

Study of Scheduling Techniques in Cloud Computing Environment

Anagha Yadav¹, S.B.Rathod²

¹M.E Student Computer Engineering , Sinhgad Academy of Engineering
Pune, India

²Assistant Professor, Sinhgad Academy of Engineering
Pune,India

Abstract: *The Cloud Computing Technology is everywhere. With the Significant advances in Information and Communication Technology, Cloud computing technology is taking a large space in IT Industry. In Simple terms Cloud Computing can be depict as means of storing and accessing data and program over internet instead of your computer's hard drive. Cloud Computing provides on demand self-service, Resource Pooling and Unambiguous network access. Scheduling is the task where the work specified by some means is assigned to certain resources that complete the work. The resources may be CPU time, IO devices, and many more. Different Schedulers are implemented to achieve load balancing, share system resources effectively, or to achieve quality of service. Scheduling of task and resources maximizes throughput , minimizes response time, minimize latency and maximizing fairness i.e. equal CPU time and workload for each process. Scheduling can be categorized as resource scheduling, task scheduling and workload scheduling. Task scheduling is the main concern in this paper. The objective of this study is to optimize workload utilization and task scheduling in cloud computing environment. This paper outlines a comparative study that has been done to assess these Scheduling models on the cloud environment.*

Keywords: *Cloud Computing, Scheduling Models, Workload, Priority, Task Scheduling, Minimized Response Time.*

I. INTRODUCTION

Cloud computing is revolutionizing the IT industry; it not only provides robust technology solutions, but also because it facilitates a change in the role of the IT organization itself. The Cloud Computing technology has changed the way the IT is managed. Cloud provides collaborated, real time views and automated services to the IT business. The general benefits of cloud are disaster recovery, automatic software updates, increased collaboration, and work from anywhere. Cloud computing also has many benefits in E-Governance sectors. The definition by the National Institute of Standards and Technology (NIST) "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage,

applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

Cloud computing is highly dynamic, therefore scheduling tasks and resources is little challenging. Scheduling can be described as the set of policies to control the order of work to be performed by a computer system. The main advantage of scheduling algorithm is to achieve a high performance computing and high system throughput. Scheduling manages availability of CPU memory and gives maximum utilization of resource. Scheduling of task is based on various parameters. Scheduling problem is NP Hard. The objective of the scheduling algorithms in cloud computing environment is to utilize the resources properly while managing the load between the resources so that to get the minimum execution time.

Scheduling process in cloud is divided into three stages [6] as described below:

- Resource discovering and filtering- Datacenter Broker discovers the resources present in the network system and collects status information about the resources.
- Resource selection - Target resource is selected based on certain requirements of task and resource. This is deciding stage.
- Task allocation -Task is allocated to resource selected.

Datacenters are the essential parts of cloud computing environment. In a single datacenter generally hundreds and thousands of virtual machines run at any instance of time, hosting many tasks and simultaneously the cloud system keeps receiving the batches of task requests. During this process, few target machines out many powered on machines, can fulfil a batch of incoming tasks. So Task scheduling is a relevant issue which is greatly influence the performance of cloud service provider. One of the incitation of scheduling is for load-balancing of the tasks and to ensuring availability of resources. Task Scheduling in cloud computing can be performed by various algorithms. The task scheduling problem can be viewed as the finding an optimal mapping/assignment of set of subtasks of

different tasks over the available set of resources (processors/computer machines) so that we can achieve the desired goals for tasks [5]. First Come First Serve, Round Robin, Minimum Execution Time, Minimum Completion Time, Genetic Algorithm, Switching Algorithm, Priority based Job Scheduling Algorithm are the few algorithm [2].

