



A Comparative study of Resource Scheduling in Cloud Computing

Kiran Bala*Department of computer science & CEC,
Landran (PB), India**Sahil Vashist**Department of computer science & CEC,
Landran (PB), India

Abstract - Cloud computing is new technique that provides resources to the users as per demands. In the cloud, there are cloud provider that provides the resources and consumers that use or consume the resources. The users can access the resources (storage, server, network bandwidth, information and so on) from cloud as per demands, only required internet connection. The user can pay for these resources as per usage of resources. In cloud, these resources must be scheduled so that efficiently used. Scheduling of resources is most challenging task. Scheduling is required for execution of task in time, for less cost and also other parameters. In this paper, compare the sequential algorithm, optimum scheduling algorithm, Urgency first parallel strategy and Priority Impact Scheduling Algorithm These algorithms performs better resource scheduling in cloud environment.

Keywords— scheduling, sequential, optimum, Urgency first parallel strategy and Priority Impact Scheduling Algorithm

I. Introduction

Cloud computing is an emerging technology in which resources are made available ubiquitously for the users-on demands whenever they needed. The resources are available to the users as with internet connection. The users can access the information, data and application from cloud at any time and at anywhere, must be internet connection without any worrying about managing the data, information, resources, infrastructure and so on. Example of cloud computing are Hotmail, Yahoo, Gmail, Face book, etc. cloud computing provides pool of communal resources like server, storage, network resources and virtual machines etc. The user can access these resources at anytime from anywhere and the most beneficial thing is user can pay only as per usage of cloud. The cloud computing uses virtualization technique for providing the resources to users. With the help of virtualization, virtual machines can be create and destroy according to requirement. In the cloud, the datacenter manages hosts and hosts run virtual machines on a physical machine and work like a real machine. In host, the virtual machine can be created and destroy at anytime by datacenter broker. Datacenter broker manage the datacenter and allocate and reallocate the resources to the requests [1] [2].

A. Cloud computing Service model

The cloud providers provide services –as a service to the consumers. So the users can access the services and also access the data and information from the cloud [3].

- IAAS: Infrastructure as a service (IAAS) provides infrastructure, storage, servers, processing power, virtual machines, switches, routers etc. The consumer or uses has no need to control over infrastructure. Consumer can access the resources storage, servers etc when required. Example of IAAS is Amazon EC2 and GoGrid.
- PAAS: Platform as a service (PAAS) provides platform to design, deploy and develop any application. Consumer does not control over the infrastructure of cloud but consumer can control the deployed application. It provides run time environment for execution of any application. Example is Google App Engine. This Google engine enables to design any application by providing a development kit for developer.
- SAAS: Software as a service (SAAS) provides software where consumers can store, access and transfer the data and information. Consumers can access the data through web with internet connection. Now no need to install or update the software on personal laptops, computers and desktops, Example of SAAS is Gmail where consumer can store, access the data and also no need to update Gmail software.

B. Cloud Computing Deployment Model

The cloud provider provides the four deployment model. These models provide services to the consumers at anytime and anywhere. These models are:

- Public model: From public model, consumer can access the data from anywhere. It is an open model where consumer can access data publically and pay as per usage. This model is managed by provider.

- Private model: This model is private to organization or company and fully control over all the services. This model is managed by owner, so this model provides more security.
- Hybrid model: This model is combination of both public and private model. This model can provide security and also no need of any hardware/infrastructure because organization can process or transfer the data in public cloud and store the data at private cloud.
- Community model: This model is used by specific group of members of organization or company for share the data.

There are number of strategy that still under research. Many strategies have described in this paper. In section 2 defined the related work of scheduling algorithm. In the next section 3 described the scheduling, process, need and types of scheduling. Rest of paper defined the algorithm in section 4. In section 5 described the pros and cons of the algorithm. At the end section 6 defined conclusions.

II. Related Work

Jing Xiao[4] has proposed Scheduling approach provides the service to the consumers by schedule the requests. In this, virtual machines are scheduled on physical hosts in cloud environment. The allocation of request in to the virtual machine according to priority algorithm. By this priority algorithm, when there is less resources than required then ranked the requests according to maximize the benefits of providers.

Karan D. Prajapati[5] has proposed many algorithms for managing the resources at cloud. In cloud, there are many different virtual machines provided by provider to the consumer as per requirement. So there is a need of scheduling these virtual machines. The Round robin, Genetic and Memory-Aware Cloud scheduling algorithm has been proposed to estimate the cost, security, time and bandwidth. Different algorithm provides different benefits.

