A review of round robin algorithm on cloud computing task allocation.

Kishan Raj[1], Shubham kumar[2]

[1]Assistant Professor, Meerut Institute Of Technology, Meerut, India
[2]Assistant Professor, Department of computer science MGCU, Bihar, India

---***---

Abstract - The cloud computing infrastructure has many issues including scheduling, budgeting load balance. The goal is to use resource efficiently and decrease resource consumption in the cloud environment. This can be achieved by increasing the LB rate when selecting the best resources for low work failures rates with low lead times. The round robin is a simplest and most efficient algorithm. It gives more accurate result than any other Algorithm to schedule VM task's in cloud computing. Many research scholars proposed various technique in order to improve the RR algorithm and selected dynamic time quantum which plays major role in improving this algorithm. Because of taking longer time in order to swap contexts many researchers provide us a dynamic time quantum strategy to overcome this type of limitations. Many dynamic time Slice is proposed, so in this paper we have compared some of the approaches with concerns of its waiting time, turnaround time, and efficiency of the algorithm. In reality, the arrived jobs are made up of a variety of interconnected tasks, some of which may run on different VMs or on different cores of the same VM. Additionally, the jobs arrive at random intervals throughout the server's operation under different load situations[2].

Key Words: CC, RR, VM, Time quantum

1.INTRODUCTION

Cloud computing has gained increased importance in the realm of IT services due to its remarkable adaptability and the proper distribution of resources. The effectiveness of a cloud computing environment is closely tied to how tasks are scheduled. Efficient task scheduling guarantees the appropriate allocation of resources for the tasks' associated requests. Cloud computing is an emerging technology in current era. It provide utility as computing like electricity, water. Cloud computing provide services like virtualization with operating system, software development, storage without any physical devices. Anyone can access these services without hardware, basically cloud is a chain of networks that can be accessed from anywhere using only internet connections. This technology reduced cost of computing and made anyone access large files or computing irrespective of hardware devices. Cloud was first introduced by Salesforce.com in late 1990s. Today major provider of cloud services are Amazon.com, Google etc. Amazon is largest giant in cloud technology with it's headquarter in San Francisco Cloud computing provide us services that any Company have facilities to rent access to anything. In this technology user can simply pay for what they use when they use. For example Gmail, Netflix, cloud backup photos uses features of cloud. The concept of renting access to computing power has resurfaced again and again in the application service. Cloud computing is a popular option for reasons such as cost saving, increased productivity, speed and efficiency, performance and security. Cloud computing is an abstract of computing, memory and network infrastructure collected as a stage on which applications and systems can be deployed swiftly. It is one of the best and emerging technology that has changed activities of Human life and fast in a big way and also providing various services, resources using the Internet and cloud-based resources in recent times through virtualization .Cloud computing in simple words means allocation of resources, retrieval of data and programs over the Internet without physical storage rather than our computer's hard drive. Those users that are working outside the working places are able to access resources through internet connecting devices. NIST (National Institute of Standards and Technology) has listed five essential required characteristics of cloud computing from virtual machine that include on- demand of cloud computing of self-service, broad network access, resource pooling, rapid elasticity and measured service. It also involves the concepts of parallel processing and distributed computing providing the distributed the resources in many consumers by means of Virtual Machines (VMs) hosted by physical servers. It is developing rapidly and it will be the next generation technology where humans can use resources remotely without concern of time. This technology has changed human life in all areas with the help of technology. This technology has expanded rapidly during covid time due to extensively use of technical resources. It is emerging internet based technology used to store the information data and accessing resource and application through web on internet. It stores very huge amount of data and information without any physical storage device and it works under certain terms and conditions. Virtual machine (VM) is one the main element of virtualization of cloud computing, when a single resource of cloud computing appeared as multiple resource this process can be achieved with the help of virtual machine. Cloud technology is emerging rapidly and used in different field of life like e-commerce, AI, education, engineering technology, data intensive applications, health, geospatial sciences and many more different scientific and business domains uses cloud-based server. Cloud computing has become an respected technology in worldwide due to offering a huge amount of storage and resource without any physical storage to different companies and organization to access this resource with proper management, rule and regulations. Some of the key features are virtualization, viability, wide network access, large information storage, automated system, economical security, scalability and many more. Cloud resources are variable with renting and releasing for each user specific rules and regulations with the help of the Internet. According to NIST, cloud computing is a model that supports on-demand network access to a shared pool of configurable computing software resources.in a cloud computing environment, there are three main categories of task scheduling algorithms which are conventional algorithms such as FIRST COME FIRST SERVE, SHORTEST JOB FIRST, ROUND ROBIN, HEURISTIC ALGORITHM, MIN- MAX and meta hieuristic algorithms such as ANT COLONY OPTIMIZATION and PARTICLE SWARM OPTIMIZATION.

