A Survey on Service Request Scheduling In Cloud Based Architecture

Vinita Tiwari[1] Shikha Agrawal[2]
[1,2]Department of Computer Science and Engineering
[1,2]Chhatrapati Shivaji Institute of Technology, Durg, Chhattisgarh

Abstract— Cloud computing has become quite popular now-a-days. It facilitates the users to store and process their data which is stored in 3rd party data centers. Today in IT sector everything is run and managed on the cloud environment. As the number of users is increasing day by day, faster and efficient processing of large volume of data and resources is desired at all levels. So the management of resources attains prime importance. While using cloud computing various issues are encountered like load balancing, traffic while computation etc. Job scheduling is one of the solution of these problems which reduces the waiting time and maximizes the quality of services. In job scheduling "priority" is an important factor. In this paper, we will be discussing various scheduling algorithms and a review on dynamic priority scheduling algorithm.

Key words: Cloud Computing, Job Scheduling, Priority, FCFS, SPSA, Dynamic Priority Scheduling Algorithm

I. INTRODUCTION

Cloud provider provides services to their customers and charge as per usage by a particular customer. In short, we can say pay per use. For establishing any kind of business, you need infrastructure and other services. It is not always possible to manage all these things on your own. Cloud computing gives us an opportunity to use remote infrastructure and diverse services economically so that we can focus on core competency of the business.

It focuses on maximizing the effectiveness of shared resources. Cloud resources are usually shared by multiple users and dynamically reallocated as per demand. Cloud computing environment are mainly categorized into three ways-

1) Private cloud- these types of cloud exists in the organization because it provides special benefits to them.
2) Public cloud- as the name suggests, it is for public .Services are offered by third parties and managed by them.
3) Hybrid cloud- it is the combination of private and public environment.

There are 3 kinds of services provided by cloud computing which are as follows:

A. Infrastructure as a Services (Iaas):

Services in the form of infrastructure like server, storage network, operating system which deploy on demand services are provided. The resources are distributed which allows for dynamic scaling.

B. Platform As A Services(Paas):

it is defined as a computing platform that concede the creation of web applications rapidly and efficiently without the complexity of buying and maintaining the software and supporting infrastructure.

C. Software As A Services(Saas):

it is fast becoming one of the most rapidly adaptive method for delivering of new technologies. Here users do not bother about up-gradation of software and patches.

There are multiple resources in cloud environment like server, network, application, software and services. Cloud environment is a delivery of computing as a service with shared resources, software and information over a network.

There are several cloud management techniques that includes infrastructure management, Cloud Resource usage management, cloud security management, management as a service, energy management over cloud and cloud management for load balancing of tasks amongst others.

D. Service Request Scheduling:

It is the most crucial area with respect to the profit of the service provider and the quality of services to the user. With respect to the service request scheduling, the entire cloud architecture comprises of three- tier architecture resource provider, service providers and the consumer.

The consumer requests for the services to the service provider and to fulfill the request the service provider needs either to get hold of new hardware or to hire it from a resource provider.

Taking resources on rent is always profitable for users rather than purchasing a new one. The resource provider is responsible for providing the instances to run on physical or virtual resources, so the charges of the running instance are based on the flat rate per time unit. As we know cloud computing is based on "pay-as-you-use" utility, service provider needs to reduce the response time and delay in processing the request. So over here service request scheduling is required to reduce the cost and maximize the performance.

II. SCHEDULING ALGORITHMS

A. FCFS:

In first come first served scheduling method allocation is based on the arrival time which ensures fairness but it takes all the task units priority at the same level. The disadvantage of fcfs is that it is non-preemptive. In general the task comes from different users and we can say that some tasks are very important and some are normal tasks. Important task must be scheduled without delay but it is not possible in fcfs.

B. Static Priority Scheduling Algorithm:

Here we consider the priority so each task is assigned with a priority and the priorities are fixed before the scheduling. It is an efficient algorithm but at times it is not efficient for lower priority tasks.

C. Round Robin:

It uses the concept of time quantum or slices. Here the processes are dispatched in a FIFO manner but are given a limited amount of time(time quantum). The resources are provided to the process on the basis of time quantum. If a process does not complete in its time quantum, the CPU is preempted and given to the next process which is waiting in a queue. The preempted process is then put in to the back of the ready queue. The advantage of this algorithm is that it utilizes all the resources in a balanced order so it is good for load balancing but the power consumption is high because each process will be kept on for a long period of time.

D. Dynamic Priority Scheduling Algorithm:

Here we provide priority at the run time. It is based on the static priority scheduling algorithm with dynamic features.

III. LITERATURE REVIEW

Lee, Wang, Zhou(2011) presented that On comparing first come first served(FCFS), static priority scheduling algorithm(SPSA), dynamic priority scheduling algorithm(DPSA), the DPSA will provide better efficiency and fairness. DPSA will be more feasible than SPSA[1].

Lokesh kumar Arya, Amandeep Verma (2014) suggested that With the help of Workflow Scheduling Algorithm, the parameters like reliability, load balancing, fault tolerance can be enhanced[2].