Chandrashekhar S. Pawar[6] has proposed a algorithm priority based scheduling algorithm(PBSA). In this algorithm multiple SLA (service level agreements) are considers. Multiple SLA is CPU time, bandwidth, memory etc. This algorithm is compare with cloud min-min scheduling algorithm and in disagreement condition, priority based scheduling algorithm perform superior results than cloud min-min algorithm.

Ramkumar N, Nivethitha S[7] has proposed algorithm Effective Resource Utilization Algorithm (ERUA) .This algorithm based upon 3 tier architecture of cloud. In three tier consumers, resources providers and service providers. The cloud users rely on quality of service and providers rely on profits. To accomplish these, need an efficient algorithm which reduce the waiting time, cost and maximizes the Qos by utilization of resources effectively. Our algorithm completes the prerequisite of the users and the providers by efficient Scheduling of resources and reassignment of priority.

Hsu Mon Kyi[8] has proposed Efficient Virtual Machines Scheduling Algorithm for allocation and scheduling of virtual machines and virtual resources. Stochastic Markov model is used for analyzing the performance of scheduling the resources. *This model using open sources system Eucalyptus*. In this measure the effects of varying workload and capacity of system on cloud. This algorithm improves the response time of the system.

Helen D. Karatza [9] has proposed Gang Scheduling algorithm for scheduling parallel jobs. The number of virtual machines available dynamically and scale up and down according to requirement. In this algorithm resulting a range of various workloads, migration of jobs, job size and handling the job starvation while takes into performance and cost parameters. This scheduling algorithm effect on cloud and improve the performance.

III. Scheduling

Scheduling is the procedure of schedule the tasks or requests of consumers according to available resources on cloud computing environment on the basis of condition, requirement and uniqueness of the tasks. In cloud computing environment, there are various tasks that are to be executed. For execution of tasks in cloud, there must be requirement of resources like bandwidth, memory, virtual machines etc and these resources are managed by resources management. So resource management manages or schedules these resources. The various tasks executed with the available resources in cloud to attain best performance, less response time of task execution, utilization of resources, less total time of execution and so on. For completion of all these objectives, need to be design, develop and deploy a scheduling algorithm. Scheduling algorithm is urbanized to complete many goals. The goals are high throughput, better quality of services and efficient use of resources.

The request of resources from consumer in a cloud will be delivered in term of virtual machine. In the datacenter there are various virtual machines that are created, destroyed and managed by virtual machine manager. When the requests of resources from consumer, then allocate and reallocate the resources to the requests and manage the resources. For effectual use of resources provides the scalability, high throughput, elasticity, less response time of task completion. [10]

A. The process of allocation of request in cloud

In the cloud environment, there are many hosts that run various virtual machines. The virtual machines allocation and reallocation is done by datacenter broker. Datacenter broker manage datacenter geographically. The number of requests are allocated and reallocated to virtual machine by scheduling process. Now the process of scheduling:

- Discovering of resources: When the request for resources, broker discovers resources in datacenter that is nearer to the location of request.
- Selection of resources: After discovering of resources select the virtual machine that perform best output for that request under certain parameters of tasks and then allocate the request to that virtual machine.
- Submission of task: After resources selected. Then task is submitted and executed.

B. Need of Scheduling

Task scheduling is allocating the particular resources to the tasks for execution in particular time. The task scheduling is done by job or task scheduler that should be dynamic in nature. In cloud computing the request of resources are dynamic so task scheduler must manage all these requests and enhance the utilization of resources and decrease in completion of request time so that resources are reallocated for another request. When resources are managed properly then less refusal of requests or tasks takes place and in clouds more and more request can be submitted.

C Types of scheduling

There are various types of scheduling depend upon criteria:

