Comparative Analysis on Scheduling Algorithms in Cloud Computing

Divyani
Department of Computer Science and Engineering
Bhilai Institute of Technology, Durg
gunjaannishu345@gmail.com

Dr. Ramesh Kumar
Department of Computer Science and Engineering
Bhilai Institute of Technology, Durg
rk_bitd@rediffmail.com

Sudip Bhattacharya
Department of Computer Science and Engineering
Bhilai Institute of Technology, Durg
202sudip@gmail.com

Abstract

Cloud computing is one of the most trending technologies in the IT industry and research. Cloud computing broadly refers to the Internet based delivery of computing resources very much similar to the delivery models of other utilities like gas, electricity and water. The pricing and billing scheme are also based on the pay-as-you-go model as practiced for other utilities. In this model the user's data is stored and processed at third party data centers. Since this model is growing in popularity day by day the load on the data centers and the servers running at the data centers is growing at a rapid rate. All the computing resources used in the data centers like the computational servers, the storage devices, networking devices etc. need to be managed for better resource utilization and optimized use of power and space at the data center establishments. About 2% of the global power is used to run cloud data centers hence efficient resource management is an essential requirement to save power .Resource management in cloud data centers poses a challenge because of the fluctuating arriving workloads. Resource management is primarily focused on scheduling incoming user jobs to Virtual machines to efficiently utilize all the resources like CPU, memory, network etc. Good scheduling algorithms should also have the feature of load balancing in order to keep the response time low. This paper discusses the various scheduling algorithms and investigates on their positive and negative aspects

Keywords: *Resource management, cloud computing, scheduling, load balancing.*

1. Introduction

One of the primary reasons for the emergence and popularity of cloud computing is resource sharing and optimal usage of computing infrastructure at datacenters. Computing resources provisioned through the cloud model are shared to clients and reallocated per request. Since cloud computing comes with the features of pay-per-use, scalability as per need and quick deployment it is now one of the most adopted models by organizations for hoisting their applications and for enterprise level storage. Cloud computing has proved to be a worthy option for start -up organizations where capacity planning can't be done completely and the capital expenditure is low. Because of these trends in the industry in favor of the cloud model there is a growing number of datacenters globally.

Data centers are the power houses of Cloud Computing. With the growing acceptance of cloud computing model the number of datacenters around the globe is also growing. Like many other technologies, data centre based Cloud computing also comes has a negative facet. Data centers consume lot of power to host the thousands of servers, networking devices and the cooling arrangements like Air conditioners etc. In general, the information technology (IT) sector nowadays consumes approximately 7% of the global

electricity, and it is forecasted that the share will rise up to 13% by 2030. Real-time video streaming, online gaming as well as mobile devices already account for 60% of all data traffic, and it is predicted that this will rise to 80% by 2020. The data centre (DC) sector in particular is estimated to account for 1.4% of the global electricity consumption (1.1–1.5% for 2011specifically , and the compound annual growth rate (CAGR) of this consumption in the period between 2007 and 2012 has been estimated as 4.4%, much higher that the projected 2.1% increase in global demand from 2012 to 2040. Apart from large power consumption data centers are also responsible for generating large amount of Carbon footprints. Since data center power consumption is an important concern for the global environment it is very important that we focus on reducing the number of data centers without compromising on the workload. This can be done by designing effective scheduling algorithms to map user jobs to servers so that the computing infrastructure is used optimally and thus reduce the need of additional servers with increasing work load. While working with job scheduling in cloud data centers one need to consider the varying nature of workloads encountered at a data center. Typically the ratio of normal and peak workloads at a enterprise level data center can be as about 1:10, which implies that peak loads can about 10 times as a normal day load. This further implies that on normal business days about 85-90% of server infrastructure goes unutilized.

Scheduling is the mapping of incoming jobs to VMs running on different physical hosts. As the number of cloud users in increasing as compared to the existing resources and the nature of jobs vary in terms of their resource needs the issue of load balancing comes into existence. Load balancing is incorporated as desirable feature in the scheduling algorithms that distributes the workload among all the available resources. There are various issues and tradeoffs that need to balance while considering load balancing in cloud. Some algorithms aim to achieve higher throughput while others aim to achieve maximum resource utilization and so on. When we go through the traffic analyzer over different geographical locations then load balancing algorithm plays a very important role to analyze the traffic flow in real time scenario over different geographical regions and then balance the overall workload. Different regions over the globe have different peak hours during which the network load is supposed to be at its peak.

