Proposing Priority based Dynamic Resource Allocation [PDRA] Model in Cloud Computing

Amit Chaturvedi, PhD
MCA Deptt
Govt. Engineering College, Ajmer
Rajasthan, India

Praveen Sengar
Ph.D. Scholar,
Bhagwant Univ., Ajmer
Rajasthan, India

Kalpana Sharma, PhD
Computer Sc. Deptt
Bhagwant Univ., Ajmer
Rajasthan, India

ABSTRACT
Scaling of resources in cloud computing is essential for the better utilization of resources. Dynamic allocation of the resources / VMs in the multi-tenant environment is the need of the cloud computing. Virtualization technologies evolved to help IT organizations and to improve the efficiency of their hardware resources by partitioning hardware to provide simultaneous support to multiple applications and their corresponding software stacks. If the resource utilization is not properly allocated to applications, it will lead to the faulty services to the customers. The Cloud is the hub of resources, and can be used by any client on rental bases and on no demand resources can left with no usage. Clients/ Brokers may request for the multiple VMs/ other resources like, applications, database, operating system etc, but the resources are limited. So, there is the need of such a system to handle this allocation and deallocation of resources or VMs. By this PDRA model, authors have presented an idea to handle the resources/ VMs allocation and deallocation system.

General Terms
Your general terms must be any term which can be used for general classification of the submitted material such as Pattern Recognition, Security, Algorithms et. al.

Keywords
Virtual Machine, dynamic allocation, elasticity, cloud, scaling..

1. INTRODUCTION
Introduction: Scaling of resources in cloud computing is essential for the better utilization of resources. As we know that the main purpose of the cloud computing is to share computing resources like applications, database, operating system, etc. to the world of computers for maximizing the utilization and return. Cloud scaling enables scale-up and down automatically. We can create thousands of server instances, brokers, virtual machines etc and allocate them simultaneously. These instances can be controlled separately by the medium of middleware known as Virtual Machine Monitors or Hypervisors. Flexibility in creating and hosting the resources and its services can provide multiple choices for instance creating and could be configured for the memory requirements, operating system, processors, virtual machines, etc. As the number of virtual machines increases, they want more operating memory, stable storage, processing capability, and obviously more bandwidth to meet the required performance. To tackle the workload on servers, there is the requirement of dynamic provisioning of different data centres with guarantee of availability of the resources and services required.

In virtualized multi-tenant cloud environment, applications encapsulate and segregate application performance by using virtual machines (VMs). Virtualization technologies evolved to help IT organizations and to improve the efficiency of their hardware resources by partitioning hardware to provide simultaneous support to multiple applications and their corresponding software stacks. If the resource utilization is not properly allocated to applications, it will lead to the faulty services to the customers. Resource scaling can resolve issues of application migration conflicts as well, which are due to compatibility issues in multi-tenant cloud environment. Resource scaling should do check of compatibility of applications during application transfer to cloud. Resource scaling is also very important in case of processor availability so that there should be no issue of service due to technical snag during execution of applications.

Multi-tenant clouds are central resources for computing power, and ultimately the cloud organization can distribute same resource among multiple clients without mingling each other's data. The Cloud is the hub of resources, and can be used by any client on rental bases and on no demand resources can left with no usage. Running huge data centres is unsustainable for organizations with server's resources being idle most of the time. Small companies can upload their data to cloud and take's cloud resources on rental bases. Resource scaling issues are caused by clients most of the time as they are not aware of the workload demand by their users. Customers can be increased at the times or stayed low sometimes. Resources allocation and de-allocation is important to schedule resources perfectly while entertaining requests from clients. The whole point of cloud computing is to achieve economies of scale by managing a very large pool of computing resources in a highly economic and efficient fashion.

By using the elasticity of computing resources, systems may occupy and release resources to dynamic workloads and paying for those only which are needed and used. This characteristic of cloud computing is basically known elasticity and the allocation and de-allocation dynamically by controlling through an automated system is the rapid elasticity. The multi-tier applications services that allocates and relieves resources in segments, such as virtual server instances of predefined sizes. It highlights on elastic control of the storage tier, in which storage and removing from machines or brick needs re-balancing stored data on all the machines .the new challenges for storage tier presents for elastic controls.

