATED WORK

Very little literature is available on this survey paper in cloud computing paradigm.

Shikharesh et al. in paper [30] describes the resource allocation challenges in clouds from the fundamental point of resource management. The paper has not addressed any specific resource allocation strategy.

Patricia et al. [25], investigates the uncertainties that increase difficulty in scheduling and matchmaking by considering some examples of recent research. It is evident that the paper which analyzes various resource allocation strategies is not available so far. The proposed literature focuses on resource allocation strategies and its impacts on cloud users and cloud providers. It is believed that this survey would greatly benefit the cloud users and researcher. Quiroz et al. [1] introduced a Decentralized; robust Online Clustering mechanism to drive workload (i.e. VM) provisioning on enterprise grids and clouds. In order to deal with inaccurate resource request that leads to overprovisioning provided by application job requests, their mechanism has demonstrated a model-based approach for estimating the application service time given its provisioning. They also presented a quadratic response surface model (QRSM), which was found to best capture the behaviour of the workload for application-specific. It is used to model the application in the cloud computing environment dynamically.

In [3] Van et al. proposed an autonomic resource manager to address the problem of autonomic virtual resource management for hosting service platforms while optimizing a global utility function that integrates the operating cost of the platform provider and the application level SLAs. They also used utility function with constraint programming approach to achieved self optimization by defining business level SLAs of the application and the resource exploitation cost of the hosting provider as constraints. The proposed system attempts to maximize the performance of the hosted applications with an optimal operating cost for the hosting provider.

3.1 Load dispatching : In this paper, we tend to depict unmistakable properties, execution, and power models of alliance servers, maintained a honest to goodness data take after assembled from the sent Windows Live voyager. Mishandle the models, we tend to style server provisioning and cargo dispatching figures and study delicate joint efforts between them. We have a tendency to exhibit that our counts will save a significant measure of essentialness while not surrendering customer experiences.

3.2. Green computing : Late work has seen that desktop PCs in certain conditions gobble up stores of vitality in blend while 'before staying lethargic rich of the time. The question is an approach to manage additional centrality by holding these machines rest however keeping up an imperative partition from client disturbance. Lite Green occupations virtualization to choose this downside, by moving idle desktops to a server wherever they will stay "continually on" while not obtaining the significance estimation of a desktop machine. The consistency offered by Lite Green licenses Joined States to unequivocally manhandle short sit periods other than as long broadens.

3.3 Load Adjusting in Data Centers: In this paper, we tend to given our style of relating Nursing deft information center with fused server relationship in Nursing stockpiling virtualization together with the execution of an end-to-end organization layer. We have a tendency to propose the best way to deal with utilize this for non-troublesome sensible load leveling inside the information center intersection different resource layers – servers, stockpiling and framework switches. To the present finish, we have a tendency to developed an absolutely fascinating Vector Dot subject to deal with the quality displayed by the information center topology and similarly the 3D nature of the masses on resources.

3.4 RESOURCE ALLOCATION STRATEGIES (RAS) AT A GLANCE

The input parameters to RAS and the way of resource allocation vary based on the services, infrastructure and the nature of applications which demand resources. The schematic diagram in Figure 1.

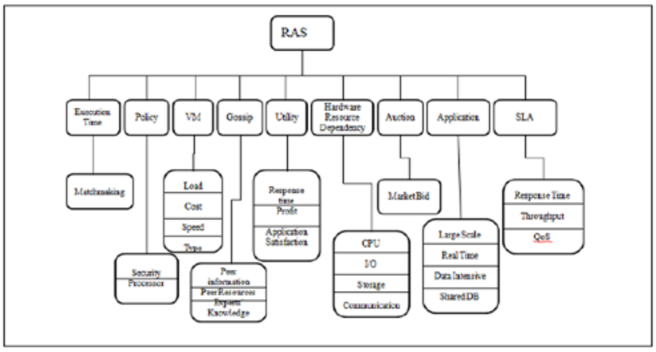

Figure 1

depicts the classification of Resource Allocation Strategies (RAS) proposed in cloud paradigm. The following section discusses the RAS employed in cloud. A. Execution Time Different kinds of resource allocation mechanisms are proposed in cloud. In the work by Jiani at.al [15], actual task execution time and preemptable scheduling is considered for resource allocation. It overcomes the problem of resource contention and increases resource utilization by using different modes of renting computing capacities. But estimating the execution time for a job is a hard task for a user and errors are made very often .But the VM model considered in [15] is heterogeneous and proposed for IaaS. Using the above-mentioned strategy, a resource allocation strategy for distributed environment is proposed by.