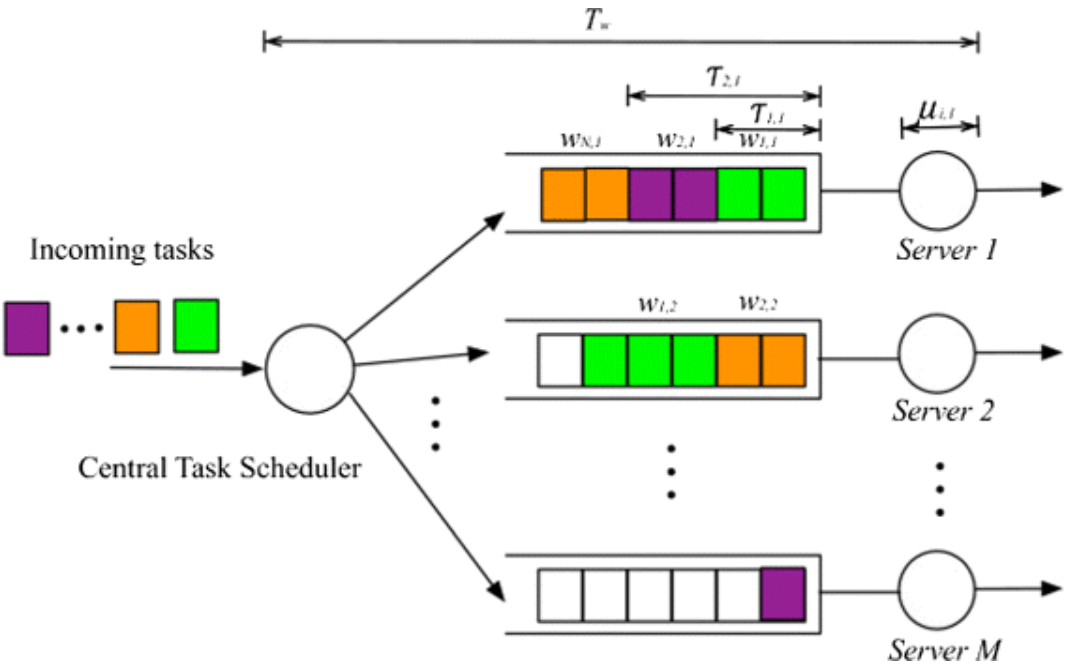

Fig 1: Mapping of various incoming jobs to different servers running as VMs on physical hosts.

The figure also shows the some unevenly distributed loads to different VMs. The job queue has been shown adjacent to each VM. The central task scheduler decides the job – VM mapping based on the scheduling policy and algorithm used Job Scheduling algorithms for cloud data centers can be classified into different categories depending upon the criteria used for scheduling. The various broad

Classifications are listed in the table below

Criteria	Type-1	Type-2
Job interdependency	**Independent** Scheduling defines that independency between tasks. Task are not dependent to each other they are totally independent	**Workflow** Work flow scheduling defines that the task which are executed, are dependent to each other.
Time of selection for scheduling algorithm	**Static Scheduling** These algorithms do not depend upon the current state of the system and have prior knowledge regarding system resources and details of all tasks in an application. These kinds of algorithms face a major drawback in case of sudden failure of system resource and tasks.	**Dynamic Scheduling** These algorithms take decisions concerning load balancing based upon the current state of the system and don't need any prior knowledge about the system. The algorithms in this category are considered complex, but have better fault tolerance and overall performance
Resource Preemption	**Preemptive** Job can be interrupted during execution. If higher priority job comes during the execution of job which has low priority then shifted to another resource.	**Non preemptive** In this scheduling job is not interrupted during execution .resources aren't being allowed or reallocated until running and scheduling job finished its execution
Response to Job	**Immediate** In this type of scheduling jobs are executed as they come, they executed one by one	**Batch mode** In this type of scheduling collecting no. of jobs and then make a schedule to schedule them
Scheduling control	**Centralized** In centralized scheduling there is one scheduler which is managing all the jobs and in this scheduling scheduler has more control on the resources and scheduler has more knowledge of the system. It has lack of scalability .	**Distributed** In this multiple schedulers who performs tasks in a co-operated manner It is less efficient but more scalable

Load balancing refers to efficiently distributing incoming network traffic across a group of backend servers, also known as a *server farm* or *server pool*. Load balancing helps enterprises

achieve high performance A load balancer acts as the router which routes the client requests across all servers capable of fulfilling those requests in a manner that maximizes speed and capacity utilization and ensures that no one server is overworked, which could degrade performance. If any server fails, the load balancer redirects traffic to the remaining online servers. When a new server is added to the server group, the load balancer automatically starts to send requests to it.