1.1 Models in cloud computing

There are two types of models in cloud computing–

1. Service model
2. Deployment model

1.1.1 CLOUD SERVICE MODELS

(1) Infrastructure as a service

(2) Platform as a service

(3) Software as a service

1.1.2 CLOUD DEPLOYMENT MODELS

(1) Public cloud

(2) Private cloud

(3) Hybrid cloud

(4) Community cloud

1.2 Round robin in cloud computing

[2]This algorithm employs a sequential method to select servers for upcoming requests. However, a significant issue with this approach lies in the inconsistent performance of servers. To address this challenge, innovative strategies have been developed. It is one of simplest, old, fairest and most popularly used scheduling algorithms designed especially for time sharing systems. In this algorithm,

Table-1 : Description of terms.

The criteria	The Description
Minimizing the context switch (CS)	It is the process of changing the task's status from one activity to another.
Minimizing Turnaround time (TAT)	It is the time period between the starting time of the process and its completion time
Minimizing the waiting time (WT)	It is the total time spent by the process in waiting in between tasks.
Minimizing Response time (RT)	It is the time period in between queue submission and its first reply.
Maximizing CPU utilization	Trying to make the CPU as busy as possible.
Maximizing throughput	It is the number of tasks which completed in a specific time unit.

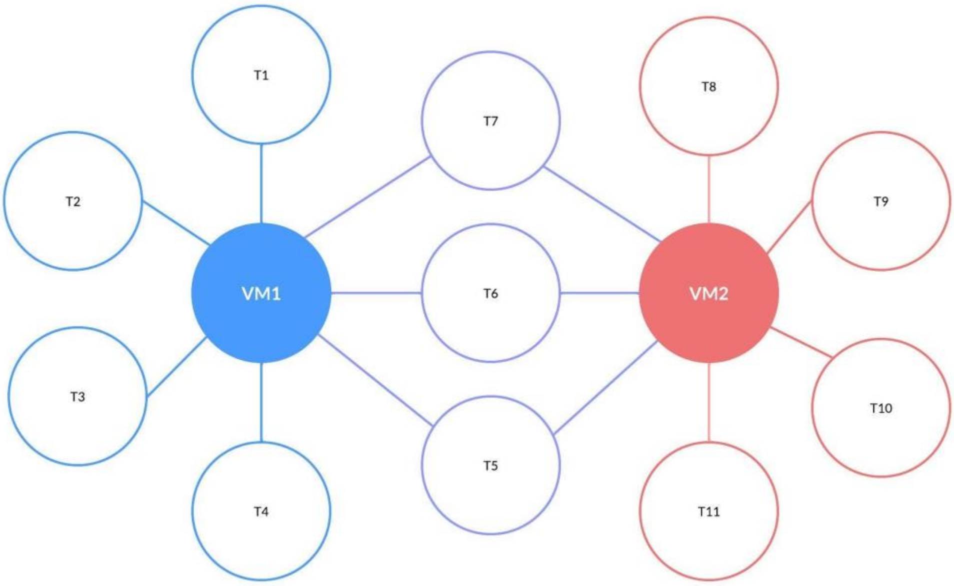
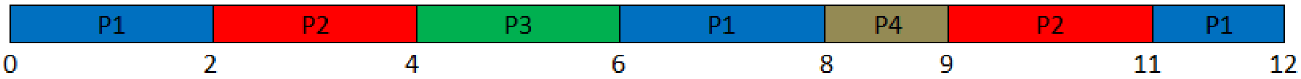


