Comparative Performance Analysis of Task Clustering Methods in Cloud Computing

Dr. Vinay Chavan[1], Dr. Kishor Dhole[2], Parag Ravikant Kaveri[2]

[1](drvinaychavan@yahoo.co.in, S. K. Porwal College, Kamptee Nagpur, India)
[2](km_phd108@yahoo.co.in, S. K. Porwal College, Kamptee Nagpur, India)
2(parag_kaveri@yahoo.co.in, P.G D C.S.T, HVPM Amravati, India)

Abstract : *The Workflow Scheduling directs in the area of workflow management systems. The efficient Scheduling in workflows helps the system to increase cost, which can be evaluation by considering various clustering methods, which is been studied in this paper. A number of parameters are there on the basis of which performance of cost is dependent and is evaluated and compared to find the efficient of workflow scheduling algorithm also. The key consideration in the paper determines the openness of various task clustering methods in cloud environment and based on detail experimentation results are drawn to evaluate the importance of methods suited with scheduling policy.*

Keywords – *Cloudsim, Clustering, Resource Scheduling , Task Scheduling, WorkflowSim*

I. INTRODUCTION

Cloud computing is a large pool of resources which interact with each other to exchange information in the form of tasks. The resources and tasks works together to exchange information from one application to another for delivering services to end user form different service provider. Task consist of instructions and data delivered in the form of services, although the most popular services provided by cloud service provider includes SaaS(Software as a Service), PaaS(Platform as a Service) and IaaS(Infrastructure as a Service) [1]. The scheduling of resources and task is also an important consideration in terms of performance point of view, which is also discussed in the paper. The main focus of the paper is to explain the functionality of different task clustering method and based on the task cluster evaluating cost vice performance in workflow based cloud system[2].

II. SCHEDULING METHODS

There are different scheduling methods explained in context of different factors in cloud computing. In this section, variety of scheduling technique reviewed briefly and broadly classified in two types, task scheduling and resource scheduling both work in support of load balancing[2],[3].
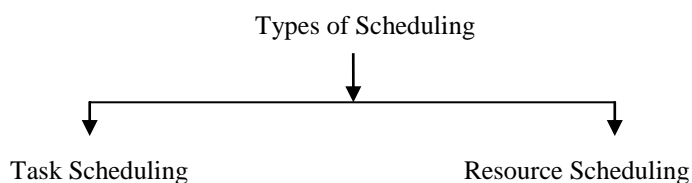

Figure 1 Types of Scheduling

- **Task Scheduling**

Task scheduling is process of assigning tasks to available resources in static or dynamic manner in real time although it depends on task scheduling algorithm that how tasks get assignee.

- **Resource Scheduling**

Cloud Computing use computing resources (hardware and software) that delivered as services over network. Cloud computing entrusts remote services with a user's data, software and computation. In the business model, using software as a service, users are provided access to application software and databases. The cloud providers manage the infrastructure and platforms on which the applications run. SaaS is sometimes referred to as "on-demand software" and is usually priced on a pay-per-use basis.

Table I. Scheduling Methods

Type	Sub-Category	Issues Resolved	Provider Oriented	Customer Oriented
Resource Allocation	At Host level At VM level	Efficient Utilization Minimize Makespan Ensure Availability	Yes	Yes
Task Scheduling	Space-Sharing Time-Sharing	Minimize overall response Time	Yes	Yes

III. CLUSTERING METHODS

Task clustering is basically used to combine small sized tasks of the workflow into a comparatively large sized task to minimize makespan of the workflow by reducing the impact of queue wait time [4],[5]. Thus task clustering restructure workflow by combining the small sized tasks into a cluster and execute this cluster as a single task. Due to clustering number of task in the workflow is reduced and there is less queue wait time as compared to small sized tasks. Task clustering are categorized as level and label based clustering, vertical, blocked and Horizontal clustering.