- *Static Scheduling*: Static scheduling means tasks are Pre-Schedule, all information about the tasks and resources available are known. Once a task is assigned to a particular resource and executes the tasks. As a result it's simpler and easier to become accustomed based on scheduler's viewpoint. Scheduling can be done at compile time is known as static or code scheduling. This scheduling supports the dynamic scheduling. The static scheduling needs intellectual compilation support. The example of this is FCFS (First-Come-First-Severed).
- *Dynamic scheduling*: Scheduling can be done at run time is known as dynamic scheduling. Tasks are scheduling by scheduler dynamically. Dynamic scheduling needs complicated hardware support. Dynamic scheduling is helped by static scheduling to get better performance and to reduce the cost of hardware. The example of this is Genetic Algorithm
- *Centralized Scheduling*: Centralized scheduling is responsibility of centralized scheduler for decision. The scheduler control, manage and maintain the resources and schedule them. The important benefits of scheduling are more control on resources and maintaining the resources and ease of implementation of resources and increased performance.
- *Distributed / Decentralized Scheduling*: In this scheduling no central manage entity. This scheduling is decentralized means local scheduler's managed and controls the requests. This scheduling has weak point as compared than centralized.
- *Pre-Emptive Scheduling*: In this scheduling, during the job execution sometimes an interruption is occurred. When this occurred then need to be migrated the task from original virtual machine to another virtual machine so that task is completed within time.
- *Non Pre-Emptive Scheduling*: In this scheduling the task are not being allocated to another virtual machines until it complete its execution. There is no migration of tasks even high priority of tasks.
- *Co-operative scheduling*: In co-operative scheduling, the system have already many schedulers and in this each scheduler execute certain activity towards common goal and cooperate each other.
- *Immediate / Online Mode*: In this scheduling, when the tasks arrive as soon as possible resources are allocated with no waiting time for tasks.
- *Batch / Offline Mode*: In this scheduling, the scheduler not allocates the resources as soon as tasks arrive. Scheduler possible to stores incoming requests as group and solved that group over consecutive intervals.

IV. Algorithms

In this section describe the comparison between the sequential algorithm and optimum scheduling algorithm.

A. Sequential algorithm

In Sequential algorithm, all the tasks/ requests for resources are executed sequentially. The sequentially means when the request from consumer for resources then resources are allocated to that request and after that when the request completed its execution then the resources are free for another request, otherwise waiting in waiting queue. The allocation of resources and execution of requests are performing sequentially.

The main drawback of this algorithm is sometime when the request executing may takes more time and required more resources but give less profit to the provider and another waiting request may take less time and give more profit waiting for resources [11].

B. Optimum scheduling algorithm

In this algorithm, the incoming requests /tasks from consumers to the cloud broker are grouped. The grouping of tasks is based on the requirement of time, cost and deadline constrained. After that grouping of initial request then they are prioritized based upon profit or deadline. The prioritized of requests is required. Without prioritized of the tasks neither checking the deadline of tasks nor gain the benefits from tasks. So this is necessary to prioritize the tasks according to shorter deadline and more profit gain from execution of task. So need to be scheduling first that have shorter deadline and gain more profit. The parameter of prioritizing is different that depend upon the kind of task that execute. For prioritized request or task the time limit is:

Turnaround time: Calculate the turnaround time for each resource and subsequent parameters are required:

- waiting time
- Processing Power of virtual machine
- Task length

C. Urgency first parallel strategy

In UFPS algorithm, there are required local and cloud service pool. In local pool of resources the consumers can get the resources locally and in cloud pool consumers can get the resources from cloud through internet connection. The service providers provide the resources to the consumers on the basis of emergency or urgency level of tasks.

The request arrives then put same type of tasks together. Special tasks put in one queue, complicated tasks put in one queue and tasks require small amount of calculation put in another queue. Then all the tasks in its own queue are arranged according to emergency level and executed one by one according to put in queue. The urgency tasks is put into the front of queue and scheduled first. The tasks that require many and complicated calculation that tasks is scheduled to that cluster whose ability is strong. The tasks that require small calculation that processed the cluster that has small ability and special tasks execute by special cluster. When tasks execute then check the idle or inactive resources in local clusters and cloud clusters at same time. When the resources are idle in local then it will quit the scheduling in cloud and similarly if idle resource in cloud cluster then stop searching resources in local cluster. Else if resources are not free from both then put the task in tail of queue. When complete scheduling round then time updated in queue [12].

D. Priority Impact Scheduling Algorithm

In PISA algorithm based on priority of tasks and execute according to priority assign. The priority is defined on the basis of fee paid by consumer or depends upon other factor. In this uses the prototype algorithm and then PISA. In prototype algorithm assign the priority to the tasks according to the fee paid. So provider classify the consumer according to priority as free consumers, VIP and so on and execute the tasks according to classify and put the value and access the resources faster. If some tasks not execute due to its low priority and put in queue for a long time then value of task in incremented by 1 after completing the round. If this value exceeds the fixed value that is already set before execution then task access the resources in next round in spite of the priority. After that PISA algorithm set up the procedure Access Policy Check that query the consumers about Access Strategy library that explain the name of policy as follow Workflow level , priority and task start and end time. The value defines the priority so that access the resources. If value of time meet the request level then request is accepted otherwise rejected [13].