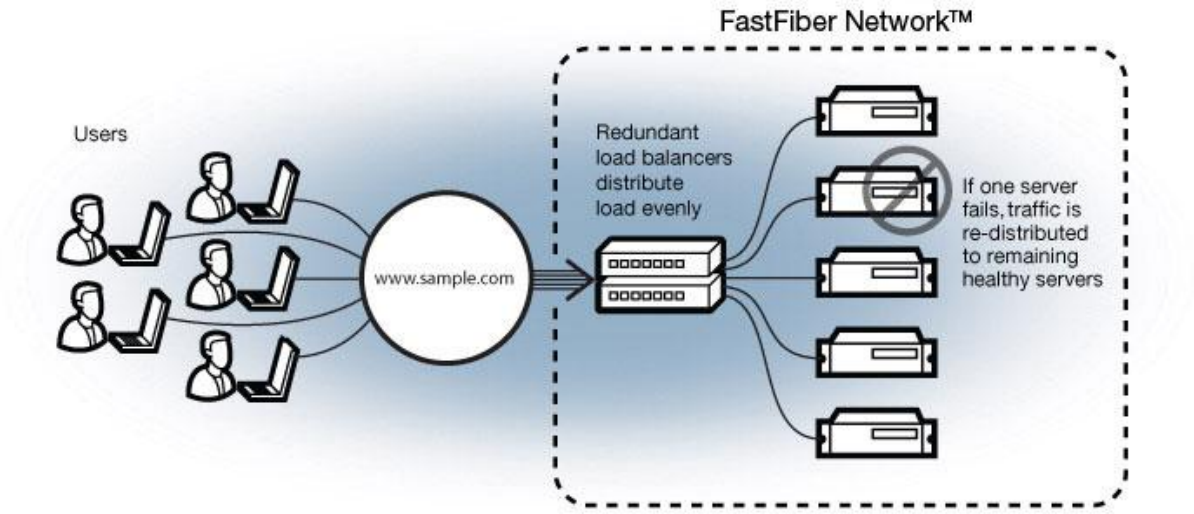

Fig 2: Load balancing in cloud Computing. Load Balancers also help to redistribute the load of a failed node.

2. Literature Survey

1. Mayank Mishra(2012)
   In this paper Mayank Mishra discusses the advantages of pay-as-use model, wherein users pay only when they want to use resources. The paper also points out the variations between peak and non peak period.[1].

2. Venkatesa Kumar and S Palaniswami
   In their paper have proposed the overall utilization of resource and processing cost and presents scheduling approaches which focuses on both high and low priority scheduling[2].

3. Liang Luo et al
   In their paper have discussed about CloudSim toolkit and VM load balancing algorithm, where CloudSim toolkit is a tool for simulation purpose. Various VMs are used and have different processing power. Task, request and application service are assigned to VM. The paper discusses the optimization of the performance parameters such as response time and data processing time, giving an efficient VM Load Balancing algorithm i.e. Weighted Active Load Balancing Algorithm in the Cloud Computing environment[3].

4. Lu huang, hai-shan
   In their paper discusses a scheduling schema that provides good performance and fairness simultaneously in heterogeneous cluster and make resource request in a cost effective manner[4].

5 Kaur and Kinger (2014)
In this paper the authors have raised issues in cloud computing resource management like heavy load balancing and traffic while computation and also give solution which job scheduling algorithm. The authors proposed various scheduling algorithm like FCFS, round robin, priority algorithm and concluded that none of the algorithms are best scheduling in each algorithm has some drawbacks like when we talk about priority

algorithm only higher priority algorithm comes first to execute and so on. so there should be a new algorithm which achieves all objective and provide better performance [5].

