COST BASED SERVER SELECTION USING CLUSTERING APPROACH

G.Kiruthiga[1], Dr. S. Mary Vennila[2]

[1]Research Scholar, PG and Research Dept. of Computer Science, Presidency College, Chennai.

[2]Associate Professor & Head, PG and Research Dept. of Computer Science, Presidency College, Chennai.

Email: g_kiruthiga@yahoo.co.in[1], vennilarhymend@yahoo.co.in[2]

Abstract

Cloud Computing deals with the benefits of business and the end users. Basic idea is to transmit all kinds of resources through the Internet such as storage resources, computing resources, bandwidth and so forth. Users do not need to purchase a large of computing systems to manage their business. Services are sold on a subscription or pay-per usage basis over internet.

Proposed a cluster oriented selection strategy which accounts the task of implementation of the relationship between time and cost. The server allocation method is an effective process when compare to the previous work. Comparing with the user's the historical cost, minimum cost, and maximum cost will be stabilized in our proposed system. The results will show that our strategy completes the task at the user's expectation of the completion time. Reduces the resource usage costs of the task and effectively enhances the scientific workflow implementation efficiency.

Keywords:-*clustering Techniques, Server selection, Cloud server, Cost reduction*

1. Introduction

Cloud Computing deals with the benefits of business and the end users. By using three types of services, only the cloud computing can be able to store the data's. These are also called as cloud computing stack because they are arranged one after the other. These services are slightly difficult to use, based on the provider, it might be easy to use in a friendly manner in some browser-based dashboard which will be easy for the IT professionals and the software developer to organize their accounts in the cloud.Basic idea is to transmit all kinds of resources through the Internet such as storage resources, computing resources, bandwidth and so forth. Users do not need to purchase a large of computing systems to manage their business. Services are sold on a subscription or pay-per usage basis over internet. Need to pay for the resources according to their needs in order to decrease the cost greatly [1].

Cloud computing has a variety of attention to large scale distributed approach. The performance of the Ali cloud with different server can provide a cost efficient by mapping the relationship of the different server. The main contribution is to implement the relationship between the time and cost; in this case the server allocation task will be effective. The scientific workflow will reduces the usage of the resources cost for the task and there will satisfy the user expectation of the task completion time. And it is mainly focuses on the business workflow in the particular time the user will fulfill their requirement regarding the task completion in the required time. It must improve in the application using the API for quick convenient configuration [2].

The main workflow of this paper is to select a server to improve the computation time and also the scheduler strategy of the server by task scheduling approach. The part of this paper is category into Section II Discussion about the Literature Survey, Section III Explore of the Proposed Work, Section IV Analysis the Experimental Result and Section V Presents the Conclusion.

2. Literature Survey

C. Ding, et al, 2012 presented a Cloud GPS: A scalable and ISP friendly server selection on cloud, there deals with the explosively increasing in the number of cloud and with the client based on the DEM component, which takes the two different kinds of the NC that is used in the cost reduction O(n) by positioning the cloud and the user. The cloud GPS are user friendly to the ISP that reduce the transmit of the traffic that leads to the low ISP operational cost and also improve the end user quality of service, which is based on MM component, that makes a closer sever selection and theinter domain transit traffic with the restriction of the server capacity.

Ao-Jan Su,et al, obtained a Drafting behind the Akamai which deals with the networking that works under the conditionof less expensive cost of the network measurement, it determine how the one can be infer and utilize the quality,along within the short time scale information. It posses five types of working there are 1. The Akamai server correlate with the network between the client and the server here it takes the

70% of the paths; in this the 10 % will be taken as a network path 2. According to the inter-redirection frequency only the given client predominantly depends and thus it responsible by the low –level Akamai DNS server. 3. There achieves both the random or round robin path for South America selection clients to get the low redirection frequencies. 4. There are heterogeneous to the edge server in the Akamai customers in this there investigate the client a large number of servers. 5. The web content distributed the user load balancing server and the server farms in small data centers which are all utilized by the CDN service for network measurement and the global server deployment. The issue here is the web performance by delivering the content to the end user from the geographical server.

Hong Xu, discussed about the geo-distributed cloud service it contribute to perform the diversity and performances well with the cost which are modeled by efficiently designing the ADMM algorithm which are used in the large scale global problems into many sub problems which are solved quickly it will be suitable for the cloud environment with the many number of the server resources it will evaluated by the trace driven algorithm which are been simulated in this algorithm for the performance metric. The algorithm will be design on the divide and the conquer strategies. It ought to improve this in the traffic scenes.