Figure 1: RR algorithm process in cloud computing

Time quantum is defined which is smallest unit of time. All the run able processes are kept in a circular queue and the CPU scheduler goes through all the processes with respect to time quantum of the CPU scheduler. New jobs are kept in tail of the queue. CPU scheduler picks the first processes from the queue and execute it with respect to time quantum. If the process does not finish in the static time quantum, the CPU is preempted and the process is added to last of the queue. If the process is finish in the given time quantum it will release CPU automatically. CPU scheduler start executing the next process of the ready queue. Every time when next process start executing, a context switch occurs, which add overhead to the process execution time.

The waiting time for each task
WT = turnaround time – burst time
P1: 12 - 5 = 7
P2: 10-4 = 6
P3: 4 - 2 = 2
P4: 5 - 1 = 4
Average wait time : 4.7

Figure 2: Gantt chart for RR



P1	0	5
P2	1	4
P3	2	2
P4	3	1

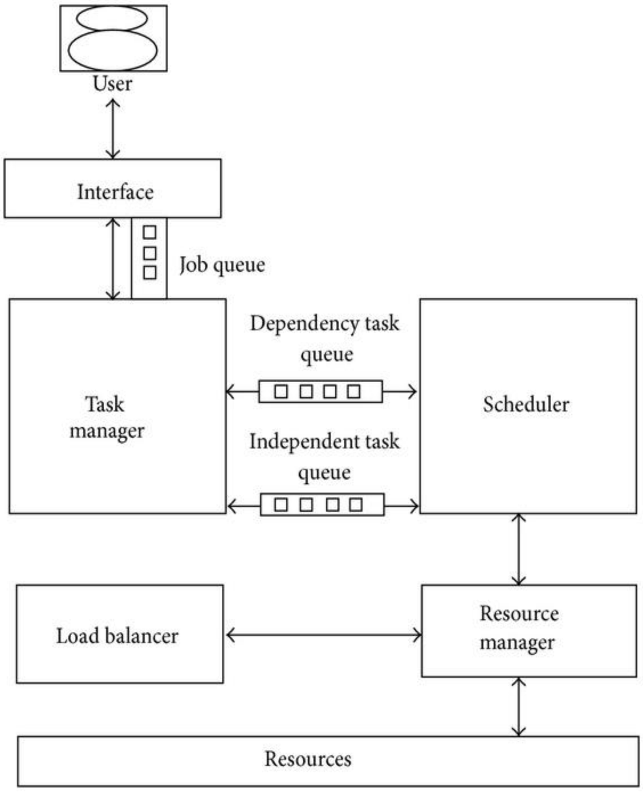


Figure 3: Design of resource scheduling

2. Literature Reviews

In paper P.Sangwan et al. suggested a new modification on RR algorithm. The processes have been stored in the ready queue in the same order as they arrived. The quantum time has been calculated based on the mean of the burst time of the processes. The evaluation results of this approach conducted using java language and it shows a good performance in term of waiting time and turnaround time. In a new algorithm was built for RR based on the following algorithms MRRA and SRBRR. It has been proposed to enhance resource allocation in cloud computing. It is a dynamic RR where the quantum time is changing in each round and it is equal to the sum of mean and median of processes divided by two. The experimental analysis shows the out performance of the proposed algorithm in term of the number of context switches, average waiting time and average turnaround time.[10]

In a paper Abdulrahman Abdulkarim et al. suggested that Time quantum is set to be equal to the average of burst time of requests for context switching which is declared as the variable 'round'. The first request in the ready queue would be assign a VM, there the VM is allocated to execute the request and the state of the VM is changed to Busy. After request is being executed, VM is de-allocated and the state changed back to Available. Then we have a context switch with a new burst time for the request.