6 "Work flow scheduling algorithm in cloud environment" author (Arya and Verma 2014)
In this paper the concept of workflow scheduling algorithm has been discussed. A workflow algorithm defines the parent child relationship. The parent task should be executed before its child task. In WSA the parameters like reliability, load balancing, fault tolerance can be enhanced [6].

7 In the paper Comparative Study on Load Balancing Techniques in Cloud Computing, N. S. Raghava and Deepti Singh (2014) has been presented. The work presents a head to head comparison between various scheduling techniques based on different algorithm respect to different performance parameter [7].

8 Efficient Resource Management for Cloud Computing Environments. In this paper the authors tell about power aware scheduling techniques and discusses about DVFS techniques and some current technology power aware and thermal aware scheduling to maximize energy saving . (using Vms)[8].

9 Efficient optimal algorithm of task scheduling in cloud computing environment Amit Agarwal, Saloni jain(2014) In this paper there are various scheduling algorithm and every algorithm has been observed to have its own characteristics, applicability, pros & cons but none of the algorithms take care of reliability and performance[9].

10. Genetic-based task scheduling algorithm in cloud computing environment (2016).
In this paper a genetic algorithm has been used for minimizing the cost and completion time and maximizing the resource utilization. The CloudSim toolkit has been used for the simulations [10]

11 Efficient load balancing in cloud computing using fuzzy logic Srinivas Sethi1, Anupama Sahu, Suvendu Kumar Jena (2012)
In this paper various techniques have been used for load balancing. The RR and FRR algorithms have been compared. Load balancing algorithm FRR have been found to be more efficient than RR(round robin algorithm)[11].

12. A survey of various scheduling algorithm in cloud computing environment Pinal Salot
This paper presents various scheduling algorithms and a comparison among them. The paper mentions that there is tradeoff between throughput and cost effectiveness. Also other issues are reliability and fault tolerance. The need for an algorithm which improve availability and reliability in cloud computing environment has been stressed[12].

13. Analysis of Different Variants in Round Robin Algorithms for Load Balancing in Cloud Computing SubasishMohapatra ,Subhadarshini Mohanty ,K.Smruti Rekha(2013)
In this paper the authors tell about various scheduling algorithm along with different variants of the Round Robin Algorithm. The finding is that Round Robin gives the best integrated performance as compare to other algorithms [13].

14. Utilizing Round Robin Concept for Load Balancing Algorithm at Virtual Machine Level in Cloud Environment Stuti Dave, Prashant Maheta (2014)
This paper discusses about the RR algorithm and concludes that RR provides simplest solution for load balancing in cloud environment[14].

15. Scheduling Virtual Machines for Load balancing in Cloud Computing Platform Supreeth S 1, Shobha Biradar
This work defines algorithms RR and WRR (Weighted Round Robin) and presents a comparison between them. The authors have found that WRR is much better than RR and it takes less time for VM scheduling and it also improves the performance [15]

16. Efficient Resource Utilization Algorithm (ERUA) for Service Request Scheduling in Cloud Ram Kumar N, Nivethitha
This paper defines the ERUA scheduling algorithm which is based on 3 tier user, service provider and resource provider. This scheduling algorithm utilizes the resources and gives benefit to both user and service provider and provide better QoS [16]

17. An Optimal Model for Priority based Service Scheduling Policy for Cloud Computing Environment Dr. M. Dakshayini, Dr. H. S. Gurupra Prasad
The work describes a priority and admission control based scheduling and compares between TSC and PSC. TSC doesn't consider admission control and priority whereas PSC considers admission control and the results found that PSC is better than TSC [17]

18 Liu,ZheQiu A Survey on Virtual Machine Scheduling in Cloud Computing, In this papers the authors analyze the VM scheduling technology in cloud computing, which includes model of scheduling, characteristics of scheduling and some general scheduling algorithms. The schedulers have been classified from five aspects. Existing scheduling algorithms give high energy and cost effective but they seldom consider reliability and availability. So there is a need for algorithms that improve availability and reliability in cloud computing environment.[18]

After going through the literature available regarding comparative study of various scheduling algorithms for cloud computing we draw an outline of the major advantages and shortcomings of the popular algorithms. These findings have been aggregated from various papers.