Jose et al. [16]. Proposed matchmaking (assign a resource to a job) strategy in [16] is based on Any-Schedulability criteria for assigning jobs to opaque resources in heterogeneous environment. This work does not use detailed knowledge of the scheduling policies used at resources and subjected to AR's (Advance Reservation). B. Policy Since centralized user and resource management lacks in scalable management of users, resources and organizationlevel security policy [6],

Dongwan et al. [6] has proposed a decentralized user and virtualized resource management for IaaS by adding a new layer called domain in between the user and the virtualized resources. Based on role based access control (RBAC), virtualized resources are allocated to users through domain layer. One of the resource allocation challenges of resource fragmentation in multi-cluster environment is controlled by the work given by Kuo-Chan et al. [20], which used the mostfit processor policy for resource allocation. The most-fit policy allocates a job to the cluster, which produces a leftover processor distribution, leading to the most number of immediate subsequent job allocations. It requires a complex searching process, involving simulated allocation activities, to determine the target cluster. The clusters are assumed to be homogeneous and geographically distributed. The number of processors in each cluster is binary compatible. Job migration is required when load sharing activities occur.

Experimental results shows that the most-fit policy has higher time complexities but the time overheads are negligible compared to the system long time operation. This policy is practical to use in a real system. Resource Allocation Strategies in Cloud Computing C. Virtual Machine (VM) A system which can automatically scale its infrastructure resources is designed in [24]. The system composed of a virtual network of virtual machines capable of live migration across multi- domain physical infrastructure. By using dynamic availability of infrastructure resources and dynamic application demand, a virtual computation environment is able to automatically relocate itself across the infrastructure and scale its resources. But the above work considers only the nonpreemptable scheduling policy. Several researchers have developed efficient resource allocations for real time tasks on multiprocessor system. But the studies, scheduled tasks on fixed number of processors. Hence it is lacks in scalability feature of cloud computing [18]. Recent studies on allocating cloud VMs for real time tasks [36], [31], [17] focus on different aspects like infrastructures to enable real-time tasks on VMs and selection of VMs for power management in the data center. But the work by Karthik et al. [18], have allocated the resources based on the speed and cost of different VMs in IaaS. It differs from other related works, by allowing the user to select VMs and reduces cost for the user Users can set up and boot the required resources and they have to pay only for the required resources [3]. It is implemented by enabling the users to dynamically add and/or delete one or more instances of the resources on the basis of VM load and the conditions specified by the user. The above mentioned RAS on IaaS differs from RAS on SaaS in cloud because SaaS delivers only the application to the cloud user over the internet.

Zhen Kong et al. have discussed mechanism design to allocate virtualized resources among selfish VMs in a noncooperative cloud environment in [44]. By non-cooperative means, VMs care essentially about their own benefits without any consideration for others. They have utilized stochastic approximation approach to model and analyze QoS performance under various virtual resource allocations. The proposed stochastic resource allocation and management approaches enforced the VMs to report their types truthfully and the virtual resources can be allocated efficiently. The proposed method is very complex and it is not implemented in a practical virtualization cloud system with real workload.

