Asian Journal of Science and Applied Technology (AJSAT)
Vol.3.No.1 2015 pp 47-53
available at: www.goniv.com
Paper Received :08-03-2015
Paper Accepted:20-03-2015
Paper Reviewed by: 1. R. Venkatakrishnan 2. R. Marimuthu
Editor : Prof. P.Muthukumar

OPTIMIZED RESOURCE SELECTION FOR TASKS USING SPECTRAL CLUSTERING ON CLOUD

[1]A.MARYANIREKA., [2]MRS.P.ARASAI
[2]Associate Professor
[1]P.G Student (AU – C)
Department of Computer Science and Engg,
Trichy Engineering College,Trichy, India
anireka@gmail.com

ABSTRACT

With the high demand of delivering services to a user has led to the development of cloud computing. Cloud computing also called as the dynamic service provider; fulfill all the requirement of a model internet. In this scenario, Task scheduling plays an important role to reduce the turnaround time and increase resource utilization. But, scheduling is a critical problem in cloud. In this paper, an efficient task grouping based approach has been proposed for task scheduling in computational cloud. The decisions made by the corresponding algorithms should be judged based not only for user satisfaction, but also based on resource related performance metrics to maintain quality of service (Qos).To propose an algorithm for assigning the tasks to resources that minimizes the violations of the tasks' time requirements while simultaneously maximizing the resource' utilization efficiency for a given number of resources. The main concept of the paper is to grouping the task before resource allocation according to resource capacity, try to achieve non-preemptive scheduling. The partitioning is performed using a spectral clustering scheduling (SCS) through normalized cuts. A simulation of proposed approach using cloudSim toolkit is conducted. Experimental results show that the proposed algorithm outperforms for different values of the granularity and the load of the task requests.

***Keywords*-Scheduling, task grouping, Graph partitioning, CloudSim, Granularity**

I. INTRODUCTION

There are several ways to express the quality of service (Qos) requirements of the tasks submitted to a cloud environment. The most common way to describe Qos in a cloud is through task deadline. To realize the full potential of cloud computing ,cloud middleware needs to support various services such as security, uniform access, resource management, task scheduling and economic computation.Though,a range of essential services are to be integrated to accomplish a real cloud environment, among them scheduling is one of the most critical service component of the cloud middleware[5].Since, it is responsible for selecting best suitable virtual machines or computing resources with a goal of maximizing resource utilization and reduce proceeding cost and time meets the user and task requirement**.** Two main objective of the paperis (1) Minimize the time overlapping of the tasks assigned toa given resource. (2) Maximize the time overlapping among tasks assigned to different resources [6]. First objective focused on task assignment and second one focused on resource allocation. To solve the mentioned dual objective, graph partitioning techniques are exploited in this paper. In this case, graph partitioning is performed based on spectral clustering through the use of normalized cuts. Use of normalized cut have many advantages such as, it support unbalanced partition; secondly, automatically estimate the partition size with respect to the diversity of the tasks' duration within each partition.

A.Problem Definition

There is several papers use various scheduling algorithm that focused on resource allocation and task assignment, but in this paper the proposed spectral clustering scheduling (SCS) algorithm in cloud environment for scheduling is based on algebraic graph theory. It can find a global optimal solution of the graph cut objective function by constructing the graph Laplacian matrix[9] and it's Eigen-decomposition for scalability in cloud computing. Use of spectral method leads to good clustering results. It acts reasonably fast for sparse data sets of several thousand elements.

Traditional clustering algorithms are unable to meet the requirements of today's requirement in cloud. When dealing with high-dimensional data, some clustering algorithms that perform well in low dimensional data space are often unable to get good clustering results, and even invalid. Task Grouping Strategy is very effective technique to reduce communication overhead problem in task scheduling and better utilization of the resources on cloud [6]. ByEfficient, we mean appropriate resources are allocated at a right time to a right task, so that the task can utilize the resources effectively.