In a paper Pandaba Pradhan et al. suggested that this algorithm begins with the time equals to the time of first request, which changes after the end of first request. When a new request is added into the ready queue in order to be granted, the algorithm calculates the average of sum of the times of requests found in the ready queue including the new arrival request. This needs two registers: (i)SR: To store the sum of the remaining burst time in the ready queue (ii)AR: To store avg. of the burst times by dividing the value found in the SR by the count of requests found in the ready queue. After execution, if request finishes its burst time, then it will be removed from ready queue or else it will move to the end of the ready queue. SR will be updated by subtract.[11]

In a paper D.Gritto et al. suggested that the time quantum selection in Round Robin algorithm is a big challenge. The inappropriate time slice may degrade the cloud system performance. When the time slice is larger the Round Robin algorithm behaves like First Come First Server (FCFS) algorithm, similarly for the smaller time slice the number of context switching is more. Context switching implies the method of VM switching between one cloudlet to the other cloudlet. While switching, the status of the cloudlet is stored in VM buffer in order to restore or resume back for the cloudlet execution in near future. More context switching degrades the performance. So the proposed work makes an attempt to estimate an optimal time quantum that reduces the response time, waiting time, turnaround time and the number of context switching. The time quantum is allocated using mid and average remaining execution time approximation method. The Enhanced Time Quantum Round Robin Algorithm (ETQRRA) arranges the cloudlets in ascending order as per their execution time. When the number of cloudlets in the ready queue RQ of the VM is odd then the mid approximation method is used. If the number of cloudlet in the ready queue is even then the average remaining execution time approximation method is used.[4]

El amrani Chaken et al. suggested that In case of a waiting list with three or less tasks.The proposed algorithm uses the Shorter Job First Algorithm design and allocates the time necessary for each task until its execution - In case of a waiting list with more than three tasks .If the total count of tasks on the list is a pair number o The proposed algorithm dynamically calculates a time quantum as the average of tasks burst time. If the total count of tasks on the list is an impair number o The proposed algorithm dynamically calculates a time quantum as the median of tasks burst time For each of the conditions that treats the case of a waiting list with more than three tasks, the proposed algorithm injects a test to verify if there are more than three reoccurring tasks with the same burst time. If it is the case, the time quantum is set to the value of the repetitive task with highest burst time on the waiting list[5].

In a paper Nermeen Ghazy et al. suggested that many improvements have been developed the Round Robin algorithm for task scheduling to be compatible with time sharing systems and real-time systems. The most important issue in the RR algorithm is the time quantum. The main contribution of this paper is enhancing the efficiency of the RR algorithm by proposing a Median Mean Round Robin algorithm (MMRR). It found an intelligent dynamic time quantum which is calculated as (median +mean)/2 and taking the remaining burst time of the current process into account. If the remaining burst time of the current process is less than half of the time quantum, it allocated again. Otherwise, it is placed at the end of the ready queue. Each cycle will have its TQ based on their burst time for these processes. By using a variable TQ depends on the burst time, it lead to minimize the average waiting time, turnaround time and number of context switches. The experimental results showed that the proposed algorithm enhances the performance of the system by reducing the average waiting time, turnaround time and number of context switches.[6]

V.Rhymend Uthariaraj et al. observed that Time quantum = square root of sum of the median + highest burst time .Time quantum = in first round highest burst time in ready queue. Calculate 0.8 proportion of the bust time and set it as the value of quantum time. If the process in the queue < quantum time task is allotted and completion of the task time quantum is the value of highest burst time.[8]