V. Pros and cons of algorithms

Technique	Pros	Cons
Sequential algorithm	<ol style="list-style-type: none"> 1. This algorithm is easy to use. 2. The requests are executed according to they are enter like FCFS (first come first serve) 	<ol style="list-style-type: none"> 1. The main drawback is sometime the important request is waiting in the queue and another less important request executes and takes more time. So it is not beneficial for providers.
Optimum algorithm	<ol style="list-style-type: none"> 1. The requests are executed according to priority. Higher priority request execute first. 2. By this algorithm utilization of resources properly and all the requests executes in time. 	<ol style="list-style-type: none"> 1. Sometimes it is difficult to assign the priority to the tasks. Priority can be depends upon number of factors like revenue, time, utilization of resources etc. 2. Another difficulty is sometimes two requests have same priority and may want to use same resources. So it is difficult to the providers to assign the resources to the requests.

Urgency first parallel strategy	1. It consume less time for execution of task. 2. Emergency tasks scheduled and execute first and give profit to the provider. 3. The resources is provided locally as well as at cloud. The consumer can access both at same time	1. This algorithm checks the resources from local and cloud need to do extra work. 2. The consumer or enterprise must maintain local cluster.
Priority Impact Scheduling Algorithm	1. Priority is assigned by consumers according to the fee paid and executes the task according to priority. 2. If some tasks don't execute due to some reason then this algorithm execute the task in next round of scheduling, does not matter the priority of task.	1.It only depends upon fee that consumers pay.

VI. Conclusion

It is covered the scheduling algorithm that is used in cloud for scheduling the resources. In this explain the optimum algorithm enhances utilization of resources for execution of the tasks as Sequential Algorithm. Also defined, the Urgency first parallel strategy and Priority Impact Scheduling Algorithm. This algorithm executes the task according to emergency and priority defined by consumer. The priority is defined on the basis of fee paid by consumers.

References

1. Sahil Vashist ,Rajwinder Singh, "Energy Cost and QoS based management of Data Centers Resources in Cloud Computing" in International Journal of Computer Science & Engineering Technology (IJCSSET), Vol. 4 No., 06 Jun 2013
2. Lohit Kapoor, Seema Bawa, Ankur Gupta, "Hierarchical Chord-Based Resource Discovery in Intercloud Environment" in IEEE/ACM 6th International Conference on Utility and Cloud Computing, Pages 464-469
3. Kiran Bala, Sahil Vashist, Rajwinder Singh, Gagandeep Singh, " A REVIEW OF THE LOAD BALANCING TECHNIQUES AT CLOUD SERVER" in International Journal of Advances in Computer Science and Communication Engineering (IJACSCE) , Vol 2 , Issue I, March 2014.
4. Jing Xiao, Zhiyuan Wang, " A Priority based Scheduling Strategy for Virtual Machine Allocation in Cloud Computing Environment" , International Conference on Cloud Computing and Service Computing, 2012
5. Karan D. Prajapati ,Miren Karamta, Miren Karamta, M. B. Potdar, " Comparison of Virtual Machine Scheduling Algorithms in Cloud Computing" in International Journal of Computer Applications (0975 – 8887) Volume 83 – No 15, December 2013
6. Chandrashekhar S. Pawar, Rajnikant B. Wagh, " Priority Based Dynamic resource allocation in Cloud Computing".
7. Ramkumar N, Nivethitha S, "Efficient Resource Utilization Algorithm (ERUA) for Service Request Scheduling in Cloud" in International Journal of Engineering and Technology (IJET)
8. Hsu Mon Kyi and Thinn Thu Naing , "STOCHASTIC MARKOV MODEL APPROACH FOR EFFICIENT VIRTUAL MACHINES SCHEDULING ON PRIVATE CLOUD" in International Journal on Cloud Computing: Services and Architecture(IJCCSA),Vol.1, No.3,November 2011
9. Ioannis A. Moschakis ,Helen D. Karatza, "Performance and Cost evaluation of Gang Scheduling in a Cloud Computing System with Job Migrations and Starvation Handling" in IEEE, 2011
10. Swachil Patel, Upendra Bhoi, " Priority Based Job Scheduling Techniques In Cloud Computing: A Systematic Review" in INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 2, ISSUE 11, NOVEMBER 2013, ISSN 2277-8616.
11. Monika Choudhary , Sateesh Kumar Peddoju, " A Dynamic Optimization Algorithm for Task Scheduling in Cloud Environment" in International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 3, May-Jun 2012, pp.2564-2568
12. Zheng-Wu Yuan, Xin-Guang Sang, "A Study on Resource Scheduling Strategy in the Enterprise Service Cloud" in International Conference on Systems and Informatics (ICSAI 2012), 2012.
13. Hu Wu, Zhuo Tang, Renfa Li, "A Priority Constrained Scheduling Strategy of Multiple Workflows for Cloud Computing" in ICACT2012