B. Scheduling model in cloud

In cloud computing, applications are submitted for use of cloud resources by users from their terminals. Theresource includes computing power, communication power and storage. An application consists of number of tasks; users want to execute these tasks in an efficient manner.In this figure1, general step of scheduling a task in cloud computing.The main components are scheduler, information repository, resources. In cloud computing, a user submits tasks to schedulers. Scheduler is connecting to the information repository, it's called as cloud information services (CIS).Scheduler is receiving tasks from user then scheduler is arranging the tasks according to criteria of tasks.

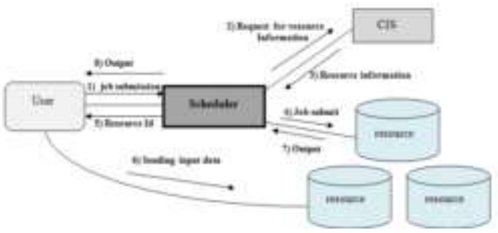

Fig.1. Scheduling process in cloud

Scheduler connects to CIS and gets information about the available resources and then scheduler compute to the resources. After compute the resources, tasks are submit on that resource which have high processing capabilities, then resource id is generated by CIS and id sent to user. So the resource is communicated to the user. The user then submits the input data or tasks directly on that resource and finally user gets the output from the resource through the scheduler. This is the general process of scheduling.

II. PROPOSED SYSTEM

In cloud computing environment, there are two players: cloud providers and cloud users. On one hand, providers hold massive computing resources in their large datacenters. On the other hand, there are users who have applications with fluctuating loads and least resources from providers to run their applications. First,a user sends a request for resources to a provider. When the provider receives the request, it looks for the resources to satisfy the request and assigns the resources to the requesting user, typically in the form of virtual machine (VM).Proper scheduling is needed for meet user's requirement and satisfies the quality of services.

Grouping strategy is very effective technique to solve the task scheduling problem and also better utilization of resources [7]. In this case, our proposed task allocation scheme as the spectral clustering Scheduling (SCS) algorithm.

We argue that a successful algorithm should take into account for both consideration, and we address the problem by proposing a SCS algorithm that, for a given number of resources, assigns tasks to processors so that (1) the time overlapping between tasks assigned on the same processor is minimized, while simultaneously (2) overallresource utilization efficiency is maximized.In cloud environment, the SCS algorithm used to find the number of resource virtual machines required to serve the grouped tasks with a given level of task time violations.

A. Proposed Task Scheduling Model

The four basic building blocks of cloud model are user,cloud information system (CIS) and resources. From the figure 2, the proposed scheduling, User submitted the tasks to the scheduler for scheduling to the resources.The scheduler is a service that resides in a user machine. When the user createsa list of tasks in the user machine, these tasks are sent to the scheduler for scheduling [5].

The scheduler obtains information of available resources from the cloud information service (CIS).Based on this information; task scheduling algorithm is used to grouping the tasks and then resource selection for groupedtasks. When all the tasks are put into groups with selected resources, the grouped tasks are dispatched to their corresponding resources for computation by the dispatcher.

The CIS service keeps track of all the resources characteristics of the cloud.CIS collects resource characteristic information like operating system, system architecture, processing capability, network bandwidth and processing cost. The information collector collects information from the cloud information service (CIS).It assembles the resource availability and processing capability to the resource information table.

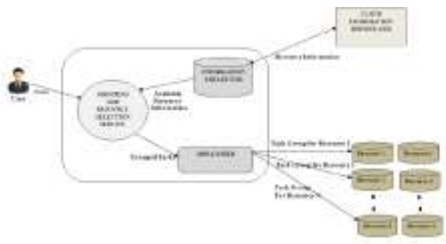

Fig. 2. Scheduling Model for task grouping.

The information collector is used by the grouping and resource selection to gather necessary information to perform task grouping. It also gathers information of the network bandwidth and processing cost of each listed resource provided by the CIS.

The grouping and resource selection service is responsible for grouping of tasks based on information collected by the information collector from CIS.In the task grouping process, user submitted tasks are collected by scheduler and tasks are grouped based on the selected available resource characteristics. The process iteratively performed until all the tasks are grouped according to the corresponding resources. The dispatcher acts as a sender that sends grouped tasks to their respective resources.