D. Gossip Cloud environment differs in terms of clusters, servers, nodes, their locality reference and capacity. The problem of resource management for a large-scale cloud environment (ranging to above 100,000 servers) is addressed in [28] and general Gossip protocol is proposed for fair allocation of CPU resources to clients. A gossip-based protocol for resource allocation in large scale cloud environments is proposed in [9]. It performs a key function within distributed middleware architecture for large clouds. In the thesis, the system is modeled as a dynamic set of nodes that represents the machines of cloud environment. Each node has a specific CPU capacity and memory capacity. The protocol implements a distributed scheme that allocates cloud resources to a set of applications that have time dependent memory demands and it dynamically maximizes a global cloud utility function. The simulation results show that the protocol produces optimal allocation when memory demand is smaller than the available memory in the cloud and the quality of the allocation does not change with the number of applications and the number of machines. But this work requires additional functionalities to make resource allocation scheme is robust to machine failure which spans several clusters and datacenters. But in the work by Paul et al. [26] cloud resources are being allocated by obtaining resources from remote nodes when there is a change in user demand and has addressed three different policies to avoid over-provisioning and underprovisioning of resources. Recent research on sky computing focuses on bridging multiple cloud providers using the resources as a single entity which would allow elastic site for leveraging resources from multiple cloud providers [19]. Related work is proposed in [24] but it is considered only for preemptable tasks.

Yang et al. [43] have proposed a profile based approach for scaling the applications automatically by capturing the experts' knowledge of scaling application servers as a profile. This approach greatly improves the system performance and resource utilization. Utility based RAS is also proposed for PaaS in [12]. Gossip based co-operative VM management with VM allocation and cost management is introduced. By this method, the organizations can cooperate to share the available resources to reduce the cost. Here the cloud environments of public and private clouds are considered. They have formulated an optimization model to obtain the optimal virtual machine allocation. Network game approach is adopted for the cooperative formation of organizations so that none of the organizations wants to deviate. This system does not consider the dynamic co-operative formation of organizations. The implication for a desktop cloud is that individual resource reallocation decisions using desktop consolidation and decision based on aggregate behavior of the system. E. Utility Function There are many proposals that dynamically manage VMs in IaaS by optimizing some objective function such as minimizing cost function, cost performance function and meeting QoS objectives. The objective function is defined as Utility property which is selected based on measures of response time, number of QoS, targets met and profit etc. There are few works [4], [38] that dynamically allocate CPU resources to meet QoS objectives by first allocating requests to high priority applications. The authors of the papers do not try to maximize the objectives. Hence the authors' Dorian et al. proposed Utility (profit) based resource allocation for VMs which use live VM migration (one physical machine to other) as a resource allocation mechanism [7]. This controls the cost-performance trade-off by changing VM utilities or node costs. This work mainly focuses on scaling CPU resources in IaaS. A few works [1}that use live migration as a resource provisioning mechanism but all of them use policy based heuristic algorithm to live migrate VM which is difficult in the presence of conflicting goals. For multitier cloud computing systems (heterogeneous servers), resource allocation based on response time as a measure of utility function is proposed by considering CPU, memory and communication resources in [10].

HadiGoudarzi et al. characterized the servers based on their capacity of processing powers, memory usage and communication bandwidth. For each tier, requests of the application are distributed among some of the available servers. Each available server is assigned to exactly one of these applications tiers i.e. server can only serve the requests on that specified server. Each client request is dispatched to the server using queuing theory and this system meets the requirement of SLA such as response time and utility function based on its response time. It follows the heuristics called force-directed resource management for resource consolidation. But this system is acceptable only as long as the client behaviors remain stationary

IV. PROPOSED SYSTEM

Here we have two fundamental objectives to give dynamic asset allotment :

4.1. Advance weights: PM ought to give all the fundamental assets required to process applications on VMs. It fulfills VM needs in light of its ability. 2. Green Computing: Advance superfluous utilization of PMs to spare the vitality The work talked about underneath in our Paper makes examinations of how to conquer these two issues in cloud. To start with we need to share the work to servers balancingly contingent on their ability. By sharing server we can play out their undertaking viably to enhance stack on it. Next, we need to upgrade the utilization of asset then no one but we can give adaptable and powerful administration to customers, for this utilization of asset Monitor is essential. By observing, we came to know underutilization and overutilization of assets in PM through VMs.Figure 1 gives the flow for green architecture. So to ascertain the use of asset we present another approach called "Skewness". checks the unevenness of application using skewness algorithms. With the assistance of already utilized asset logs, we need to estimate intermittently for future asset needs. A customer can interest for exceptionally asset arrangement. At the time there might be a chance for inadequate asset, while giving that support of the planned customer, asset and also memory determining is vital. For this we plan "asset guaging calculation".