2. RELATED WORK
In cloud computing various types of resources like CPU, Memory, OS, Application Software etc. are used. A cloud server, which has sufficient resources all the time for its clients as resource pools, efficiently and dynamically allocates or deallocates these resources, is considered good for its

clients. Here, in this paper, we have presented a study on various types of resource pools like physical server pools, virtual server pools, storage pools, network pools etc available in cloud computing, various allocation strategies, and virtualization tools used in cloud computing to manages these resources for client's request [1].

Elastic resource allocation in cloud computing paradigm is managing the resource elastically from anywhere and in anytime automatically. Elasticity is the fundamental characteristic by combining on-demand self service and resource pooling techniques in cloud, distributed, grid and utility computing circumstance. Realizing the significance of elasticity in cloud computing, this paper gives comprehensive analyses of elastic resource management and scheduling in cloud. In addition to, this paper discourse open challenges and future directions connected with role of elasticity mechanism in cloud computing [2].

Cloud computing offers utility-oriented IT services to users worldwide. Based on a pay-as-you-go model, it enables hosting of pervasive applications from consumer, scientific, and business domains. However, data centers hosting Cloud applications consume huge amounts of electrical energy, contributing to high operational costs and carbon footprints to the environment. The basic principle of cloud computing is that user data is not stored locally but is stored in the data center of internet. To gain the maximum degree of the benefits, the services offered in terms of resources should be allocated optimally to the applications running in the cloud. The following section discusses the significance of resource allocation. This research work was focused on the design and implementation of an automated resource management system that achieves a good balance. Main objective of this work is implementation of a resource allocation policy that can avoid overload in the system effectively while minimizing the number of servers [3].

Cloud Computing environment provisions the supply of computing resources on the basis of demand, as and when needed. It builds upon advances of virtualisation and distributed computing to support cost efficient usage of computing resources, emphasizing on resource scalability and on-demand services. It allows business outcomes to scale up and down their resources based on needs. Managing the customer demand creates the challenges of ondemand resource allocation. Virtual Machine (VM) technology has been employed for resource provisioning. It is expected that using virtualized environment will reduce the average job response time as well as executes the task according to the availability of resources. Hence VMs are allocated to the user based on characteristics of the job. Effective and dynamic utilization of the resources in cloud can help to balance the load and avoid situations like slow run of systems. This paper mainly focuses on allocation of VM to the user, based on analyzing the characteristics of the job. Main principle of this work is that low priority jobs (deadline of the job is high) should not delay the execution of high priority jobs (deadline of the job is low) and to dynamically allocate VM resources for a user job within deadline [4].

The 'Pay-as-per-use' model of cloud computing requires a user to rent cloud resources or services for a period of time and make payment accordingly. This paper proposes a demand-based preferential resource allocation technique that designs a market-driven auction mechanism to identify users for resource allocation based on their payment capacities and implements a payment strategy based on a buyer's service preferences. A comparison is drawn between the proposed allocation technique and the famous off-line VCG auction mechanism and results show a performance benefit in revenues to service provider, payments of cloud users besides ensuring an optimum resources use [5].

A large number of resources are integrated into a data center to provide various resource services in cloud computing. A major challenge is how to provide resources timely and accurately to satisfy users' demands. However, users' resource demands change constantly and sometimes fluctuate very strong. The resource provision may be not performed in time. And even, sometimes the active physical resources may be too insufficient to satisfy users' demands because some of them are shut down in order to reduce energy. So it is important to provide a proactive resource provision to guarantee good users' experiences in cloud computing. The key is to predict the future resource demands accurately to support resource provision in advance. In this paper, we propose a resource demand prediction method EEMD-ARIMA based on ensemble empirical mode decomposition (EEMD) in cloud computing. This method decomposes the non-stationary users' resource demands into a plurality of intrinsic mode function components (IMFs) and residual component (RES) through EEMD method to improve the prediction accuracy. The experimental results show that our method has a higher prediction accuracy compared with the existing ARIMA prediction model in the short term prediction of cloud resource demands [6].