B. Architecture of Scheduler

The architecture of the scheduler system is described in figure 3.The system accepted tasks from the users specified by the TASK ID, TASK LENGTH (in Million Instructions (MI)), TASK INPUT FILE SIZE (in Mb) and the total number of tasks submitted by the user. After gathering details of user tasks, scheduler collects all the available resources information specified by their RESOURCE ID, RESOURCE MIPS (computational power of the resource in Million Instruction per second), RESOURCEBANDWIDTH (in Mb/sec) and RESOURCE COST (in cost/sec)[5].

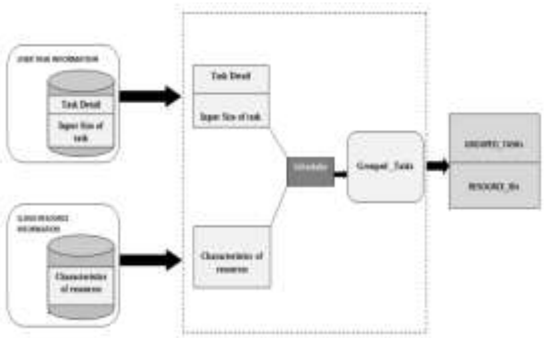

Fig.3. Architecture of Scheduler

After gathering the details of the user tasks and the available resources, the scheduler selects the resource and multiplies the resource MIPS with the given granularity time, which is the time within which a task is processed at the resource. The value of this calculation produces the total million instructions (MI) for that particular resource to process within the granularity time. New ids are assigned to grouped tasks, results go back to the corresponding users and the resource is again available to scheduler system.

III. PROPOSED ALGORITHM-SPETRAL CLUSTERING SCHEDULING

The proposed algorithm is in good accordance with the cloud computing paradigm, where users pay for the resource usage with a specific number of computation units (e.g., Amazon EC2).In this scenario, a user requests the time period, they want to use the cloud resources, expressed as requested start and finish times.

Then, the resource provider activates a set of virtual machines for serving the tasks. Since the activation of a virtual machine entails a cost for the resources provider, anoptimization strategy able on the other hand, to maximize the resource utilization and, on the other, minimizes the task overlaps and the consequent time shifts are of primary importance for a cloud computing provider.

Algorithmic Steps for the Spectral Clustering Scheduling:

- Step1 : Find the non-overlapping measures for all tasks. (Form a graph structure).
- Step2 : Form the respective matrices for the directed graph.
- Step3 : Compute the Eigen vectors and Eigen values for the above matrix.
- Step4 : Normalized cut uses the above Eigen value and vector for partition the graph.
- Step5 : Estimate the continuous matrix for continuous domain.
- Step6 : Refine the partitioning by running kernel k-means algorithm for clustering.

A.Algorithmic description

From the SCS algorithmic step one, find the measure of all non overlapping tasks and then form the directed graph for a graph partitioned problem, from these tasks. Dependencies among tasks are modeled using a directed Acyclic Graph $G = \{N, A\}$, where a node $u \in N$ represents a task, and an arc $u \rightarrow v \in A$ represents the dependence of task v on u; v cannot be executed before u finishes execution[6].

Secondly, from the directed graph,form the adjacency matrix as shown in the figure 4.An adjacency matrix is a means of representing which vertices (or nodes) of a graph are adjacent to which other vertices. Matrix representation for a graph is called incidence matrix. The relationship between a graph and the eigenvalues and eigenvectors of its adjacency matrix is studied in spectral graph theory.

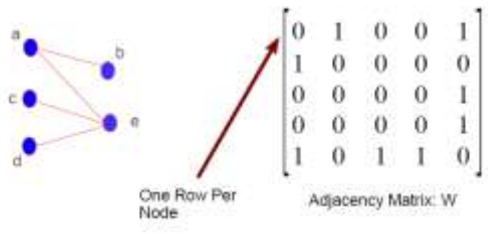

Fig.4. Form a Adjacency Matrix from a Graph Structure

Thirdly, a nonzero vector v of dimension *N*is an eigenvector of a square ($N \times N$) matrix A .i.e. AV= λV where λ is a scalar. Eigenvectors are the vectors that the lineartransformation A merely elongates or shrinks; Amount that they elongate/shrink depends upon the Eigen value. The set of solutions, *i.e.* the eigenvalues, is sometimes called the spectrum of A. Above computation are suitable for cloud Environment.

