ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 1, Issue 8, October 2012

HOST SELECTION METHODOLOGY IN CLOUD COMPUTING ENVIRONMENT

Pawan Kumar, Pijush Kanti Dutta Pramanik
Computer Science & Engineering Department
B. T. Kumaon Institute of Technology
Dwarahat, Almora, (Uttarakhand), India

***Abstract*—Cloud computing is a paradigm in which IT (information technology) application provide as a service. It allows users to utilize on-demand computation over internet, which is helpful for storage of data and services from around the world in commercialize manner. In cloud environment, applications need access to mass datasets that may each be replicated on different resources (or data centers) and Mass data moving from user to host and hosts to user. Based on the above two points, how to select best host for accessing resources and creating a virtual machine(VM) to execute applications to make execution efficiency high and access cost low as far as possible simultaneously is a challenging and urgent problem. In this paper, a host selection model based on minimum network delay using WSCP combined with Max-Min Heuristic. To select the host and schedule multiple jobs on multiple machines in an efficient manner is proposed, the objective is to minimize propagation time of input and output data by selecting nearest host into the network and finally it minimize the execution time of cloudlet.**

***Index Terms*—cloud Computing, WSCP, Max-Min Heuristic.**

I. INTRODUCTION

Following distributed computing, parallel computing, grid computing, utility computing, Web 2.0. etc., the computer industry and academia put forward cloud computing model [1], which achieves generalization and commercialization of these previous models in some sense [2]. Cloud computing, the long-held dream of computing, has the potential to change a large part of the IT industry, making it even more attractive as a service and shaping the way IT hardware is designed and purchase[3]. No doubts it would increasingly change the way people live and work. Cloud computing can be defined as "a type of parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service providers and consumers"[1].

The cloud computing is still at its infant stage and a very new technology for enterprises. Cloud computing is term used to describe both a platform and types of application. As a platform its supplies, configure and reconfigures servers, while servers can be physical machine or virtual machine. On the other hand, cloud computing describes application that are extended to be accessible through the internet and for this large data centers and powerful servers are used to host the web application and web services.

There are some important points in the definition to be discussed regarding cloud computing. Cloud computing differs from tradition computing paradigm as it is scalable, can been capsulated as an abstract entity which provides different level of services to the clients, driven by economies of scale and the service are dynamically configurable.

As a very new technology for enterprises there are many benefits stated of cloud computing by different researchers which make it more preferable to be adopted by enterprises. Cloud computing infrastructure allows achieving more efficient use of there IT hardware and software investments. This is achieve by breaking down the physical barrier inherent in isolated systems, automating the management of the group of the systems as a entity.

Manuscript received Oct, 2012.
***Pawan Kumar**, Computer Science and Engg., UTU/Bipin Chandra Tripathi Kumaon Institute of Technology. Almora, India.*

***Pijush Kanti Dutta Pramanik**, Computer Science and Engg., UTU/Bipin Chandra Tripathi Kumaon Institute of Technology. Almora, India.*

Cloud computing can also be describe as the ultimately virtualized system and a natural evolution for data centers which offer automated systems management.

The paper concentrates on the selection of resources, that is, to select a data center for creating VM to submit the task and several other data centers for accessing replicas required by the task. The method for the problem adopted here is: firstly, to find a set of data centers for the task to access all the replicas required, and then to find an appropriate data center among them for creating a VM to execute the task. Here we select the one who has the minimum transfer time from other data centers in the set of all. Our aim is to reduce data transfer times and access cost of data by selecting an appropriate set of data centers. In this paper, we propose a host selection methodology using WSCP coupled with Max-Min Heuristic for finding a set of data centers such that every data center selected contains replicas required as more as possible for reducing transfer times, while the total access cost of these replicas is as low as possible. That is to select data resources with lowest average access cost of replicas.

The rest of the paper is organized as follows: Section II introduces previous work in replica selection strategy in data-intensive environment. Section III introduces the Heuristic based Scheduling. Section IV introduces the evaluation model used in this research. Section V having the WSCP Based Max-Min heuristic details. Section VI having the simulation and result details. Finally, the paper was concluded in section VII.

II. RELATED WORK

In this section, we present some background knowledge and literature review on host selection method. The problem of resource selection in distributed environment has received lots of attention in recent years. In many previous works, resource selection refers to the selection of computational resource in grid environment. In [4], the authors presented a resource selection model using decision theory for selecting the best machine to run a task. This paper presents a resource selection model using decision theory for combining these influential factors in the resource selection process. This model deploying distributed and parallel processing for job execution prediction. They have presented appropriate functional behaviors and positive performance results.