One approach for scheduling is to analyse different properties of virtual machines by means of assigned attributes, and then scheduling them while respecting conflicts expressed over those attributes. The Major factors of scheduling are CPU Utilization time, Computational Complexity (task length, processing power), response time and waiting time. A generalized attribute-based constraint specification framework for virtual resource to physical resource scheduling is for IaaS clouds. Cloud task scheduling is a NP complete problem [2]. There are number of users that share the cloud resources by submitting their computing task on the centralized cloud system. Scheduling these tasks is a challenge to cloud computing environment. Different scheduling strategies are proposed in [2] and [3]. These strategies consider different parameters like cost, response time, quality of Service (QoS) and workloads. Optimal resource allocation or task scheduling in the cloud computing environment should decide optimal number of systems required in the cloud so that the total cost is minimized. Cloud computing is highly dynamic, and therefore, resource allocation problems should to be continuously directed, as servers become available or non-available Thus this study focuses on scheduling algorithms in cloud computing environment considering above mentioned parameters, and strategies.

Need of Scheduling:

1. Quality of Service: Quality of services is important in cloud environment. The cloud is mainly to provide users with computing and cloud storage services, resource demand for users and resources supplied by provider to the users in such a way so that quality of service can be achieved. When job scheduling management comes to job allocation, it is necessary to guarantees about QoS of resources.
2. Load Balancing: Task scheduling algorithm can maintain load balancing. So load balancing becomes another major important parameter in cloud.
3. The throughput of the system: Throughput is measure of system task scheduling optimizing performance, and it is also a target which has to be considered in business model development. Increase throughput for users and cloud providers would be benefit for them both

4. The best running time: Tasks can be divided into different categories according to the needs of users, and then set the best running time on the basis of different goals for each task. It will improve the QoS of task scheduling indirectly in a cloud environment [14].

II RELATED WORK

This section, describe the related work of task scheduling in cloud computing environment. Cost of each task in cloud resources is different with one another. A task cannot be executed on two or more processors simultaneously. Therefore there is need to schedule these tasks effectively. Scheduling of user tasks in cloud computing is important. A good scheduling technique is one that helps in efficient utilization of resources. Proper scheduling can make all tasks meet their deadlines. Task scheduling can be assessed in terms of having smaller Makespan, cost, etc. The performance and efficiency of cloud computing services always depends upon the performance of the user tasks submitted to the cloud system. Scheduling of the user tasks plays significant role in improving performance of the cloud services. Task scheduling is one of the main types of scheduling performed [2]. Task scheduling can be categorized as distributed and centralized. Distributed scheduling can be further classified as heuristic and hybrid algorithm.

a. Heuristic Task Scheduling Approaches

Heuristic Task Scheduling are of two types Static and Dynamic. First we will discuss static method of scheduling.

1) Static Scheduling

In Static Scheduling all tasks are present at the same time. Static Scheduling includes scheduling strategies like First Come First Serve, Round Robin methods.

FCFS method initially keeps all the tasks in a queue. When the resources are free and available they are allocated to the tasks. It organises the queue, where the head of the queue is processed first. The task leaves the queue in the order in which they arrive. The policy implemented by the VM provisioned is a straightforward policy that allocates a VM to the Host in First-Come-First-Serve (FCFS) basis [4].

Advantages of FIFO:

- Fewer Complexes in nature.
- Simple to understand and easy to design.
- Scheduling overhead can be minimal.
- Throughput is low, since long processes can hold the CPU
- Turnaround time, waiting time and response time is significantly high.

- It is based on Queuing

Round Robin method works in FIFO manner but the scheduler assigns a fixed time slots for each task and then cycles through them one by one. Time slices or time quantum is assigned to each task in equal portions. Round-robin scheduling is simple and easy to implement. A round-robin scheduler gains the advantage of time-sharing. It gives each task a particular time slot or time quantum and if the task is not able to complete within that time quantum it interrupts the task. The task is kept in the waiting queue and latter the task is allotted with the next time a time slot and then executes the remaining portion of the task.

Advantages of Round Robin:

- Round Robin (RR) algorithm focuses on the fairness.
- Little Complex as compared to FCFS.
- RR scheduling involves comprising overhead, usually with the small time unit.
- Average response time.
- Waiting time is dependent on number of tasks in the system.
- Order of time unit allocation is based upon task arrival time, similar to FCFS.