B. Normalized cut Spectral Graph Partitioning

Fourth, for normalized cut graph partitioning, Set up a weighted graph G= {V, E} and set the weight on the edge connecting two nodes. Solve for Eigen vectors with the smallest Eigenvalues[5]. Use the Eigen vector corresponding to the second smallest Eigenvalue to bipartition the graph into two groups. Recursively repartition the segmented parts if necessary. Normalized cut partitioning is an Eigen vector based algorithm. Uses of normalized cut method are 1) support unbalanced partition and gives spectral connection to k means clustering 2) Used to minimize the overlapping of tasks better than min cut partitioning.

C.Optimization in the continuous Domain

Weighted kernel k-means and normalized cuts using spectral clustering appear to be quite different.. We first unite these two forms of clustering under a singleframework. The normalized cut problem can be expressed as a trace maximization problem.As well as weighted kernel k-means also express as a trace maximization problem.So it makes a spectral connection[6].

Allow the matrix to take values in the continuous domain, if necessary.It is suitable for cloud environment for adding of the resources .Then shift to discrete domain within the interval, to fulfill the objective.

D. Kernel K-means clustering algorithm

A major drawback to *k*-means is that it cannot separate Clusters those are non-linearly separable in input space. Tworecent approaches have emerged for tackling such a problem.One is kernel *k*-means, where, before clustering, pointsare mapped to a higher-dimensional feature space using anonlinear function, and then kernel *k*-means partitions thepoints by linear separators in the new space. The otherapproach is spectral clustering algorithms, which use theeigenvectors of an affinity matrix to obtain aclustering ofthe data.

A popular objective function used in spectral clusteringis to minimize the normalized cut. Specifically, we can rewrite the weighted kernel k-means objective function as atrace maximization problem whose relaxation can be solvedwith eigenvectors [7].

Lastly, after partitioning through normalized cut, they are to be Cluster. From the figure 5, Clustering means grouping of data or dividing a large data set into smaller data set of some similarity. Nonlinear functions are used in k-means are suitable for cloud. The theoretical connection between spectral clustering and kernel k means helps in obtaining higher quality results.

The k means clustering algorithmic steps are

- Step1: Given N objects, K cluster centers
- Step2 : Assign each object to its closest cluster Center.
- Step3 : Update the center for each cluster
- Step4 : Repeat 2 & 3 until no change in each Cluster center.

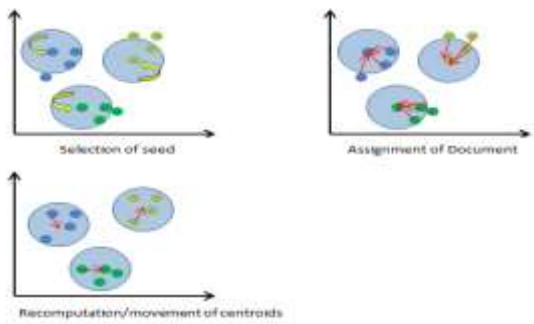

Fig.5. K means Clustering

IV. SIMULATED RESULTS

This section shows simulation and their result obtained from this work. The simulation can run with the help of JAVA because it's also collection of package.

A. Simulation Tool

CloudSim is a java based simulation tool, for our work, we select Net beans IDE to implement proposed VM allocation policy. Since Net beans is java base platform, so a java run time environment is needed before installing it. When this installation is completed, latest CloudSim package is extracted and it is imported in the NetBeans.In cloud computing, cloudSim is used as a simulation tool to implement this process.