Designing market-based mechanism that benefits both the cloud customer and cloud provider in a cloud market is a fundamental but complex problem. Double auction is one such mechanism to allocate re- sources that prevents monopoly and is used to design an unbiased optimal market strategy for cloud market. This work proposes a truthful combinatorial double auction mechanism for allocation and pric- ing of computing resources in cloud. For resource allocation, utilitarian social welfare maximization prob- lem is formulated using Integer Linear Programming (ILP) and a near optimal solution is obtained using Linear Programming based padded method. For payment, truthful and novel schemes are designed for both customers and providers. Moreover, the proposed mechanism is individual rational, computationally tractable, weakly budget-balance and asymptotic efficient. Performance evaluation and comparative study exhibit that the proposed mechanism is effective on various performance metrics such as utilitarian so- cial welfare, total utility, customers' satisfaction, providers' revenue and hence is applicable in real cloud environments [7].

In cloud computing environment, resources can be dynamically provisioned on deman for cloud services. The amount of the resources to be provisioned is determined during runtime according to the workload changes. Deciding the right amount of resources required to run the cloud services is not trivial, and it depends on the current workload of the cloud services. Therefore, it is necessary to predict the future demands to automatically provision resources in order to deal with fluctuating demands of the cloud services. In this paper, we propose a hybrid resource provisioning approach for cloud services that is based on a combination of the concept of the autonomic computing and the reinforcement learning (RL). Also, we present a framework for autonomic resource provisioning which is inspired by the cloud layer model. Finally, we evaluate the effectiveness of our approach under two real world workload traces. The experimental results show that the proposed approach reduces the total cost

by up to 50%, and increases the resource utilization by up to 12% compared with the other approaches [8].

Users and providers have different requirements and objectives in an investment market. Users will pay the lowest price possible with certain guaranteed levels of service at a min- imum and providers would follow the strategy of achieving the highest return on their investment. Designing an optimal market-based resource allocation that considers the ben- efits for both the users and providers is a fundamental criterion of resource management in distributed systems, especially in cloud computing services. Most of the current market- based resource allocation models are biased in favor of the provider over the buyer in an unregulated trading environment. In this study, the problem was addressed by proposing a new market model called the Combinatorial Double Auction Resource Allocation (CDARA), which is applicable in cloud computing environments. The CDARA was prototyped and simulated using CloudSim, a Java-based simulator for simulating cloud computing environ- ments, to evaluate its efficiency from an economic perspective. The results proved that the combinatorial double auction-based resource allocation model is an appropriate market- based model for cloud computing because it allows double-sided competition and bidding on an unrestricted number of items, which causes it to be economically efficient. Further- more, the proposed model is incentive-compatible, which motivates the participants to reveal their true valuation during bidding [9].

Cloud computing offers on-demand network access to the computing resources through virtualization. This paradigm shifts the computer resources to the cloud, which results in cost savings as the users leasing instead of owning these resources. Clouds will also provide power constrained mobile users accessibility to the computing resources. In this paper, we develop performance models of these systems. We assume that jobs arrive to the system according to a Poisson process and they may have quite general service time distributions. Each job may consist of multiple numbers of tasks with each task requiring a virtual machine (VM) for its execution. The size of a job is determined by the number of its tasks, which may be a constant or a variable. The jobs with variable sizes may generate new tasks during their service times. In the case of constant job size, we allow different classes of jobs, with each class being determined through their arrival and service rates and number of tasks in a job. In the variable case a job generates randomly new tasks during its service time. The latter requires dynamic assignment of VMs to a job, which will be needed in providing service to mobile users. We model the systems with both constant and variable size jobs using birth–death processes. In the case of constant job size, we determined joint probability distribution of the number of jobs from each class in the system, job blocking probabilities and distribution of the utilization of resources for systems with both homogeneous and heterogeneous types of VMs. We have also analyzed tradeoffs for turning idle servers off for power saving. In the case of variable job sizes, we have determined distribution of the number of jobs in the system and average service time of a job for systems with both infinite and finite amount of resources. We have presented numerical results and any approximations are verified by simulation. The results of the paper may be used in the dimensioning of cloud computing centers [10].

Cloud computing is an attractive computing model since it allows for the provision of resources on-demand. Cloud computing has emerged as a new technology that has got huge potentials in enterprises and markets. Clouds can make it possible to access applications and associated data from anywhere. Companies are able to rent resources from cloud for storage and other computational purposes so that their infrastructure cost can be reduced significantly. Hence there is no need for getting licenses for individual products. Cloud Computing offers an interesting solution for software development and access of content with transparency of the underlying infrastructure locality. The Cloud infrastructure is usually composed of several data centers and consumers have access to only a slice of the computational power over a scalable network. The provision of these computational resources is controlled by a provider, and resources are allocated in an elastic way, according to consumers' needs. However one of the major pitfalls in cloud computing is related to optimizing the resources being allocated. The other challenges of resource allocation are meeting customer demands and application requirements. In this paper, modified round robin resource allocation algorithm is proposed to satisfy customer demands by reducing the waiting time [11].