4..1.1 Advantages of proposed system:

- ✓ We build up an asset portion framework that can maintain a strategic distance from over-burden in the framework viably while limiting the quantity of servers utilized.
- ✓ We present the idea of "skewness" to gauge the uneven usage of a server. By limiting skewness, we can enhance the general usage of servers despite multidimensional asset imperatives.

4.2 System Design

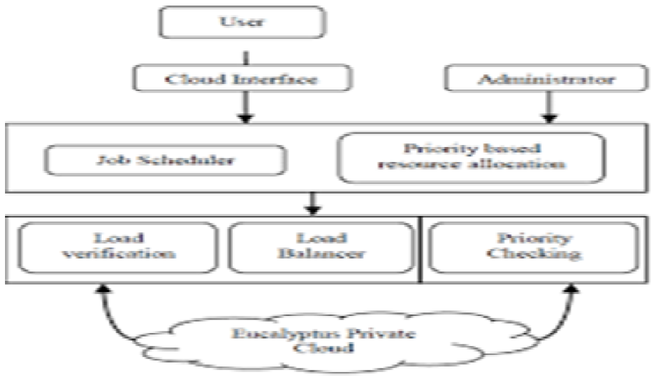

Figure. 2. Cloud Architecture

V. CONCLUSIONS

We have presented an approach for implementation and evaluation of a resource management system for cloud computing services. We have also shown in our paper of how we can multiplex virtual resource allocation to physical resource allocation effectively based on the fluctuating demand. We also make use the skewness metric to determine different resource characteristics appropriately so that the capacities of servers are well utilized. We can apply our algorithm to achieve both overload avoidance and green computing for systems which support multi-resource constraints. Cloud computing technology is increasingly being used in enterprises and business markets. In cloud paradigm, an effective resource allocation strategy is required for achieving user satisfaction and maximizing the profit for cloud service providers. This paper summarizes the classification of RAS and its impacts in cloud system. Some of the strategies discussed above mainly focus on CPU, memory resources but are lacking in some factors. Hence this survey paper will hopefully motivate future researchers to come up with smarter and secured optimal resource allocation algorithms and framework to strengthen the cloud computing paradigm.

VI. REFERENCES

[1]. A.Singh ,M.Korupolu and D.Mohapatra. Server-storage virtualization:Integration and

Load balancing in data centers. In Proc.2008 ACM/IEEE conference on supercomputing (SC'08) pages 1- 12, IEEE Press 2008.

[2]. ndrzejKochut et al. : Desktop Workload Study with Implications for Desktop Cloud Resource Optimization,978-1-4244-6534-7/10 2010 IEEE.

[3]. Atsuo Inomata, TaikiMorikawa, Minoru Ikebe, Sk.Md. MizanurRahman: Proposal and Evaluation of Dynamin Resource Allocation Method Based on the Load Of VMs on IaaS(IEEE,2010),978-1-4244-8704-2/11.

[4]. D. Gmach, J.RoliaandL.cherkasova, Satisfying service level objectives in a self-managing resource pool. In Proc. Third IEEE international conference on self-adaptive and self organizing system.(SASO'09) pages 243-253.IEEE Press 2009 .

[5]. David Irwin, PrashantShenoy, Emmanuel Cecchet and Michael Zink:Resource Management in Data-Intensive Clouds: Opportunities and Challenges .This work is supported in part by NSF under grant number CNS-0834243.

[6]. Dongwan Shin and HakanAkkan :Domain-based virtualized resource management in cloud computing.

[7]. Dorian Minarolli and Bernd Freisleben: Uitlity –based Resource Allocations for virtual machines in cloud computing(IEEE,2011),pp.410-417.

[8]. DusitNiyato, Zhu Kun and Ping Wang : Cooperative Virtual Machine Management for Multi-Organization Cloud Computing Environment.

[9]. FetahiWuhib and Rolf Stadler : Distributed monitoring and resource management for Large cloud environments(IEEE,2011),pp.970-975.

[10]. HadiGoudaezi and MassoudPedram : Multidimensional SLA-based Resource Allocation for Multi-tier Cloud Computing Systems IEEE 4th International conference on Cloud computing 2011,pp.324-331.