In CloudSim, first of all data centers are created, then create broker [8] andthen created virtual machines (VM) are submitted to respective brokers. Number of datacenters is assumed to be known in advance.In cloudSim, the tasks are added to scheduling algorithm. In the scheduler, tasks or cloudlets are grouped according to the priority of normal task or speed task forms and run the tasks with the help of virtual machine through datacenter broker. Since our concept is to allocate the virtual machines dynamically, as per the task grouping given by the user request. So a graphical interface is prepared for selecting the position of virtual machines. When these lists are submitted, simulation with CloudSim starts.

B. Implementation of clustering algorithm in virtual machineusing cloudSim.

K-Means follows the normalized cut partitioned clustering approach. It involves partitioning the given data set using normalized cut graph partitioning and then into specific number groups called Clusters [2]. Each cluster in K means is associated with a center point called centroid. In the CloudSim, Datacenter Broker class is the base class for any VM level event [3].When createVmsInDatacenter (int datacenter) function is called, it performs the VM allocation to the respective datacenter.It consists of several datacenter for storing data. In a network, simulation contain several virtual machine can run in a single physical system.

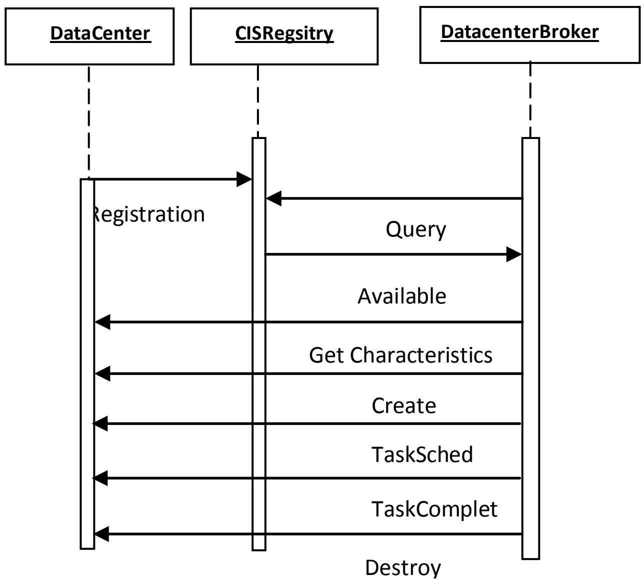

Fig. 6. Simulation Data Flow

After grouping cloudlet (user tasks) it is allocated to virtual machine, cloudSim is used and allow the tasks for execution.Figure 6 depicts the flow of communication among core CloudSim entities. In the beginning of the simulation, each Datacenter entity registers itself with the CIS (Cloud Information Service) Registry. CIS provides information registry type functionalities such as matchmakingservices for mapping user/brokers requests to suitable Cloud providers [10].

Datacenter Brokers acting on behalf of usersconsult the CIS service about the list of Clouds who offer infrastructure services matching user's application requirements. In case the match occurs the broker deploys the application with the Cloud that was suggested by the CIS.

The communication flow described so farrelates to the basic flow in a simulated experiment. Some variations in this floware possible depending on policies. For example, messages from Brokers to Datacenters may require a confirmation, from the part of the Datacenter, about the execution of the action, or the maximum number of VMs a user can create may benegotiated before VM creation.

C.Result Analysis

We were providing the stability comparisons of SCS from experimental results from the figure 7, to validate the results presented in the previous sections. Then, with the recognition data set, we show that using eigenvectors to initializekernel K-means gives better initial and final objective function values and better clustering results [10].

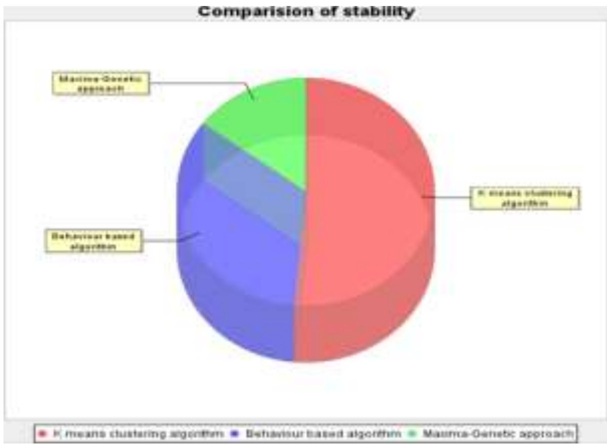

Fig.7. Comparison of SCS with Maxima Genetic & Behavior based algorithm

Thus the theoreticalconnection between spectral clustering and kernel k meanshelps in obtaining higher quality results. Finally, weshow that our SCS algorithm with other existing algorithm save a considerableamount of computation time, verifying the scalability of our approach.