Comparison of algorithms

Table1. Comparison of different scheduling algorithms

Algorithm	Parameters	Objective / Allocation	Waiting time
PSJN	Cost and Time	Effective and Fast execution	Lesser
Shortest Job Scheduling	Arrival time, process time, deadline and IO requirement	Effective resource allocation under defined parameter	Lesser
Optimized algorithm	Cost, profit and priority	Measure the cost and performance more accurately	More
Cost based algorithm	Cost and task grouping	Minimizing the cost and completion time	Lesser
Min Min Algorithm	Makespan	Promised the guarantee regarded the provided resources	Lesser
Ant algorithm	Pheromone updating rule	Enhance the performance of basic ACO	More
ACO	Cost and time	Improve the efficiency and reliability in allconditions	More
FCFS	Simplest scheduling Algorithm	CPU is allocated in the order in which the process arrive	More
SJF	Difficult to understand	CPU is allocated to the process with least CPU burst time	Lesser than FCFS
Priority algorithm	Difficult to understand	Based on priority, the higher priority job can run first	Lesser
RR algorithm	Performance heavy depends	The preemption take place after a fixed	More than all

	upon the size of time quantum	interval of time	
Genetic algorithm	Complexity depends on the task to be scheduled	There is a greedy algorithm and pick the best job to allocate the CPU	Less

The above table shows the number of the scheduling algorithms and comparison between them on the basis of complexity, allocation, waiting time and type of system. Complexity defines which type of algorithm is simple or easy to use in processing. Allocation defines how the jobs are assigned to the resources. Waiting Time defines which of the algorithm takes more time for processing. Type of System defines which algorithm is suitable for which type of system. The First Come First Serve algorithm is a simplest algorithm for scheduling but waiting time to process the tasks is much more in this. This algorithm is suitable for batch type of systems. In Shortest Job First algorithm the waiting time is less than FCFS. This algorithms process the task first having least CPU burst time. Priority algorithm is difficult to understand because how priority can be assigned to the task is a difficult task. Here waiting time is less because task with higher priority processed first. This algorithm is suitable for both batch and time sharing systems. In RR algorithm waiting time is more than all because after a fixed time interval the next task will execute. So problem faced when one task is very heavy and other one is with very simple and small calculations. Genetic algorithm is a bio inspired artificial intelligent scheme and its complexity depends on the task. The best selected task executes first so waiting time is less here[19].

3. Problem identification

A numerous algorithms have been proposed for job scheduling in cloud computing. Each of the algorithm have its own set of advantages and limitations. For instance the drawbacks of two of the most common algorithms are listed below.

FCFS –It is a non preemptive scheduling algorithm, the request which comes first is served first. The algorithm is simple and easy to implement but it comes with an inherent drawback.

- The waiting time can be large for any job because of lot of jobs being in the queue

Round robin - In this algorithm the concept of time slicing is used and it gives fix amount of time to the jobs which is called as time quantum. If we consider the positive side of algorithm it utilizes all resources in balanced order and equal number of VMs are allocated to all node which ensures fairness but it has some drawbacks too.

- It selects the load on random basis and leads to situation where some nodes are heavily loaded and some are lightly loaded.

- The 2nd drawback is power consumption is high as many node will be kept turned on for a long time.

The literature found presents comparative studies but the different authors have used different parameters for the comparison. In this work our primary objective was to compare the different algorithms and their variations and compare the algorithms on same parameters.

4. Methodology

For simulation of the different cloud scheduling algorithms we used the CloudSim simulation toolkit. CloudSim is a Java based command line tool that can be used to simulate the activities of a data center like VM allotment to incoming user tasks through different service brokers. It has been developed and is distributed by the Cloud computing and Distributed Systems (CLOUDS) laboratory , University of Melbourne. CloudSim provides the following functionalities

- support for modeling and simulation of large scale Cloud computing data centers
- support for modeling and simulation of virtualized server hosts, with customizable policies for provisioning host resources to virtual machines
- support for modeling and simulation of application containers
- support for modeling and simulation of energy-aware computational resources
- support for modeling and simulation of data center network topologies and message-passing applications
- support for modeling and simulation of federated clouds
- support for dynamic insertion of simulation elements, stop and resume of simulation
- support for user-defined policies for allocation of hosts to virtual machines and policies for allocation of host resources to virtual machines[20].