Opportunistic load balancing is another heuristic method of scheduling. In this method it tries to schedule the tasks to the next available machines based on their expected completion time. Minimum Execution Time and Minimum Completion Time are other two heuristic strategies. MET assigns tasks on the machines based on which machine it takes less execution time. It selects the best machine for execution but do not consider the availability of resources at the time of scheduling so load imbalance can occur. Minimum Completion Time Algorithm selects machines for scheduling the tasks based on the expected minimum completion time of tasks among all the machines available [2].

Min-Min and Max-Min are two other heuristic methods for task scheduling. Min-min heuristic selects the smallest task first from all the available tasks and assigns it to a machine which gives the minimum completion time for that task. Another heuristic method Max-Min is similar to min-min except that it selects the longest task(with maximum completion time) first to schedule on the best machine available based on the minimum completion time of that particular task on all available machines. Genetic Algorithm and Simulated Annealing are two other general methods in heuristic approach which is used to perform near optimal scheduling [2].

2) Dynamic Scheduling

Tasks are dynamic in nature within dynamic scheduling. Tasks arrive at different points of time and it is based on the system machine's state. Dynamic scheduling algorithms are classified as: (1) online mode and (2) batch mode. In online mode tasks are assigned instantly once they arrive in the system like most-fit task scheduling algorithm where as in batch mode tasks are collected as a group and scheduled at predefined times[2]. K-Percent Best is one heuristic of same kind in which, a subset of k computationally higher ranking machines is first selected during the scheduling process. Min-min, max-min, round robin are some examples for batch mode. MCT, MET, OLB belongs to online mode. Switching algorithm is another type of algorithm in which it switches between MET and MCT as per the load of the system. Dynamic scheduling must depend on not only the submitted tasks to cloud environment but also the current states of system and computer machines to make scheduling decision.

b. Hybrid Scheduling Algorithms

Many of these algorithms are developed on certain scheduling parameters to improve the performance. Some of the existing algorithms under this category are given below:

Tasks are scheduled based on their cost to different resources. The cost of services varies for different tasks based on their complexity [2]. The algorithm considers resource cost and processing capability of resources. They group the tasks based on the processing capacity and selects some best resources to schedule them in such a way to reduce cost. This algorithm reduces the makespan and the processing cost when compared to other scheduling method [9]. Generally task's priority is calculated for scheduling them on various resources. Based on different attributes of tasks, priorities are calculated for the tasks and they are sorted based on that. Then they are assigned on the machine which produces the best completion time [9]. Hence this algorithm improves performance by having better completion time. In [10]; an algorithm is developed based on traditional min-min algorithm which includes scheduling based on load of the servers as well as considering the user priority. The users are classified into two categories as VIP users as well as the ordinary users. Load is balanced based on the maximum loaded resource and the makespan of the system. The method has advantage in user satisfaction, make span and resource utilization ratio.

In Priority based job scheduling, priorities are assigned to the individual tasks in the applications. Parameters considered in this are priority of the tasks and expected completion time. Advantages of this method are priority is considered for scheduling. Design is based on multiple criteria decision making

model. Disadvantage of this method are Makespan, consistency and complexity of the proposed method can be considered for improvement.

User Priority guided Min-Min is another method [2] for task scheduling in cloud environment. Parameters for this method are Priority, Load balancing, Makespan and Utilization of resources. Priority is given to users improving load balancing and without increasing the total completion time. These are few advantages of this method. Disadvantage is rescheduling of tasks to perform load balancing can increase complexity and time.

III PROPOSED METHOD

Based on the above discussion priority based job scheduling has certain drawbacks. Priority based task scheduling by mapping conflict-free resources and workload in cloud computing environment is the proposed scheduling method in this paper to overcome them. In the proposed priority scheduling algorithm a parameter named “priority” has been introduced. The tasks are classified into high, medium and low based on the priority. The priority is assigned based on the computational complexity of the task and by mapping the conflict-free resources. In this scheduling method, a higher priority is assigned to task of higher computational power. The fastest conflict-free resource available is assigned to the task of high priority. This priority algorithm optimizes the computational speed of the cloud and shows a consistent performance during execution of the assigned task.