[11]. HadiGoudarzi and MassoudPedram: Maximizing Profit in Cloud Computing System Via Resource Allocation: IEEE 31st International Conference on Distributed Computing Systems Workshops 2011: pp,1- 6.

[12]. Hien et al. ,'Automatic virtual resource management for service hosting platforms, cloud'09,pp 1-8. 13Hien Nguyen et al.: SLA-aware Virtual Resource Management for Cloud Infrastructures: IEEE Ninth International Conference on Computer and Information Technology 2009, pp.357-362.

[13]. I.Popovici et al,"Profitable services in an uncertain world". In proceedings of the conference on supercomputing CSC2005.

[14]. Jiyani et al.: Adaptive resource allocation for preemptable jobs in cloud systems (IEEE, 2010), pp.31-36.

[15]. Jose Orlando Melendez &shikhareshMajumdar: Matchmaking with Limited knowledge of Resources on Clouds and Grids.

[16]. K.H Kim et al. Power-aware provisioning of cloud resources for real time services. In international workshop on Middlleware for grids and clouds and e-science, pages 1-6, 2009.

[17]. Karthik Kumar et al.: Resource Allocation for real time tasks using cloud computing (IEEE, 2011), pp. 19Keahey et al., "sky Computing",Intenet computing, IEEE,vol 13,no.5,pp43-51,sept-Oct2009

[18]. Kuo-Chan Huang &Kuan-Po Lai: Processor Allocation policies for Reducing Resource fragmentation in Multi cluster Grid and Cloud Environments (IEEE, 2010), pp.971-976.

[19]. Linlin Wu, Saurabh Kumar Garg and Raj kumarBuyya: SLA –based Resource Allocation for SaaS Provides in Cloud Computing Environments (IEEE, 2011), pp.195-204 .

[20]. Lizhewang,JieTao,Kunze M.,Castellanos,A.C,Kramer,D.,Karl,w, "High Performance Computing and

Communications",IEEE International Conference HPCC,2008,pp.825-830.

[21]. M.SuhailRehman,MajdF.Sakr : Initial Findings for provisioning Variation in Cloud Computing(IEEE,2010),pp.473-479 .

[22]. P.Ruth,J.Rhee, D.Xu, R.Kennell and S.Goasguen, "Autonomic Adaptation of virtual computational environments in a multi-domain infrastructure", IEEE International conference on Autonomic Computing, 2006,pp.5-14.

[23]. Patricia Takako Endo et al. :Resource allocation for distributed cloud :Concept and Research challenges(IEEE,2011),pp.42-46 .

[24]. Paul Marshall, Kate Keahey& Tim Freeman: Elastic Site(IEEE,2010),pp.43-52.

[25]. PenchengXiong,Yun Chi, Shenghuo Zhu, Hyun Jin Moon, CaltonPu&HakanHacigumus: Intelligent Management Of Virtualized Resources for Database Systems in Cloud Environment(IEEE,2011),pp.87-98.

[26]. RerngvitYanggratoke, FetahiWuhib and Rolf Stadler: Gossip-based resource allocation for green computing in Large Clouds: 7th International conference on network and service management, Paris, France, 24-28 October, 2011

[27]. Zhen Xiao, Senior Member, IEEE, Weijia Song, and Qi Chen, "Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment", VOL. 24, NO. 6, JUNE 2013

[28]. L. Siegele, "Let It Rise: A Special Report on Corporate IT," The Economist, vol. 389, pp. 3-16, Oct. 2008.

[29]. P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, "Xen and the Art of Virtualization," Proc ACM Symp. Operating Systems Principles (SOSP '03), Oct. 2003.

[30]. "Amazon elastic compute cloud (Amazon EC2)," http://aws. amazon.com/ec2/, 2012.

[31]. C. Clark, K. Fraser, S. Hand, J.G. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Warfield, "Live Migration of Virtual Machines," Proc. Symp.Networked Systems Design and Implementation (NSDI '05), May 2005.

[32]. M. Nelson, B.-H. Lim, and G. Hutchins, "Fast Transparent Migration for Virtual Machines," Proc. USENIX Ann. Technical Conf., 2005.