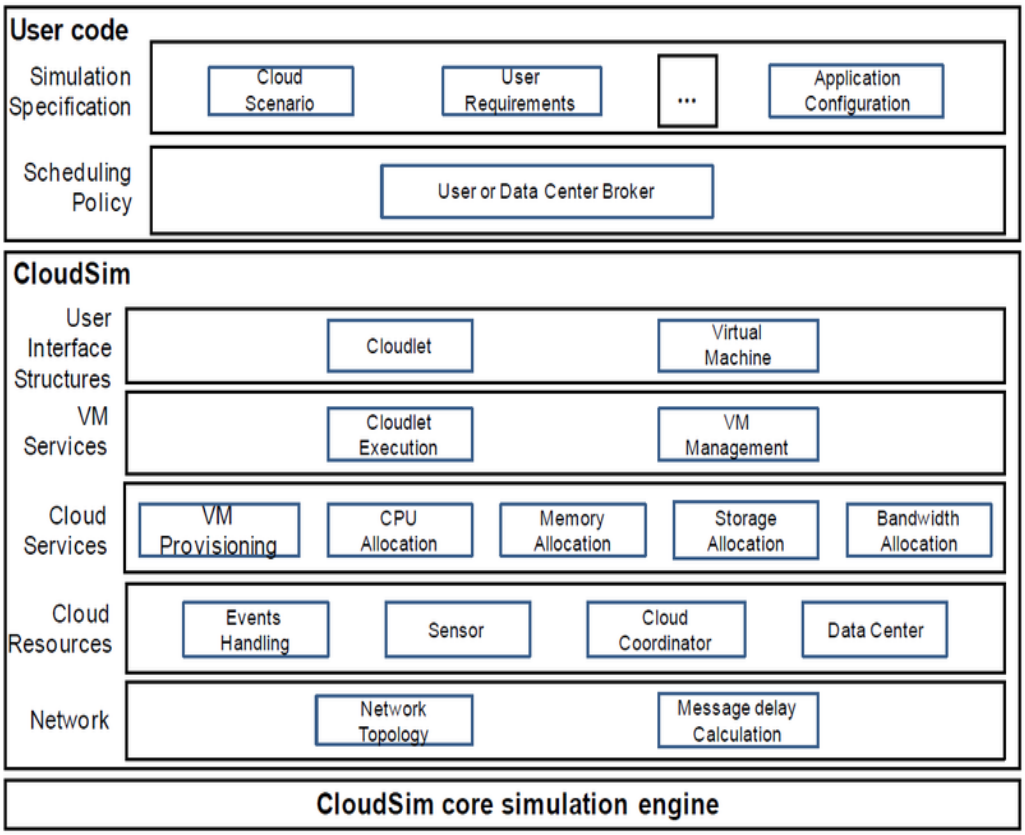

Fig 3: CloudSim architecture. User code runs on top of the Cloud simulation Engine which has the cloud infrastructure and the service details.

There is GUI based tool called Cloud Reports that works on top of CloudSim and presents an easy to use GUI for users to quickly configure the parameters for simulation..

5. Results and Discussions

We have started our experiments on the CloudSim tool. The initial experiments include getting the experimental results for simulations for the variations of Round Robin and FCFS algorithms. At the next level we will move to other algorithms like preemptive priority and min-max. Next we propose to use a novel algorithm that can be adaptive to incoming workload patterns and can balance the tradeoff between performance, cost and efficiency. At the time of writing this paper we have got initial results from the CloudSim tool.

6. Conclusion

With the growing adoption of the cloud computing more and more users are now moving to cloud for running their tasks. This has led to a massive growth in the number of data centers around the globe. As data centers are power hungry and produce a large carbon footprints it's a high concern for all of us to use the resources at data centers to optimum levels to check the growing demand for new data centers. For this we need to focus on our job scheduling and load balancing algorithms used at data centers to make them more and more efficient. The present work examined the various scheduling algorithms and considered their pitfalls with respect to performance and cost parameters. We concluded that no algorithm clearly dominates the scheduling space and we need to device novel algorithm for handling next generation cloud workloads

7. References

[1] Mayank Mishra, Anwesha Das, Purushottam Kulkarni, and Anirudha Sahoo, "Dynamic Resource Management Using Virtual Machine Migrations", Sep 2012, 0163-6804/12, IEEE Communications Magazine, page no: 34-40

[2] Venkatesa Kumar. V and S. Palaniswami," A Dynamic Resource Allocation Method for Parallel Data Processing in Cloud Computing", 2012, Journal of Computer Science 8 (5), ISSN 1549-3636, Science Publications, page no: 780-788.