Zhangjun Wu,et al studied about the hierarchical scheduling strategy, It make possible for distribution in e-business and the e-science. The important aspect of the cloud workflow from the other counterpart is a business model in the marketing here the user can change according to the consumption such as computing, storage and the network. It affords to the two market based strategy 1. It must be best based scheduling with the QOS 2. So it must be first required for the QOS in the cloud workflow system instead of the conventional strategy. The overall cloud workflow can deals with services whether there are within or outside their own data center and also the ability to handle to the workflow applications. The scheduling strategy must be improved.

Yang Zhang, et al, presented about the fault Tolerance and Scheduling Techniques for computational Grids, which slot in the balance approach towards the reliability and performance that are not very sensitive to the underlying resource prediction reliability. This proposes the new algorithm for the replication for the whole DAG (WDO) onto the several clusters. One of the widely used fault tolerance techniques for the grid means the application is resubmitted on the resource in the case of the failure. And it is demonstrated that the fault tolerance techniques will be very effective > and the reliability of the execution will be of 200% which will not affects the performance more than 10 %. The drawback is that there are degradable in the performance when there is several of resource availability.

3. Proposed System

In this section, we have explained the working model of our proposed model named, "Priority based clustering" which ensures a better QOS constraint to select efficient servers from pool of servers. The proposed approach composed of segments which are explained as follows:

3.1 Priority based Task Scheduling

The cloud computing it innovatively deals with the programming works, in this one of the trendy analysis is the Priority Based Scheduling shown in Fig.1. In this task scheduling is classified into three levels, the first one is the Objective level where the Scheduling (Goal) will assign the number of resource that are listed. Second is the Feature level and finally the Substitute Level where the number of task or jobs are been assigned for the Resources shown in the Fig.2. According to the priority scheduling, the incoming task prioritized based on memory. If the incoming task memory is greater, when compare to the current task running in the process means it will saves the current task and update the incoming task and executes. And the higher priority will be allocated and run on the cloud. If the upcoming task with the lower priority means it is been queued on the buffer. Let T= {T1, T2, T3…. Tn} it represents n number of task, and R= {R1, R2, R3….Rm} resembles the Resource allocation the input in cloud atmosphere, is m<<n. The priority of each task will be compared independently.

In this scheduling strategy, there prioritize the task using the certain condition. Thus the make span is important while scheduling where the total time of the task are scheduled and completed in cloud. If it is lesser the makespan, the service and the quality of the scheduling are better. It improves the computational time and efficiency of the resources. In order to derive this the equation (1) is been used where, τi represents the complete time taken for the task ti , N : Total number of task, M : number of resources available and i: represents the task scheduler. Thus the quality will be maintained from this process.

$$M = \frac{\min \sum_{i=1}^{N} * \tau i}{N} (1)$$

Task scheduling is very significant approach in the cloud computing, and it assigns the task to the resources at the particular time interval.

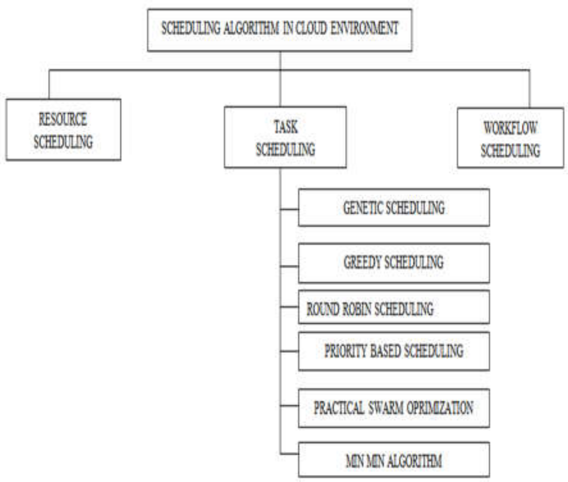

Fig.1. Priority based Scheduling in cloud

The main goal of scheduling is to maximize the resource utilization, minimizing the waiting time.

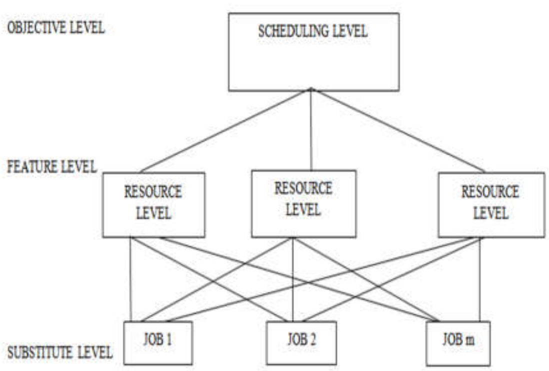

Fig.2. Scheduling classification in cloud

When the performance is good then the system is said to be a superior scheduler. There are two types of scheduling Static Scheduling and Dynamic Scheduling. By doing this distribution of task to the resources, will appears in the long distance from one resource to another resource. In this aspect the performance degrades.