Cloud computing is a hybrid paradigm which makes use of utility computing, high performance cluster computing and grid computing and it offers various benefits such as flexibility, expandability, little or almost no capital investment, disaster recovery, moveable work space and much more. However, due to constantly increasing number of data centers worldwide, the issue of energy consumption by these data centers has attracted attention of researchers. Resource allocation and resource utilization are the major criterion in which the problem of energy efficiency can be addressed. In this research, we aim to provide energy-efficient resource allocation using Multi-Objective Optimization (MOO) method. Further, We propose MOO-based virtual machine (VM) allocation policy and implement it in CloudSim environment. The results are compared with existing policies. The results depict that MOO-based policy leads to saving in energy due to efficient resource allocation, without compromising performance of data center operations [12].

Compared to traditional distributed computing paradigms, a major advantage of cloud computing is the ability to provide more reliable, affordable, flexible resources for the applications (or users). The need to manage the applications in cloud computing creates the challenge of on-demand resource provisioning and allocation in response to dynamically changing workloads. Currently most of these existing methods focused on the optimization of allocating physical resources to their associated virtual resources and migrating virtual machines to achieve load balance and increase resource utilization. Unfortunately, these methods require the suspension of the cloud computing applications due to the mandatory shutdown of the associated virtual machines. In this paper, we study the resource allocation at the application level, instead of studying how to map the physical resources to virtual resources for better resource utilization in cloud computing environment. We propose a threshold-based dynamic resource allocation scheme for cloud computing that dynamically allocate the virtual resources (virtual machines) among the cloud computing applications based on their load changes (instead of allocating resources needed to meet peak demands) and can use the threshold method to optimize the decision of resource reallocation. The proposed threshold-based dynamic resource allocation scheme is implemented by using CloudSim, and

experimental results show the proposed scheme can improve resource utilization and reduce the user usage cost [13].

Cloud computing is a new generation of computing based on virtualization technology. An important application on the cloud is the Database Management Systems (DBMSs). The work in this paper concerns about the Virtual Design Advisor (VDA). The VDA is considered a solution for the problem of optimizing the performance of DBMS instances running on virtual machines that share a common physical machine pool. It needs to calibrate the tuning parameters of the DBMS's query optimizer in order to operate in a what-if mode to accurately and quickly estimate the cost of database workloads running in virtual machines with varying resource allocation. The calibration process in the VDA had been done manually. This manual calibration process is considered a complex, time-consuming task because each time a DBMS has to run on a different server infrastructure or to replace with another on the same server, the calibration process potentially has to be repeated. According to the work in this paper, an Automatic Calibration Tool (ACT) has been introduced to automate the calibration process.

Also, a Greedy Particle Swarm Optimization (GPSO) search algorithm has been proposed and implemented in the VDA instead of the existed greedy algorithm to prevent the local optimum states from trapping the search process from reaching global optima. The main function of this algorithm is to minimize the estimated cost and enhance the VMs configurations.

The ACT tool and the GPSO search algorithm have been implemented and evaluated using TPC-H benchmark queries against PostgreSQL instances hosted in Virtual Machines (VMs) on the Xen virtualization environment [14].

Growing interest in Cloud Computing places a heavy workload on cloud providers which is becoming increasingly difficult for them to manage with their primary datacenter infrastructures. Resource limitations can make providers vulnerable to significant reputational damage and it often forces consumers to select services from the larger, more established companies, sometimes at a higher price. In order to increase their capacity, cloud providers need to invest heavily in costly hardware. Funding limitations commonly prevent emerging and even established providers from making continual investments of this type speculatively assuming a certain level of growth in demand. As an alternative, they may strive to use the current inter-cloud resource sharing platforms. These however mainly rely on monetary payments which can put pressure on already stretched cash flows and transaction costs may reduce profitability. To address such issues, we have designed and implemented a new multi-agent based Cloud Resource Bartering System (CRBS) that fosters the management and bartering of pooled resources without requiring costly financial transactions between providers. Unlike existing systems, CRBS assigns resources by considering resource urgency which comparatively improves customers' satisfaction and the resource utilization rate by more than 50%. The evaluation of CRBS provides evidence that it assists providers to timely acquire the additional resources and to maintain sustainable service delivery. We conclude that the existence of such a system is economically beneficial for cloud providers and enables them to adapt to fluctuating workloads [15].