[3]Liang Luo, Wenjun Wu, Dichen Di, Fei Zhang, Yizhou Yan, Yaokuan Mao, "A Resource Scheduling Algorithm of Cloud Computing based on Energy Efficient Optimization Methods", 2012, 978-1-4673-2154-9/12, IEEE.

[4] Lu Huang, Hai-shan Chen, Ting-ting Hu, "Survey on Resource Allocation Policy and Job Scheduling Algorithms of Cloud Computing", JOURNAL OF SOFTWARE, VOL. 8, NO. 2, FEBRUARY 2013, ACADEMY PUBLISHER,doi:10.4304/jsw.8.2.480-487, Page no: 480-487.

[5]Kaur, R., & Kinger, S. "Analysis of Job Scheduling Algorithms in Cloud Computing" International Journal of Computer Trends and Technology, *9*, 379-386.(2014)

[6] Author Arya and Verma" work flow scheduling algorithm in cloud environment" Proceedings of 2014 RAECS UIET Panjab University Chandigarh, 06 – 08 March, 2014

[7] N. S. Raghava* and DeeptiSingh "Comparative Study on Load Balancing Techniques in Cloud Computing" OPEN JOURNAL OF MOBILE COMPUTING AND CLOUD COMPUTING Volume 1, Number 1, August 2014

[8], Jeffrey M. Galloway, Karl L. Smith, Susan S. Vrbsky "Power Aware Load Balancing for Cloud Computing" Proceedings of the World Congress on Engineering and Computer Science 2011 Vol I WCECS 2011, October 19-21, 2011, San Francisco, USA

[9] Dr Amit Agarwal, Saloni jai "Efficient optimal algorithm of task scheduling in cloud computing environment" (2014) International Journal of Computer Trends and Technology (IJCTT) – volume 9 number 7– Mar 2014

[10] Safwat A. Hamad, Fatma A. Omara "Genetic-based task scheduling algorithm in cloud computing environment (2016)." (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 4, 2016

[11] , Anupama Sahu, Suvendu Kumar Jena Srinivas Sethi1 "Efficient load balancing in cloud computing using fuzzy logic " (2012 IOSR Journal of Engineering (IOSRJEN) ISSN: 2250-3021 Volume 2, Issue 7(July 2012), PP 65-71 www.iosrjen.org

[12] Pinal Salot "survey of various scheduling algorithm in cloud computing environment" IJRET: International Journal of Research in Engineering and Technology ISSN: 2319-1163 2013

[13] Computing Subasish Mohapatra ,Subhadarshini Mohanty ,K.Smruti Rekha" Analysis of Different Variants in Round Robin Algorithms for Load Balancing in Cloud" (2013 , International Journal of Computer Applications (0975 – 8887) Volume 69– No.22, May 2013

[14] Stuti Dave,Prashant Maheta" Utilizing Round Robin Concept for Load Balancing Algorithm at Virtual Machine Level in Cloud Environment "International Journal of Computer Applications (0975 – 8887) Volume 94 – No 4, May 2014 (2014).

[15] Platform Supreeth 1, 2013Shobha Bir adar **"**Scheduling Virtual Machines for Load balancing in Cloud Computing "International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064 .

[16] Ramkumar N, Nivethitha ,Ramkumar)"Efficient Resource Utilization Algorithm (ERUA) for Service Request Scheduling in Cloud" ISSN 2013 N et al. / International Journal of Engineering and Technology (IJET) ISSN Ramkumar N, Nivethitha ,Ramkumar N et al. / International Journal of Engineering and Technology (IJET)

[17] Dr. M. Dakshayini,Dr. H. S. Gurupra sad "An Optimal Model for Priority based Service Scheduling Policy for Cloud Computing Environment" International Journal of Computer Applications (0975 – *8887) Volume 32– No.9, October 2011*

[18] Liu,ZheQiu Z" A Survey on Virtual Machine Scheduling in Cloud Computing, "Scheduling 2016 2nd IEEE International Conference on Computer and Communications.

[19] Shameer A.P Dr. A C Subhajini "International Journal of Advanced Research in Computer Science and Software Engineering" Volume 6, Issue 5, May 2016 ISSN: 2277 128X

[20]CloudSim The CLOUDSLab:FlagshipProjects-Gridbusand www.cloudbus.org/cloudsim

.