Simple Workflow   Horizontal Clustering

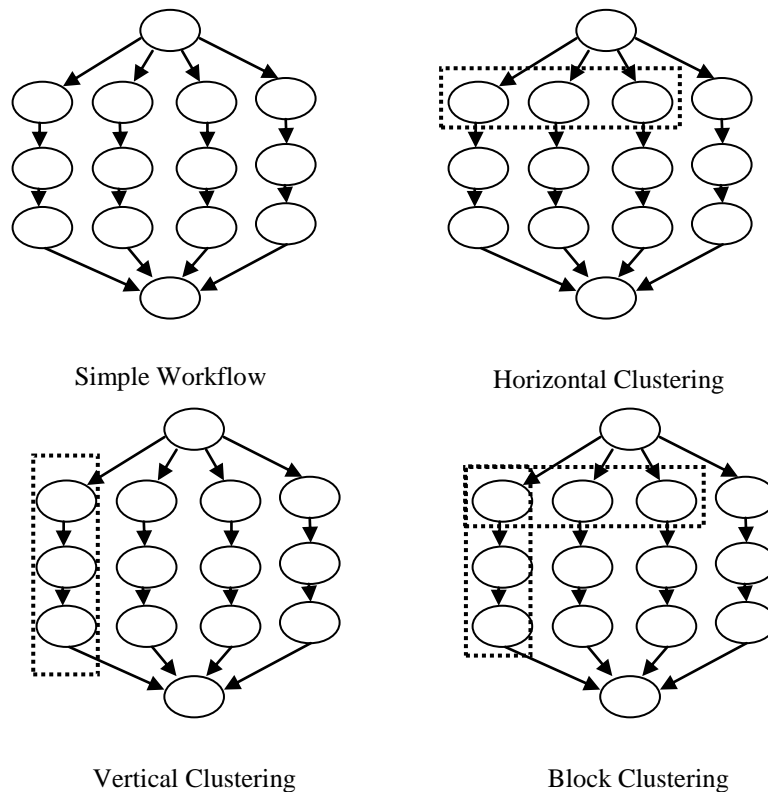Vertical Clustering   Block Clustering

Figure 3.1 Types of Scheduling

- **Horizontal Clustering:**
Level-based clustering is also called horizontal clustering. In this type of clustering, the tasks are independent and the tasks of the same level of the workflow combined together. In vertical clustering, tasks of the same pipeline can be combined together.

- **Block Clustering:**
Blocked clustering represents the combination of both horizontal and vertical clustering in the same workflow.

- **Vertical Clustering:**
Vertical clustering, in which tasks in the same pipeline can be combined together to form a cluster. Each cluster of vertical clustering contains three tasks.
Tasks in the same level of workflow can have different execution time and whenever these tasks are combined without the consideration of their runtime variance then it causes the problem of load imbalance, i.e., some cluster may contain the smaller tasks and other may have larger tasks.

IV.   EXPERIMENTATION

The proposed work is executed on a cloud simulation tool WorkflowSim. WorkflowSim extends the functionality of CloudSim simulator by adding a higher layer. This simulation supports in the distributed environment and techniques for implementing cloud with open access programming for modeling and simulating cloud computing infrastructures and services supports for the execution of workloads. The proposed work is based on different datasets known as workflows which are already available with WorkflowSim and workflows are generated using workflow generator. The experimental evaluation of the significance research problem is based on execution of a workflow. The parameters considered are time, total cost, VM deployed total number of tasks and jobs etc. The work is executed by taking montage[6] data set with 1000 tasks. These workflows worked as tasks submitted to 10 Virtual machines respectively and are deployed on single data centre the configuration of virtual machine is 1024 MB of RAM, 10000 MB of image size and 2 processing element (C.P.U). Host with High configuration is deployed in data centre to provide ample space to create these virtual machines. Min-Min scheduling policy is defined for scheduling tasks in the desired experimentation. The policy works in manner to find the virtual machine and resources which can execute the tasks by taking minimum time.

V.   RESULT AND ANALYSIS

The results of the experiments are tabulated in Table II which shows the performance of clustering method based on cost and time parameter. Here the experiment is executed with 1000 tasks in a montage dataset which support horizontal clustering as maximum number of task in workflow are at horizontal level. The number of job created in horizontal clustering is less and very less in block clustering because it considered both horizontal and vertical level tasks. Vertical cluster doesn't worked well here due to the montage data set. Block clustering shows good result in time compared to other clustering methods. The maximum time taken to complete same number of tasks is with without clustering. Block clustering also show good results by minimizing the processing cost to complete the same number of tasks.

Table II. Experimentation Results

Clustering Type	Number of Jobs Created	Total Time (ms)	Total Cost ($)
NONE (Without Clustering)	1000	596770.77	35984.09
Horizontal Clustering	21	37454.05	35460.5
Vertical Clustering	996	591030.09	35983.88
Block Clustering	17	27993.79	35342.1

VI.   CONCLUSION

Conclusion drawn from results points to introducing clustering methods for higher performances in cloud environment. Here the performance is majored by considering two important parameters in server oriented cloud architecture which are cost and time. cost plays an important role in user oriented cloud model as by paying minimum cost, users get profitable service. The results indicates to cluster tasks into jobs at different level to minimize cost in cloud environment. The second import parameter considered was time which is quality for service parameter. Time is minimized after clustering task into jobs for different clustering methods. High improvements are noticed after block clustering. The clustering method shows results based on the type of clustering and the ability of tasks to get clustered.

REFERENCES

Books:

[1]. Rajkumar Buyyaames BrobergAndrzej Goscinski "Cloud Computing Principles and Paradigms" wiley publication 2011 by John Wiley & Sons.

Journal Papers:

[2]. Santhosh B, Raghavendra Naik, Balkrishna Yende, Dr D.H Manjaiah, "Comparative Study of Scheduling and Service Broker Algorithms in Cloud Computing, International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 5, October 2014 ISSN(Online): 2320-9801

[3]. Komal Sood, "Comparative Study of Scheduling Mechanisms In Cloud Computing", IOSR Journal of Engineering (IOSRJEN), Vol. 04, Issue 05 (May. 2014), ||V7|| PP 30-33, ISSN (e): 2250-3021, ISSN (p): 2278-8719

[4]. Weiwei Chen, Rafael Ferreira da Silva, Ewa Deelman, Rizos Sakellariou, "Balanced Task Clustering in Scientific Workflows"

[5]. Santhosh B, Harshitha, Prachi Kaneria A, Dr D.H Manjaiah" "Comparative Study of Workflow Scheduling Algorithms in Cloud Computing. , International Journal of Innovative Research in Computer and Communication Engineering ISSN(Online): 2320-9801

[6]. Gideon Juve, Ann Chervenak, Ewa Deelman, Shishir Bharathi, Gaurang Mehta, Karan Vahi "Characterizing and Profiling Scientific Workflows".