3. IMPORTANT FINDINGS

The cloud computing technology is based on virtualization. Virtualization is a technology that separates computation functions from physical hardware. It allows the users to partition and multiplex physical machine infrastructure (e.g., CPU, memory, I/O, storage, and network interface cards). The applications are running on virtual machines instead of physical ones. The Virtual Machine (VM) is a software implementation of a computing environment to simulate a physical machine directly executing on physical hardware. The Virtual Machine Monitor (VMM) is used to create and manage the VMs (e.g., Xen, VMware, VirtualBox, and KVM). The virtual machine configuration or resource allocation controls the sharing of physical resources (CPU, memory, I/O bandwidth) allocated to VMs. The problem of optimizing the performance of the virtualized applications (i.e., the applications that run on VMs) is critical to the success of the cloud computing paradigm, because VM configuration affects the application performance.

On the other hand, The Database Management System (DBMS) is considered one of the applications deployed on the cloud. Each DBMS instance has its own tuning parameters. The tuning parameters interact with cost model in DBMS' query optimizer to change the performance (e.g., CPU parameters and buffer parameters). how DBMS instances can get a benefit of resource allocation for each VM in the shared physical pool, this called Virtualization Design Problem (VDP). Virtual Design Advisor (VDA) is a technique that offers a solution for such problem. It gives recommended configurations for multiple VMs running different workloads among shared resources [2,7–9]. It explores the characteristics of workloads to distinguish their intensity (e.g., CPU or I/O intensive, etc.) and makes a decision for best resource allocation for VM which run this workload. The DBMS has a query optimizer tool to choose the best execution plan based on the estimated cost.

The query optimizer's cost model is not aware of virtualized environment because it takes the default values of tuning parameters. So, the query optimizer parameters are needed to be calibrated in order to be aware of different resource allocation in virtualized environment. An Automatic Calibration Tool (ACT) has been introduced to tune parameters of DBMS query optimizer in virtualized environment to solve the manual calibration problem in the VDA. A Particle Swarm Optimization (PSO) is considered a modern evolutionary algorithm which is used to explore the search space of a given problem. It is used to find optimal or near-optimal solutions for maximization/minimization search problems. Search algorithm called Greedy Particle Swarm Optimization (GPSO) has been proposed to overcome the local optimum problem of the existed greedy algorithm in the VDA. The proposed GPSO algorithm is considered an amalgamation of heuristic greedy search and particle swarm optimization to optimize configurations based on the workload profile in virtualized environments. The GPSO algorithm has been implemented in the VDA enumerator module, which initially makes an equal resource allocation of VMs and adapts these allocations based on the estimated cost obtained by cost models of the database system query optimizer. To evaluate the ACT tool and the GPSO search algorithm, prototype experiments have been conducted based on the optimal CPU allocation for the different virtual machines.

4. PROPOSED MODEL AND METHODOLOGY

When a Broker or client have some job or task to perform on the cloud server, it will request to allot VMs [virtual machine]. If the VMs are available with the server, it will

assign a VM to the client or the broker and job is preceded with the provided resources as requested. If the resources or the VM, requested by the new job having high priority is not available, i.e. all the VMs / resources are already allotted to the clients/ brokers. Then the server has to check the priority status of the new job, in comparison with the already continuing job's priority. In the proposed Priority Based Dynamic Resource Allocation [PDRA] Model, the job may fall under any of the three lease types i.e. Cancellable, suspendable, and non-preemptable. Hence, the server will check the lease status among the low priority jobs, already in process, in comparison with the new job. The preference should be given to the cancellable jobs than the suspendable jobs, the server should avoid the non-preemptable jobs.