3.2 Propsed Algorithm

Proposed Scheduling Algorithm

1. All the service request jobs are gathered in the cloud server.

2. After the request are submitted they are rearrange the request in sorted (ascendi ng) order based on the length.

3. Then make in to groups (for example 5 jobs) the request according to the availability of resources.

4. Find the Standard deviation(S) for all the request is done according to their respective lengths.

5. Calculate the Mid value for total number of available lengths. Allocate the resources to the request in the following method.

5.1. If S is greater than m, then mid of the total numbers of request in the groups with the longer length will be allocated first to the resources in descending order.

5.2 Else the request in groups with the minimum length will be allocated first to the resources with the resource in ascending order.

6. All the resources will be deleted at end.

In this enhanced algorithm compare with the Shortest Job First algorithm. This algorithm by which request will be executed according to their length and to the single resource for the execution. Standard deviation is calculated using the following one,

$$\text{Standard Deviation(S)}=\sqrt{\sum(X-X)^2/N-1}$$

The Standard Deviation is greater than the value at mid of the total numbers of requests, then the request in groups with the longer length is allocated to the resources for execution.

Table 1:Possibility 1

Index	1	2	3	4	5	6	7	8	9	10
Job	10	20	40	80	120	240	480	960	1920	3840

From the above case we have 10 number of jobs and standard deviation is 390.4426.Based on our algorithm SD>s/2 i.e. 390.4426 > 100 so we will take the maximum possibility i.e. the job in groups with the longer length is allocated to the resources for execution.

If standard deviation is less than the value at mid of the total number of jobs, then the tasks in groups with the shorter length is allocated to the resources for execution.

Table 2:Possibility 2

Index	1	2	3	4	5	6	7	8	9	10
Job	10	20	40	50	60	70	80	90	110	130

In this possibility we have same 10 number of jobs and tandard deviation is 12.03698. Based on the algorithm SD<s/2 i.e. 12.03698 < 65. The jobs in groups with theshorter length is allocated to the resources for execution.

4. Analysis and Experimental Results

To implement the enhanced algorithm that has used the CloudSim Framework for the practical purpose. In this study, we have taken ten jobs (T1,T2…T10) and they have five resources for the same (R1,R2,…R10). They have taken two

different-different scenarios for the same number of tasks with different lengths but same resources. These are the five resources they are having for the tasks. They have evaluated the results for these scenarios on existing algorithm and our enhanced algorithm.
These are the five resources they are having for the tasks. They have evaluated the results for these scenarios on existing algorithm and our proposed algorithm.

They have the following problem set in which two different problem set are mentioned with different-2 task lengths.

Table 3:Different length of the Job

Problem	Number of jobs	Length of the job
P1	T1	3
	T2	80
	T2	90
	T4	4
	T5	50
	T6	343
	T7	300
	T8	9930
	T9	3000
	T10	10
P2	T1	4500
	T2	3000
	T2	7500
	T4	12000
	T5	10000
	T6	14000
	T7	30000
	T8	18000
	T9	20000
	T10	2000

These are the two problem sets with different-2 task lengths. In both the task sets that has to be grouping of two-two tasks i.e. each resource will process only two tasks. In the P1 i.e. in the first problem set our enhanced algorithm will run the maximum possibility i.e. tasks with the maximum length will be getting executed first. In the P2 our proposed algorithm will run the Minimum possibility i.e. tasks with the minimum length will be getting executed first. When evaluated on the CloudSim framework we get the following output when compared with the existing SJF algorithm.

In the below Figure.3 we can see that our proposed algorithm is taking less processing time when compared with the existing algorithm.