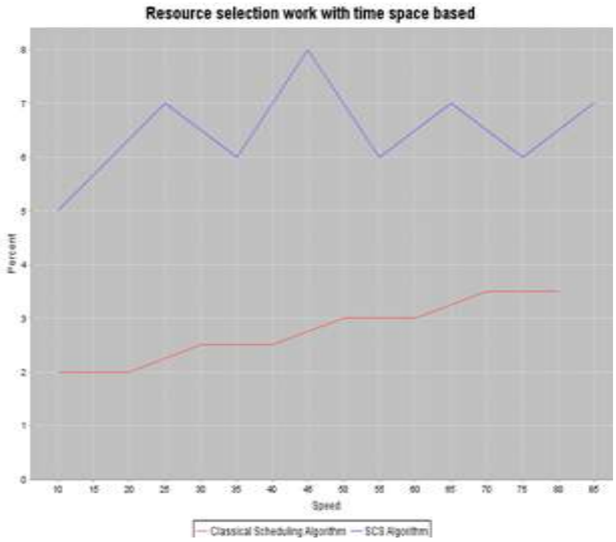

Fig.8. Performance comparison of Proposed Algorithm with classical scheduling.

V. CONCLUSION

Scheduling is a critical problem in cloud computing, because a cloud provider has to serve many users in cloud computing system. So scheduling is the major issue in establishing cloud computing systems. In this paper, we proposed an efficient task grouping based scheduling algorithm of spectral clustering scheduling scheme aims at maximizing processor utilization efficiency, while simultaneously minimizing the tasks processing time. The algorithm defines a no overlapping measure between tasks and then uses a matrix representation, the notation of generalized Eigenvalues to perform scheduling (graph partitioning) in the relaxed continuous space.

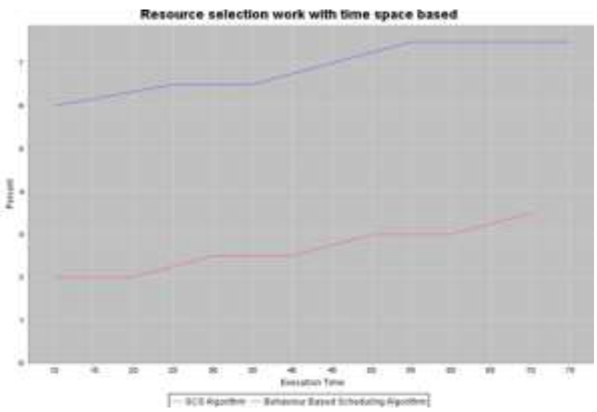

Fig. 9. Performance comparison of Proposed Algorithm with Genetic scheduling Algorithm.

In this paper we have analyzed various scheduling algorithm which efficiently schedules the computational tasks in cloud. We have created behavior based scheduling algorithm, classical uniprocessor scheduling algorithm and new proposed Scheduling algorithm is (SCS) Spectral graphical algorithm. Priority is an important issue of job scheduling in cloud environments. The experiment is conducted for varying number of Virtual Machines and workload traces. The experiment conducted is compared with other algorithm like uniprocessor scheduling and adaptive algorithm from the figure 8&9. The result shows that the proposed algorithm is more efficient than exciting algorithm.