If the resources / VMs of the cancellable job are allotted to the new high priority job, then there is no problem for further allotment. But if the resources/VMs of the suspendable job are allotted to the new high priority job, then this suspended job should be kept in the suspended queue by marking the status of the job executed to continue afterwards, when resources/VMs again collected. This suspended job will continue from the marked status done when it was suspended. This PDRA model will help to properly regulate the job scheduling on the server side in PaaS layer of the cloud computing. The steps followed in the PDRA model are illustrated by the figure 1 given below:

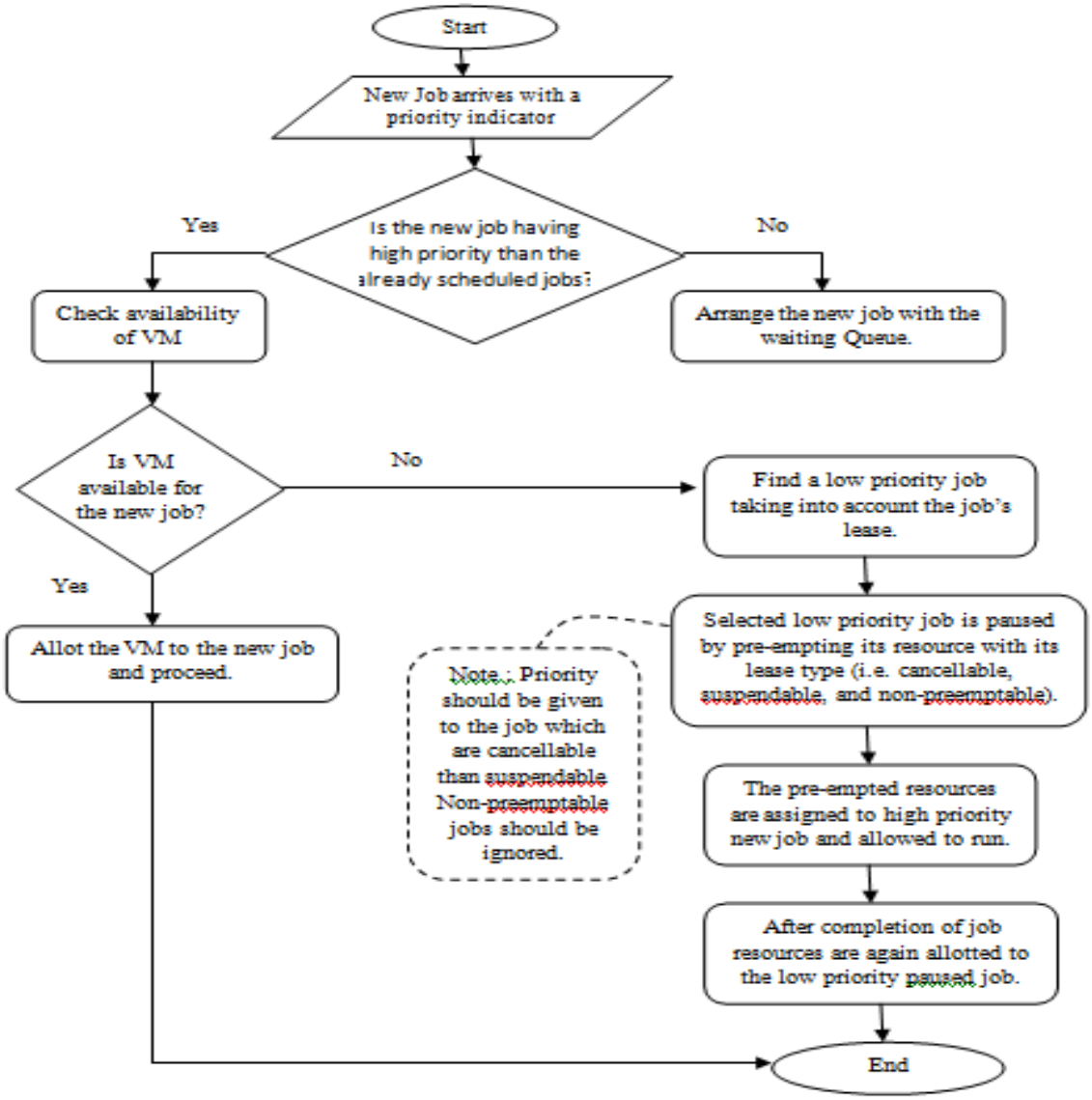

Figure 1: Priority Based Dynamic Resource Allocation [PDRA] Model

5. CONCLUSION

Allocation or deallocation of cloud resources like virtual machine, database, applications, operating systems etc, in a rapid manner with scale up or down is the prime requirement to provide a good virtual cloud environment. In the PDRA Model, author attempts to provide an idea to handle the limited resources or virtual machines, by marking the incoming jobs with any of three categories either cancellable, suspendable, or non-preemptable. If there will be the situation arises when the resource/VM of the already assigned job has to be taken back for the high priority job, then first it will be select the cancellable jobs, than the suspendable jobs. The non-preemptable jobs should be ignored, according to the PDRA model.