In [5], they proposed an algorithm for resource selection problem of computational grids, based on the resource-availability prediction using frequent workload patterns. Resource selection is an important issue of grid computing. However, most of the proposed methods are not effective enough to resolve the problem of resource selection in grids. The reason behind is that these methods usually make use of current workload state or short-term prediction in available CPU time to be the basis of resource selection while most of grid jobs require a long execution time.

Recently, with the rapid development of data intensive computing, many researchers turned their attention to resource selection of data-intensive environment, such as data grid [6]. In data-intensive environment, besides computational resources, resources to be selected include data resources selection, which is equivalent to replica selection in data grid.

In [8], the author proposed Economy-Based File Replication Strategy for a Data Grid. It used an auction protocol to select the cheapest replica of a data set by a job running on computing element, which is lack consideration of the selection of computational resource.

In [9], the author proposed the datacenter selection based on number of PE available in to the host. So that it selects that host which has maximum no of free PE. It doesn't consider delay of transferring data.

In this paper, Weighted Set Covering Problem (WSCP) based on Max-Min heuristic is proposed. For the model, author applies a WSCP based on Max- Min to produce an approximately optimal resource set for each task. The result shows that WSCP based on Max-Min heuristic can produce an approximately optimal solution in most cases to meet both execution efficiency and economic demands simultaneously, compared to other strategies large-scale.

III. HEURISTIC BASED SCHEDULING

There are various scheduling techniques, but here we discuss only two of them.

A. MIN-MIN

Min-Min begins with the set MT (MetaTask) of all unassigned tasks and has two stages. In the first stage, the set of minimum expected completion time for each task in MT is found. In the second stage, the

task with the overall minimum expected completion time from MT is chosen and assigned to the corresponding machine. Then this task is removed from MT and the process is repeated until all tasks in the MT are mapped. However, the Min- Min Algorithm first finishes the shorter tasks and then executes the long task.

B. MAX-MIN

Max-Min is almost same as the min-min algorithm except the following: in this after finding out the completion time, the minimum execution times are found out for each and every task. Then among these minimum times the maximum value is selected which is the maximum time among all the tasks on any resources. Then that task or jobs is scheduled on the resource on which it takes the minimum time and the available time of that resource is updated for all the other tasks. The updating is done in the same manner as for the Min-Min. All the tasks are assigned resources by this procedure.

IV. EVALUATION MODEL

A cloud computing environment can be considered as a set of P data centers $D = \{d_1, d_2, ..., d_M\}$, which are connected by high speed Internet. For an application made up of a set of N independent tasks or jobs $J = \{j_1, j_2, .., j_N\}$,(N>>M),each job $j \in J$, require a set of k data set, denoted by F^j , that are spread on a subset of D.

Consider a set of N independent tasks(jobs) submitted to a VM, which is created on data center *D*. This is shown below fig. no. 1