Maysoon A. mohammad et al. proposed a method of dynamic time quantum in which Processes are arranged in increasing order in the ready queue. New priorities and quantum times are assigned according to the CPU bursts of the processes; the process with the lowest burst time is set with the highest priority. Choose the first process in the RR fashion and assign its burst time as quantum time for this round, and allocate CPU to this process only for one quantum time. Calculate quantum times of all the processes in this round depending on their priorities and the quantum time of this cycle by using a simple formula, which is q = k + p– 1, where q is the new quantum time,k is the old quantum time and p is the priority of the processes in the ready queue.Set different quantum times for the processes according to their priorities. The highest priority process will get the largest quantum time, which is q, and the lowest priority process will get the smallest quantum time, which is k. Each process gets the control of the CPU until they finish their execution. Repeat the steps above until all processes are complete and ready queue is assigned NULL. Apply the original RR and IRRVQ and our PDQT with the priorities and new, different quantum times.[7]

In a paper suggested by Md. Sohrawordi et al. proposed a method that the proposed RRDTQ algorithm uses dynamic time quantum in CPU scheduling. It takes the integer average value of the burst time of the remaining waiting processes in the ready queue as time quantum. After each cycle, the time quantum is updated by calculating the integer average value of the burst time of the processes in the ready queue. If the ready queue contains only one waiting process, then the time quantum will be the burst time of that processes. When the execution of the processes is completed, then it will be removed from ready queue and otherwise it will be added at the tail of the ready queue with the remaining required burst time.[1]

[2]Niraj Patel, Sandip Chauhan observe and used a new approach with 5 Server for Clustering. It can be run on dedicated machine or single machine. 4 of them work as Cluster Member (S01, S02, S03, or S04).Same application (client's request) will deploy into all 4 Serve. Remaining 1 work as Load Balancer.Server (LBS).Calculate Weight of Each Server based on Processing Power, Bandwidth and Memory. For example:The firstserver can handle 100 request/sec, the second can handle 300 request/sec, the third can handle 200 request/sec and the fourth can handle only 25 request/sec.Any one Server reserve

for long completion time request and other 3 server for normal request. If long completion time request found. First check Server (S01) status, if free allocate to it.If server is busy, then check other servers status, if anyone is free then allocate to it.If all servers are busy, wait for any of the server become free.If short request found, it will first check status of S02, S03, and S04 any of these server is free then allocate to it. Otherwise check S01 server status. If these servers is free then allocate to it. If all servers are heavily loaded,then only long completion time request allocate to S01 and all normal request allocate between S02, S03, S04. Now here, step of our proposed algorithm shown as below: 1.Create Map (weight Map) and add server and according weight. 2.Create Map (load Map) to track current load of each Server.3.Create Map (avgTimeMap) to track averagge time of each request. 4.Create Map (waiting Map) to track not completed request. 5.Whenever client's requests come, first it will forward to the Load Balancer. 6.Balancer looks up in avgTimeMap for average time of Request and put it in waiting Map. 7.Check for Server status as per type of request. 8.Load Balancer redirect request to Server and increment current pointer of Server. 9. After Processing of request decrement pointer of server, update Average time of request in Map (AvgTimeMap).10).Load Balancer will take next request from Map (waitingMap) and repeat 5 through Load Balancing using DWARR Algorithm in Cloud Computing.

Paper name	Publication date	Simulation tools	Evaluation result
DWRR algorithm (niraj patel, sandeep chauhan)	2015	Cloud analyst tools	Throughput increases and resource utilizes efficiently
DTQA (Bhavin fataniya)	2018	Eclipse	less avgTAT, less avgWT
DRSS (m. satish, A. prakash)	2015	Euclyptus	average TAT and no. of context switch reduced to 20 percent
ANPDQ (Abid ur rehman khattak)	2010	Uni processor environment	drastically decrease context switching
RPJS (k.sutha)	2016		satisfy better in all parameter
NRRTSA	2023	Notebook kernel	reducing theaverage turn-around time, minimizing average waiting time, and lessening the number of contexts switching
a shared approach of dynamic loadbalancing (snehlata mishra)	2016	Cloudsim simulator	cost of each data center is more accurately than throttled and equal load sharing algorithm
m-Throttled:DLBA (Amrutanshu Panigrahi)[9]	2020	Cloudsim simulator	increased UB response time and the DC processing time
ISRBRR (P.S.Varma)	2013	C++ language	decreased the number of CS, average WT and average TAT
SRBRR (S.Saeidi)	2012	Lingo 8.0 software	The experimental shows that SRBRR achieves a good performance compared with the others
Proposed Algorithm (S.Banerjee)	2018	Cloudsim	outperformance of the proposed algorithm in term of the number of CS, average WT and average TAT
TSPBRR (S.Elmougy)	2017	Cloudsim	good performance in reducing the WT and RT compared with SJF, RR and Time Slice Priority Based RR