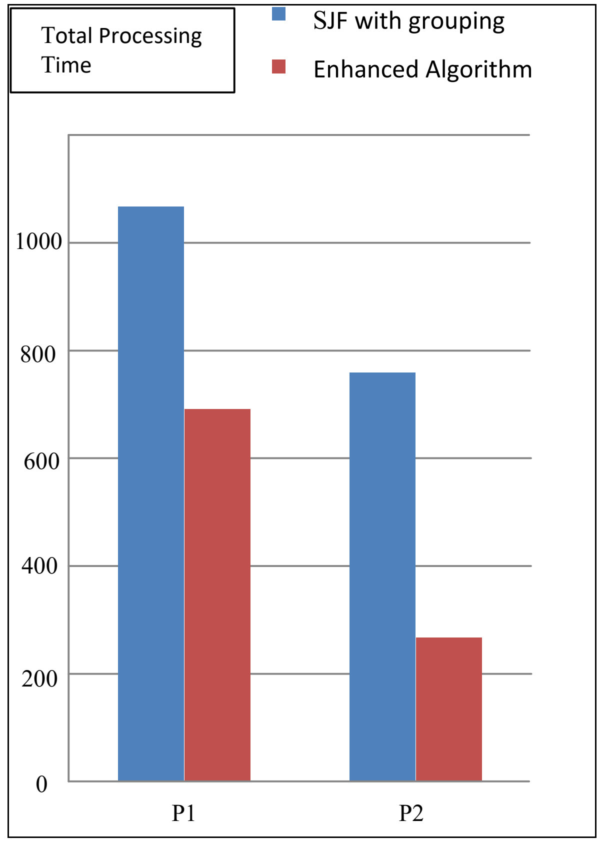

Fig3: Existing Vs Enhanced Algorithm

We have also evaluated the results with the randomly generated values of the length of the tasks. We have made the scenarios of various randomly generated values like i.e. of 10,20,30,40, 50, 60,70,80,90 and 100 tasks and compared it with the existing algorithm. In all the cases we got better results in our proposed algorithm. Here I am just putting results of three cases with i.e. 10, 50 and 100 tasks.
In the first problem set i.e. in P1 with ten tasks, we have made the groups of two-two tasks.
We have compared the existing algorithm with the proposed algorithm and got the following results.

5. Conclusion and Future Work

In this paper we have enhance an improved scheduling algorithm which is working efficiently and when evaluated using the CloudSim framework it is performing best as compared with the existing shortest job first algorithm with task grouping. Our enhanced algorithm is continuously taking less processing time i.e. time taken by the resources to complete the execution when compared with the existing algorithm. The results will show that our strategy completes the task at the user's expectation of the Processing time. Reduces the resource usage costs of the task and effectively enhances the scientific workflow implementation efficiency.

In future work can be done on this to reduce the average waiting time and processing time of the execution.

6. References

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, et al., "A view of cloud computing," Communications of the Acm,, 53, Pp. 50-58, 2010.

[2] Futian Wang, Ke Fang, Jin Tang, cheng Zhang, "A Server Selection Strategy about Cloud Workflow based on QoS Constraint", IEEE Computer Society, IEEE SERA, 2017.

[3] C. Ding, Y. Chen, T. Xu and X. Fu, "CloudGPS: A Scalable and ISPFriendly Server Selection Scheme in Cloud Computing Environments," in IEEE/ACM IWQoS, 2012.

[4] A. -J. Su, D. R. Choffnes, A. Kuzmanovic and F. E. Bustamante, "Drafting behind Akamai (Travelocity-based detouring)," ACM SIGCOMM, 2016.

[5] H. Xu and B. Li, "Joint Request Mapping and Response Routing for Geo-distributed Cloud Services," IEEE INFOCOM, 2013.

[6] Z. Wu, X. Liu, Z. Ni, D. Yuan, and Y. Yang, "A market-oriented hierarchical scheduling strategy in cloud workflow systems," The Journal of Supercomputing, 2013, 63, Pp. 1-38.

[7] Y. Zhang, A. Mandal, C. Koelbel, and K. Cooper, "Combined Fault Tolerance and Scheduling Techniques for Workflow Applications on Computational Grids,"IEEE/ACM International Symposium on CLUSTER Computing and the Grid, CCGRID 2009, Shanghai, China, 18-21, 2009, Pp. 244-251.

[8] C. Moretti, H. Bui, K. Hollingsworth, B. Rich, P. Flynn, and D. Thain, "All-Pairs: An Abstraction for Data-Intensive Computing on Campus Grids," *IEEE Transactions on Parallel & Distributed Systems,* 2010, 21, Pp. 33-46.

[9] Mohamed Nabeel, "Privacy Preserving Policy Based Content Sharing in Public Clouds", *IEEE transactions on knowledge and data engineering,*2013, 25(11), Pp. 2602-2614.

[10] Ke Liu, Hai Jin, Jinjun Chen, Xiao Liu, Dong Yuan and Yun Yang," A Compromised-Time-Cost Scheduling Algorithm in SwinDeW-C for Instance-Intensive Cost-Constrained Workflows on a Cloud Computing Platform", *International Journal of High Performance Computing Applications,*2010, 24(4), Pp. 445-456.

[11] D. D. Oliveira, E. Ogasawara, K. Ocaña, F. Baião, and M. Mattoso, "An adaptive parallel execution strategy for cloud-based scientific workflows," *Concurrency & Computation Practice & Experience, 2012,* 24, Pp. 1531-1550.

[12] F. Wang, X. Liu, and Y. Yang, "Necessary and sufficient checkpoint selection for temporal verification of high-confidence cloud workflow systems," *Sciece China Information Sciences,* 2015.