So, we can say the ongoing jobs are well handled in the PDRA model, and new jobs are also taken care with a good

virtual environment. The high priority jobs are attended by providing the necessary resource requirements. The lease based priority system will leads less crowded suspended queue. So, the overall performance of the system will be improved with satisfaction of proper handling and take care.

6. ACKNOWLEDGMENTS

The authors are thankful to all the contributors directly or indirectly for the preparation of this research paper especially Dr. K. K. Goyal for his continuous support.

7. REFERENCES

[1] A. Rashid, Dr. Amit Chaturvedi, "A Study on Resource Pooling, Allocation and Virtualization Tools used for Cloud Computing", International Journal of Computer Applications (0975 – 8887), Volume 168, No. 2 , June 2017

[2] Dr. M. Durairaj, P. Kannan, "Elastic Resource Allocation Challenges in Cloud computing System", nternational Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS), Vol. 12, issue 2, March -May 2015, pp. 180-184

[3] Nisha Sharma, Mamta Dhanda, "Improving Resource Allocation in Virtualized Cloud Environment", International Journal of Computer Science Trends and Technology (IJCST) –Volume 2, Issue 4, Jul-Aug 2014

[4] Saraswathi AT, Kalaashri Y R A, Dr. S. Padmavathi, "Dynamic Resource Allocation Scheme in Cloud Computing", Elsevier, Procedia Computer Science 47 (2015), pp. 30-36.

[5] N. Kumar, S. Saxena, "A Preference-based Resource Allocation In Cloud Computing Systems", 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015), Elsevier, Procedia Computer Science 57 (2015), pp. 104 – 111

[6] J. Chen, Y. Wang, "A Resource Demand Prediction Method Based on EEMD in Cloud Computing", 8th International Congress of Information and Communication Technology (ICICT -2018), Elsevier, Procedia Computer Science 131 (2018), pp. 116–123

[7] D. Kumar, G. Baranwal, Z. Raza, D.P. Vidyarthi, "A truthful combinatorial double auction-based marketplace mechanism for cloud computing", The Journal of Systems and Software 140 (2018), pp. 91–108

[8] M.G. Arani, S. Jabbehdari, M. A. Pourmina, "An autonomic resource provisioning approach for service-based cloud applications: A hybrid approach", Future Generation Computer Systems, ISSN : 0167-739X, 2017, pp. 1-20

[9] P. Samimi, Y. Teimouri, M. Mukhtar, "A combinatorial double auction resource allocation model in cloud computing", Information Sciences 357 (2016), pp. 201–216

[10] S. Vakilinia, M. M. Ali, D. Qiu, "Modeling of the resource allocation in cloud computing centers", Elsevier, Computer Networks (2015), ISSN : 1389-1286, pp. 1-18

[11] P. Pradhan, P. K. Behera, B N B Ray, "Modified Round Robin Algorithm for Resource Allocation in Cloud Computing", International Conference on Computational Modeling and Security (CMS 2016), Procedia Computer Science 85 (2016), pp. 878 – 890

[12] B. Shrimali, H. Patel, "Multi-objective optimization oriented policy for performance and energy efficient resource allocation in Cloud environment", Journal of King Saud University, Computer and Information Sciences (2017), ISSN : 1319-1578, pp 1-10

[13] W.Lin, J.Z.Wang, C. Liang, D. Qi, "A Threshold-based Dynamic Resource Allocation Scheme for Cloud Computing", PEEA 2011, Procedia Engineering 23 (2011),ISSN: 1877-7058, pp. 695-703.

[14] F. A. Omara, S.M. Khattab, R.Sahal, "Optimum Resource Allocation of Database in Cloud Computing", Egyptian Informatics Journal (2014) 15, pp. 1–12

[15] S. Z. Goher, P. Bloodsworth, R.Ur Rasool, R. McClatchey, "Cloud provider capacity augmentation through automated resource bartering", Future Generation Computer Systems (2017)