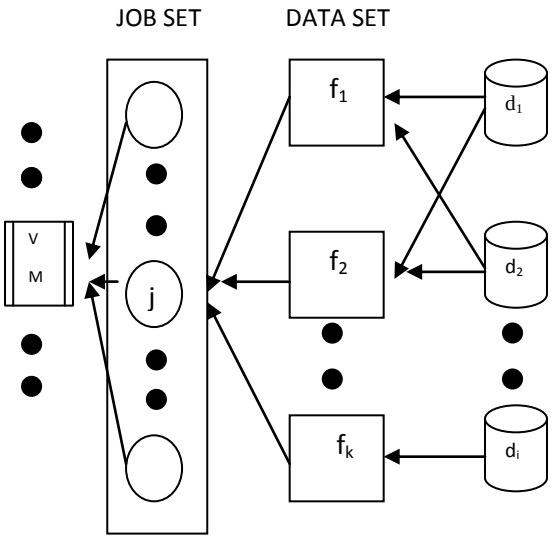

Fig1. Resource Selection Model

V. WSCP BASED MAX-MIN HEURISTIC

In cloud computing environment cloud application can be considered as a set of independent tasks (jobs), each of which require for job *j* to be submitted to a VM that is created on data center *D* , the *K* datasets required, denoted by F_j , are spread on *m* different data centers at different costs, It can be translated into a form of adjacency matrix $A = [a_{ik}]$, $1 \le i \le P$, $1 \le k \le K$ wherein $a_{ik} = w_{ik}$ ($w_{ik} > 0$) if VM can access dataset f_k from data center d_i at a cost of w_{ik} , which is abstracted as weight of f_k ; and otherwise $a_{ik} = 0$, that is d_i doesn't contain f_k . The rows that contain a w_{ik} in a particular column are said to "cover" the column at cost of w_{ik}. The problem of finding an optimal data centers set such that each can "cover" replicas as more as possible, and the total access cost of replicas is as "cheap" as possible, can be considered as the problem of finding a set of data center, each of which has lowest average access cost of replicas. This problem is equivalent to the problem of finding an optimal set of rows to cover all the columns with the lowest average weight representing access cost. While the mapping heuristic finds a resource set for a single job, the overall objective is to minimize the total makespan, the total time from the start of the scheduling to the completion of the last job, of the application consisting of *N* such dataintensive jobs. At the end, we apply the well-known Max-Min, proposed by Maheswaran et al. [11], for dynamic scheduling of jobs on heterogeneous computing resources. Our whole algorithm is shown below.

ALGORITHM1. WSCP BASED HEURISTIC

Begin Main

1. For a task *j*, create the adjacency matrix *A* with data centers forming the rows and datasets forming the columns
2. Initial solution set $B = \Phi$, $E = \Phi$, $L = \Phi$ and $z = \infty$; a data center $d = NULL$
3. ***Search(L, T, B, E, z))***
4. $S^j \leftarrow \{\{r\}, L\}$ where $r \in R$ such that S^j produces ***MCT (B)***

End Main

Search *(L, T, B, E, z)*

5. Find the minimum k, such that $f_k \notin E$. Let T_k be the block of rows in *T* corresponding to f_k.. set a pointer q to the of T_k
6. **While** q doesn't reach the end of T_k **do**

7. $F_T \leftarrow \{ f_i | t_{qi}=1, 1 \le i \le k \}$
8. $B \leftarrow B \cup \{d^k_q\}$, $E \leftarrow E \cup F_T$
9. **if** $E=F^j$
10. **if** z > MCT(*B*) **then**
11. $L \leftarrow B$, $z \leftarrow MCT(B)$
12. **Else Search** ***(L,T, B, E, z)***
13. $B \leftarrow B-\{d^k_q\}$, $E \leftarrow E- F_T$
14. Increment *q*

    End

    MCT (B)

15. Find $r \in R$ such that the completion time is minimum for the resource set $S^j=\{\{r\},R\}$ and return value

---

ALGORITHM2. WSCP BASED MAX-MIN HEURISTIC

Begin Main

1. **Repeat**
2. **foreach** $j \in J_u$ **do**
3. Find the resource set by WSCP that achieve the MCT for j
4. **end**
5. Find the job $j \in J_u$ with maximum value of $T_{ct}(j)$
6. Assign *j* to its selected resource set and remove *j* from j_u
7. Update the resource availability based on the allocation performed in the previous step
8. **Until** j_u is empty

End Main

VI. SIMULATION AND RESULT

CloudSim leaded by Buyya, allows cloud customers to test their services in repeatable and controllable environment free of cost, and to turn the performance bottlenecks before deploying on real clouds. It can provide a generalized and extensible simulation framework that enables modeling, simulation and experimentation of emerging cloud computing infrastructures and application services. It is designed for studying various resource management approaches and scheduling algorithms in cloud environment. The CloudSim toolkit supports both system and behavior modeling of Cloud system components such as data centers, virtual machines (VMs) and provisioning policies of resource. It implements generic application provisioning techniques that can be extended with ease and limited efforts In CloudSim, users is modeled by a DatacenterBroker, which is responsible for mediating between users and service providers depending on user's tasks across Clouds.

In our experiments, we have used CloudSim as a simulator for checking the performance of our improved algorithm. We have considered Virtual Machines as resource and Cloudlets as tasks/jobs. We have checked the performance of the algorithm by fixed the number of virtual machines and varied the number of cloudlets. The makespans that the algorithms produce are shown in fig no. 2 we have fixed the number of virtual machines as 20 and we are varying the number of cloudlets from 20 to 120 with a difference of 20.