3. CONCLUSIONS

Cloud computing is emerging technology with lots of concerns like task congestion . It is necessary to resolve this problem with best method. After observing all the parameters and reviewing many dynamic approaches used by researchers regarding task scheduling using round robin CPU scheduling methods we concluded in this paper that a bit of changes in conventional approaches gives us better result in context with improved turnaround time, improved waiting time, decrease number of context switches, minimum throughput. There are also many approaches other than round robin to allocate tasks in in virtual machine in cloud computing. we have only concluded round robin method in this paper, in future we will observe other method with respect to all parameters.

REFERENCES

[1] Sohrawordi a. A modified round robin cpu scheduling algorithm with dynamic time quantum. International Journal of Advanced Research, 7:422–429, 02 2019.

[2] Taghreed Balharith and Fahd Alhaidari. Round robin scheduling algorithm in cpu and cloud computing: A review. In 2019 2nd International Conference on Computer Applications Information Security (ICCAIS), pages 1–7, May 2019.

[3] Dipto Biswas, Md Samsuddoha, Md. Rashid Al Asif, and Md. Manjur Ahmed. Optimized round robin scheduling algorithm using dynamic time quantum approach in cloud computing environment. International Journal of Intelligent Systems and Applications, 15:22–34, 02 2023.

[4] et. al. D.Gritto. Enhanced time quantum based round robin algorithm for cloudlet scheduling in cloud environment. 2021.

[5] Hicham Gibet Tani and Chaker El Amrani. Smarter round robin scheduling algorithm for cloud computing and big data. Journal of Data Mining Digital Humanities, Special Issue on Scientific...,01 2018.

[6] Sardar Iqbal, Hina Gull, Saqib Saeed, Madeeha Saqib, Mohammed Alq, Yasser Bamarouf, Gomathi Krishna, and May Aldossary. Relative time quantum-based enhancements in round robin scheduling. Computer Systems Science and Engineering, 41:461–477, 10 2021.

[7] Maysoon Mohammed, Mazlina Abdulmajid, and Balsam Mustafa. A proposed priority dynamic quantum time algorithm to enhance varying time quantum round robin algorithm a proposed priority dynamic quantum time algorithm 185. Int. J. Computer Applications in Technology, 54:184–191, 10 2016.

[8] Neetu and Subhash Chandra. Dynamic divisible load balancing algorithm for balancing workload in cloud computing. International Journal of Computer Applications, 178(20):42–49, Jun 2019.

[9] Amrutanshu Panigrahi, Bibhuprasad Sahu, Saroj Kumar Rout, and Amiya Kumar Rath. Mthrottled: Dynamic load balancing algorithm for cloud computing. In Debahuti Mishra, Rajkumar Buyya, Prasant Mohapatra, and Srikanta Patnaik, editors, Intelligent and Cloud Computing, pages 3–10, Singapore, 2021. Springer Singapore.

[10] Pandaba Pradhan, Prafulla Behera, and B.N.B. Ray. Modified round robin algorithm for resource allocation in cloud computing. Procedia Computer Science, 85:878–890, 12 2016.

[11] Pandaba Pradhan, Prafulla Ku. Behera, and B.N.B. Ray. Modified round robin algorithm for resource allocation in cloud computing. Procedia Computer Science, 85:878–890, 2016. International Conference on Computational Modelling and Security (CMS 2016).

Author Photo