VI. FUTURE WORK

SCS can be reviewed in this work, that there are different scope to be researched in cloud computing. Although integration of K-Means clustering works well than the existing methodologies in cloud environment but this also can be replaced with other clustering techniques. This K-Means algorithm is having best results for large collection of data. So clustering technology can also be implemented in the grouping of resource just like task grouping. Resource grouping is also a solution for improving the performance. It allows fast execution of task grouping avoid preemptive scheduling. Hence grouping of tasks and resources both can be used for cloud datacenters to make data clusters hence improve performance.

REFERENCES

[1] *ZhenXiao, Weijia Song, and Qi Chen“DynamicResource Allocation using VirtualMachines for Cloud”,IEEE Transaction on parallel anddistributed systems.*

[2] *Tejinder Sharma, Vijay Kumar Banga , “Efficient and Enhance Algorithm in Cloud Computing “,International Journal of SoftComputing and Engineering (IJSCE).*

[3] *MayankaKatyal and Atul Mishra,” Application of Selective Algorithm for Effective ResourceProvisioning In Cloud Computing Environment”, International Journal on Cloud Computing: Services and Architecture (IJCCSA)*

[4] *RamKumarSharma,NageshSharma,”ADynamic Optimization for Task Scheduling in Cloudwith Resource Utilization “, International Journal of Scientific Engineering andTech..*

[5] *B. Kruekaew and W. Kimpan ,” Virtual Machine Scheduling Management on Cloud Computing Using Artificial Bee Colony “Proceedingsof International MultiConference of Engineers and Computer Scientists 2014.*

[6] Nikolaos D. Doulamis, Panagiotis Kokkinos, and Emmanouel (Manos) Varvarigos," Resource Selection for Tasks with Time Requirements Using Spectral Clustering", IEEE Transsaction on computer VOL. 63, NO. 2, February 2014

[7] S.P.Jeno Lovesum, K.Krishnamoorthy, Blessed Prince. P."A Scheduling Technique forQOS Sensitive jobs in cloud", S.P.Jeno Lovesum et al. /International Journal of EngineeringTechnology (IJET).

[8] Rajkumar Buyya, Saurabh Kumar Garg, and Rodrigo, N. Calheiros "SLA-Oriented Resource Provisioning for Cloud Computing: Challenges, Architecture, and Solutions",2011 International Conference on Cloud and Service Computing.

[9] Arash Ghorbannia Delavar, Yalda Aryan," A Synthetic Heuristic Algorithm for Independent Task Scheduling inCloud Systems",IJCSI International Journal of ComputerScience Issues, Vol. 8, Issue 6, No 2, November 2011 ISSN (Online): 1694-0814.

[10] Dr. Amit Agarwal, Saloni Jain," Efficient Optimal Algorithm of Task Scheduling in Cloud Computing Environment",InternationalJournal of Computer Trends and Technology (IJCTT) – volume 9 number 7– Mar 2014.

[11] Rajkumar Buyya, Saurabh Kumar Garg, and Rodrigo,N. Calheiros "SLA-Oriented Resource Provisioning for CloudComputing: Challenges, Architecture, and Solutions",2011 International Conference on Cloud and Service Computing.

[12] Marchand, M. Silly-Chetto," On QoS Scheduling Services for Real-Time Periodic Tasks "Preprint submitted to Elsevier Science 30 June 2006, Journal of Real-Time Systems.

[13] Hong Sun, Shi-ping Chen, Chen Jin, Kai Guo,"Research and Simulation of Task Scheduling Algorithm in Cloud Computing"Research and Simulation of Task Scheduling Algorithm in Cloud Computing (Hong Sun).

[14] Bhupendra Panchal , Prof. R. K. Kapoor," Dynamic VM Allocation Algorithm using Clustering in Cloud Computing",

International Journal of Advanced Research in Computer Science and Software Engineering.

[15] Monika Choudhary, Sateesh Kumar Peddoju "Optimization Algorithm for Task Scheduling in Cloud", International Journal of Engineering Research and Applications (IJERA).

[16] Dr Ajay jangra, Tushar Saini ," Scheduling Optimization in Cloud Computing ",International Journal of Advanced Research in Computer Science and Software Engineering

[16] Yogita Chawla1 and Mansi Bhonsle," A Study on Scheduling Methods in CloudComputing", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)