In the process of task scheduling, users submit their jobs to the cloud scheduler. The cloud scheduler inquires the cloud information service to get the status of available resources and their properties. They will allocate the various tasks on different resources as per the task requirements and based on the priority of the tasks. It will also map the resources to the task if the resources are free (conflict-free resources). Cloud scheduler will assign multiple user tasks to multiple virtual machines. Good scheduling always assigns the virtual machines in an optimal way [2]. This will help to balance the workload in optimal way. Efficiency of scheduling mechanism in cloud environment depends on how efficient it is in managing the processes and increases the performance of the server as well as resources. As discussed earlier in above sections there are various problems in previous scheduling mechanism, so there is need to minimize them in all possible ways, in order to increase the efficiency. Priority based task scheduling has task present in queue in the arrival order. In Priority based task scheduling, we will take input from user i.e. the user tasks. We need to maintain queue of the user task. Assign priorities to the task. Higher priority is given to the tasks with

high power or complexity. Request for the resources required for task and schedule them. Allocate the conflict-free resources to the user task.

Priority Assignment: In Priority based task scheduling higher priority is assigned generally to a process which needs high computational power and which exhibits high parallelism. A task, which requires low computational power for execution is given a low priority. A task, which exhibits a medium level of parallelism and needs medium computational power, is allocated medium priority. The fastest conflict-free resource available in the cloud is allocated to the task which has high priority. This method of prioritizing tasks enhances the rate of completion of task with a greater accuracy and with proper usage of resources. Priority based algorithm has good performance then Round Robin, FCFS, MIN-MIN, MAX-MIN and many more due to its enhanced usage of resources [11].

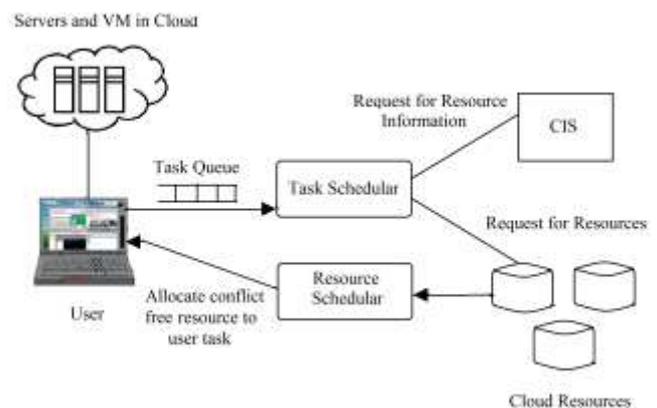


Figure1. Scheduling in Cloud Computing Environment

The main goal of Priority based task scheduling by mapping conflict-free resources:

- To process the tasks having higher priority keeping in the mind that task with the low priority should not starve.
- Conflict-free resource utilization.
- Minimizes the completion time (make span) of the tasks.
- Optimized workload scheduling.
- Minimizing the waiting time.
- Improve the utilization of servers allocated to the task.
- Optimized workload utilization in cloud computing.

IV CONCLUSION

In cloud computing environment scheduling is a wide concept. In this paper we have specifically focused on task scheduling. In cloud computing, designing a scheduling algorithm with an aim to perform scheduling task to a resource in an optimized way has been a complex task. During scheduling there are various techniques and constraints applied but as the cloud computing is too vast all aspects are not being able to capture at the same time. This paper explores various methods of scheduling done in cloud computing. The management and scheduling of resources in Cloud environment is complex, and therefore demands for optimized scheduling algorithm before applying them to the real system. The proposed task scheduling method for cloud computing environment is based on Priority based task scheduling by mapping conflict-free resources. The method also optimizes the number of physical resources while satisfying the conflicts. This method has better makespan than other task scheduling algorithm. And hence improve the throughput of the system. Proposed method can be further improved by considering more QoS parameter. A better scheduling algorithm can be developed from the existing methods by adding more number of parameters. By combining these different parameters an efficient scheduling algorithm can be obtained which can improve the overall performance of the cloud services as further enhancement.