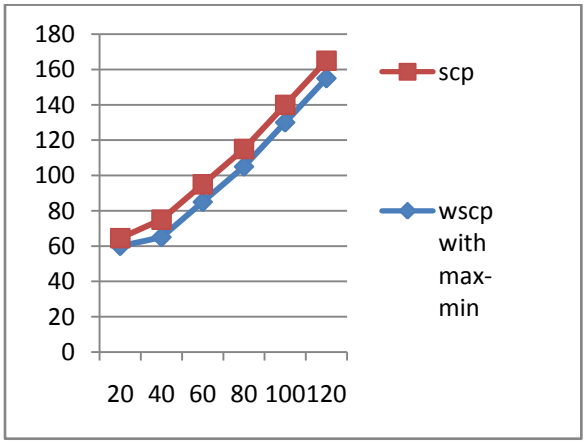

Fig2. Graph for Makespans

VII. CONCLUSION

We have designed and tested an algorithm which is made by WSCP coupled with Max-Min Heuristic. The main goal of it, to select the host and to schedule multiple jobs on multiple machines in an efficient manner such that the jobs take the minimum time for the completion.

VIII. REFERENCES

[1] I. Foster, Y Zhao, I. Raicu, and S. Lu, "Cloud Computing and Grid Computing 360-degreecompared[C]", in *Grid Computing Environments Workshop*, 2008, pp. 1-10.

[2] Daniel Nurmi, Rich Wolski, Chris Grzegorczyk, Graziano Obertelli, Sunil Soman,Lamia Youseff, Dmitrii Zagorodnov, "The Eucalyptus Open-source Cloudcomputing System ", *2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid*, CCGRID 2009, pp: 124-131.

[3] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy H. Katz, Andrew Konwinski, Gunho Lee, David A. Patterson, Ariel Rabkin, Ion Stoica, Matei Zaharia, "Above the Clouds: A Berkeley View of Cloud Computing", *Technical Report No. UCB/EECS-2009-28*, 2009.

[4] Lilian Noronha Nassif, José Marcos Nogueira, Flávio Vinícius de Andrade, "Distributed Resource Selection in Grid Using Decision Theory", in *Seventh IEEE International Symposium on Cluster Computing and the Grid*(*CCGrid 2007*,pp:97-102).

[5] Tyng-Yeu Liang Siou-Ying Wang I-Han Wu, "Using Frequent Workload Patterns in Resource Selection for Grid Jobs", *DOI 10.1109/APSCC.*2008.217.

[6] D.G. Cameron, R. Carvajal-Schiaffino, A.P. Millar, C. Nicholson, K.Stockinger, F. Zini, Evaluating scheduling and replica optimisation strategies in OptorSim, in:*Proceedings of the Fourth International Workshop on Grid Computing (Grid2003)*, IEEE CS Press, Los Alamitos,CA, USA, Phoenix, AZ, USA, 2003.

[7] Rajkumar Buyya, Rajiv Ranjan, Rodrigo N. Calheiros, "Modeling and Simulation of Scalable Cloud Computing Environments and the CloudSim Toolkit: Challenges and Opportunities", in *The 2009 International Conference on High Performance Computing and Simulation, HPCS 2009*, pp:1-11.

[8] Juefu Liu, Peng Liu, "Status and Key Techniques in Cloud Computing", in *Proceedings of 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)* , pp: V4-285–V4-288.

[9] H. Baghban and M. Rahmani, "A Heuristic on Job Scheduling in Grid Computing Environment", in *Proc. 7th Inter. Conf. on Grid and Cooperative Computing (GCC'08),* 2008, pp. 141-146.

[10] Huang Q.Y., Huang T.L., "An Optimistic Job Scheduling Strategy based on QoS for Cloud Computing", IEEE International Conference on Intelligent Computing and Integrated Systems (ICISS), 2010, Guilin, pp. 673-675, 2010.

[11] M. Maheswaran, S. Ali, H. J. Siegel, D. Hensgen and R. F. Freund, "Dynamic Matching and Scheduling of a Class of Independent tasks onto Heterogeneous Computing Systems," *Journal of Parallel and Distributed Computing*, Vol. 59, No. 2, pp. 107-131,1999.

[12] E. Munir, J. Li , S. Shi and Q. Rasool, " Performance Analysis of Task Scheduling Heuristics in Grid", in *Proc. 6th Inter. Conf. on Machine Learning and Cybernetics*, 2007, pp. 3093-3098.