Rajveer Kaur, Supriya Kinger(2014) in this paper a number of existing algorithms for job scheduling are discussed and compared with each other[5].

Nimisha Singla, Seema Bawa (2013) suggested that the algorithm implemented till now will provide optimal solution but if the load increases in cloud, there is a need of more efficient algorithm[4].

Mahendra Singh Sagar, Babita Singh, Waseem Ahmad (2013)

Proposed that Cloud Services are not only for data storage and software provision, but also it is about intelligent utilization of the available computing resources. The resource utilization analysis is based on memory usage [3].

Rizwan Mian, Patrick Martin (2012) presented that Estimate the expense of executing a workload in a cloud. It reduces time to result by exploiting rapid provisioning of clouds resources [6].

Dr. Amit Agrawal,Saloni Jain (2014) proposed a generalized priority algorithm and the experiment is conducted for varying number of virtual machines and workload traces than it also compared with FCFS and Round Robin[7].

A. Related Work:

In paper[1] author proposed dynamic priority scheduling algorithm[dpsa]. It is based on the static priority scheduling algorithm with dynamic features.

The important components in service request scheduling are-

1) Classifier:

It receives the request from the users, analyzes and classifies them into smaller task units. Before scheduling the task unit, it requires to get assigned with random priority.

2) Scheduler:

It contains several schedule units and put every task into the appropriate schedule unit. The task units are executed in scheduler based on the algorithm

3) Summary Component:

It summaries the task units from different schedule unit in a cycle time and send it to resource provider.

The DPSA structure contains several queues and the number of queues depends on how many priorities the task unit have. Before the task units are sent to the scheduler, it has its own initial priorities. Each task comes into the scheduler and is divided into task units, which are put into their corresponding queues with the cycle time. The schedule unit puts a new task unit at the end of its corresponding queue. Every schedule unit does the schedule process with the cycle time. The purpose of the introduction of dynamic way is to avoid a task unit with low priority wait for so long. To achieve this goal, each queue has a threshold A_k which means the limited time a task unit can wait in a queue of the schedule unit, and the k means the task unit's priority. If a task unit has waited for an A_k time, the schedule unit will move the task unit to a high priority queue. If the task unit in the highest priority queue has waited for an A_k time, the schedule unit will send this task unit to summary component immediately.
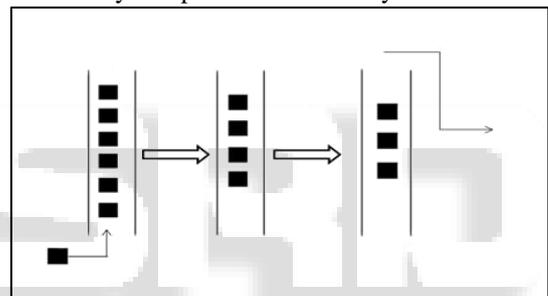

Fig. 1: schedule unit with priority queue
Here the author concludes that the DPSA is fairer than FCFS and SPSA but DPSA is less efficient than SPSA.

IV. CONCLUSION AND FUTURE WORK

In this paper a number of existing scheduling algorithms are discussed. First come first serve algorithm has some disadvantages like average waiting time is often more. In round robin scheduling also doesn't give special priority to important tasks and also the power consumption is high. In static priority scheduling algorithm, higher priority task are treated very well and its efficiency is also high but it is not good for lower priority tasks. DPSA is also based on SPSA and has a dynamic feature. Its efficiency is less but it is good for lower priority tasks also. This describes that there is a scope of improvement so that all the objectives can be achieved. So in future, the author will analyze existing algorithms and explore the possibilities of dynamic priority algorithm.

REFERENCES

[1] Zhongyuan Lee, Ying Wang, Wen Zhou "A dynamic priority scheduling algorithm on service request scheduling in cloud computing" IEEE-2011.

[2] Lokesh kumar Arya , Amandeep Verma " Analysis of job Scheduling algorithms in cloud computing" IEEE-2014.

[3] Mahendra Singh Sagar, Babita Singh, Waseem Ahmad, "Study On Cloud Computing Resource Allocation Strategies", International Journal Of Advance Research And Innovation, Volume-1, Issue-3, (2013) 107-114.

[4] Nimisha Singla, Seema Bawa "Review of Efficient Resource scheduling algorithms in cloud computing" International Journal Of Advance Research And computer science software engineering-2013.

[5] Rajveer Kaur, Supriya Kinger , "Analysis of job Scheduling algorithms in cloud computing" , International Journal of Computer Trends and Technology volume-9 number 7 Mar 2014.

[6] Rizwan Mian, Patrick Martin, "Executing Data Intensive Work Load In A Cloud" IEEE/ACM International Symposium on Cluster Cloud and Grid Computing, 2012/ 12th

[7] Dr. Amit Agrawal,Sonali Jain, "Efficient optimal algorithm of task scheduling in cloud computing environment" in international journal of computer trends and technology(IJCTT) vol.9no.7,mar-2014

[8] http://www.cs.uic.edu/~jbell/CourseNotes/OperatingSystems/6_ CPU_Scheduling.html.