ACKNOWLEDGMENT

Anagha Yadav would like to thank to her guide Assistant Prof. S.B.Rathod for his guidance and instructive comments in this paper work. Lastly, I offer my regards to all of those who supported me in any respect during the completion of this paper.

REFERENCES

- [1] Arun Chandrasekaran and Mayank Kapoor – “Increasing Acceptance of Cloud Computing in the Public Sector” Frost & Sullivan 2011.
- [2] Teena Mathew, K. Chandra Sekaran, John Jose "Study and Analysis of Various Task Scheduling Algorithms in the Cloud Computing Environment". 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI).
- [3] Khalid Bijon, Ram Krishnan, Ravi Sandhu "Mitigating Multi-Tenancy Risks in IaaS Cloud through Constraints-Driven Virtual Resource Scheduling" acm.org.SACMAT'15, June1–3, 2015, Vienna, Austria 2015 ACM.
- [4] Dr. Amit Agarwal, Saloni Jain “Efficient Optimal Algorithm of Task Scheduling in Cloud Computing Environment” International Journal of Computer Trends and Technology (IJCTT) – volume 9 number 7– Mar 2014
- [5] Raja Manish Singh, Sanchita Paul, Abhishek Kumar, Raja Manish Singh et al, “Task Scheduling in Cloud Computing: Review” (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 7940-7944
- [6] Ram Kumar Sharma and Nagesh Sharma, "A Dynamic Optimization Algorithm for Task Scheduling in Cloud Computing With Resource Utilization", International Journal of Scientific Engineering and Technology, Volume No.2, Issue No.10, pp. 1062- 1068.
- [7] S.Rekha and R.Santhosh Kumar “Priority Based Job Scheduling For Heterogeneous Cloud Environment” IJCSI International Journal of Computer Science Issues, Vol. 11, Issue 3, No 2, May 2014 IJCSI.org
- [8] S.Nagadevi, K.Satyapriya, Dr.D.Malathy, “A Survey On Economic Cloud Schedulers For Optimized Task Scheduling”, International Journal of Advanced Engineering Technology, 2013.
- [9] Qi Cao, Zhi-BoWei, Wen-Mao Gong, “An Optimized Algorithm for Task Scheduling Based On Activity Based Costing in Cloud Computing”, IEEE, 2009.
- [10] Huankai Chen, Frank Wang, Na Helian, Gbola Akanmu, “User-Priority Guided Min-Min Scheduling Algorithm For Load Balancing in Cloud Computing”, IEEE, 2013.
- [11] Lipsa Tripathy, Rasmi Ranjan Patra “Scheduling in Cloud Computing” International Journal on Cloud Computing: Services and Architecture (IJCCSA) ,Vol. 4, No. 5, October 2014
- [12] Gowtham Gajala, “Cloud Computing: A State of Art of the Cloud” International Journal of Computer Trends and Technology- volume4Issue1- 2013 ISSN: 2231-2803 <http://www.internationaljournalsrsg.org>
- [13] S.Sathyavani M.C.A, T.P.Senthilkumar M.C.A.,M.Phil., “Survey on Cloud Computing” International Journal of Computer Trends and Technology (IJCTT) – volume 4 Issue 9– Sep 2013 ISSN: 2231-2803 <http://www.ijcttjournal.org>
- [14] Rajveer Kaur, Supriya Kinger, “Analysis of Job Scheduling Algorithms in Cloud Computing” International Journal of Computer Trends and Technology (IJCTT) – volume 9 number 7 – Mar 2014 ISSN: 2231-2803 <http://www.ijcttjournal.org>