A review of different scheduling algorithm in cloud Computing architecture

Avinash Verma[#1], Saurabh Sharma[#2]
[#]*Research Scholar, Computer Technology & Applications*
**Professor, Deptt. of Computer Science & Engg.*
Gyan Ganga College of Technology
Jabalpur (M.P.), India
[1]avinash_21nov@yahoo.com
[2]saurabhsharma@ggct.co.in

Abstract: **Load Balancing is essential for efficient operations in distributed environments. As Cloud Computing is growing rapidly and clients are demanding more services and better results, load balancing for the Cloud has become a very interesting and important research area. Many algorithms were suggested to provide efficient mechanisms and algorithms for assigning the client's requests to available Cloud nodes. These approaches aim to enhance the overall performance of the Cloud and provide the user more satisfying and efficient services. In this paper, we investigate the different algorithms proposed to resolve the issue of load balancing and task scheduling in Cloud Computing. We discuss and compare these algorithms to provide an overview of the latest approaches in the field.**

Index Terms: **cloud computing, scheduling, algorithm.**

1. INTRODUCTION

Cloud Computing became very popular in the last few years. As part of its services, it provides a flexible and easy way to keep and retrieve data and files. Especially for making large data sets and files available for the spreading number of users around the world. Handling such large data sets require several techniques to optimize and streamline operations and provide satisfactory levels of performance for the users. Therefore, it is important to research some areas in the Cloud to improve the storage utilization and the download performance for the users. One important issue associated with this field is dynamic load balancing or task scheduling. Load balancing algorithms were investigated heavily in various environments; however, with Cloud environments, some additional challenges are present and must be addressed. In Cloud Computing the main concerns involve efficiently assigning tasks to the Cloud nodes such that the effort and request processing is done as efficiently as possible [1], while being able to tolerate the various affecting constraints such as heterogeneity and high communication delays. Load balancing algorithms are classified as static and dynamic algorithms. Static algorithms are mostly suitable for homogeneous and stable environments and can produce very good results in these environments. However, they are usually not flexible By On-line mode heuristic scheduling algorithm, Jobs are scheduled when they arrive in the system. Since the cloud environment is a heterogeneous system and the speed of each processor varies quickly, the on-line mode heuristic scheduling algorithms are more appropriate for a cloud environment. Most fit task scheduling algorithm (MFTF) is suitable example of On-line mode heuristic scheduling algorithm.

and cannot match the dynamic changes to the attributes during the execution time. Dynamic algorithms are
Job scheduling is one of the major activities performed in all the computing environments. Cloud computing is one the upcoming latest technology which is developing drastically. To efficiently increase the working of cloud computing environments, job scheduling is one the tasks performed in order to gain maximum profit.

The goal of scheduling algorithms in distributed systems is spreading the load on processors and maximizing their utilization while minimizing the total task execution time Job scheduling, one of the most famous optimization problems, plays a key role to improve flexible and reliable systems. The main purpose is to schedule jobs to the adaptable resources in accordance with adaptable time, which involves finding out a proper sequence in which jobs can be executed under transaction logic constraints. There are main two categories of scheduling algorithm.1) Static scheduling algorithm and 2) Dynamic scheduling algorithm. Both have their own advantage and limitation. Dynamic scheduling algorithms have higher performance than static algorithm but have a lot of overhead compare to it.

2. SCHEDULING

There has been various types of scheduling algorithm exist in distributed computing system. Most of them can be applied in the cloud environment with suitable verifications. The main advantage of job scheduling algorithm is to achieve a high performance computing and the best system throughput. Traditional job scheduling algorithms are not able to provide scheduling in the cloud environments. According to a simple classification, job scheduling algorithms in cloud computing can be categorized into two main groups; Batch mode heuristic scheduling algorithms (BMHA) and online mode heuristic algorithms. In BMHA, Jobs are queued and collected into a set when they arrive in the system. The scheduling algorithm will start after a fixed period of time. The main examples of BMHA based algorithms are; First Come First Served scheduling algorithm (FCFS), Round Robin scheduling algorithm (RR), Min–Min algorithm and Max–Min algorithm.

a. First Come First Serve Algorithm:

Job in the queue which comes first is served. This algorithm is simple and fast.

b. Round Robin algorithm:

In the round robin scheduling, processes are dispatched in a FIFO manner but are given a limited amount of CPU time

called a time-slice or a quantum. If a process does not complete before its CPU-time expires, the CPU is pre-empted and given to the next process waiting in a queue. The pre-empted process is then placed at the back of the ready list.

c. Min–Min algorithm:

This algorithm chooses small tasks to be executed firstly, which in turn large task delays for long time.

d. Max – Min algorithm:

This algorithm chooses large tasks to be executed firstly, which in turn small task delays for long time.

e. Most fit task scheduling algorithm:

In this algorithm task which fit best in queue are executed first. This algorithm has high failure ratio.

f. Priority scheduling algorithm:

The basic idea is straightforward: each process is assigned a priority, and priority is allowed to run. Equal-Priority processes are scheduled in FCFS order. The shortest-Job-First (SJF) algorithm is a special case of general priority scheduling algorithm. An SJF algorithm is simply a priority algorithm where the priority is the inverse of the (predicted) next CPU burst. That is, the longer the CPU burst, the lower the priority and vice versa. Priority can be defined either internally or externally. Internally defined priorities use some measurable quantities or qualities to compute priority of a process.

2.1 Scheduling Process

Scheduling process in cloud can be generalized into three stages namely–

- Resource discovering and filtering – Datacenter Broker discovers the resources present in the network system and collects status information related to them.
- Resource selection – Target resource is selected based on certain parameters of task and resource. This is deciding stage.
- Task submission -Task is submitted to resource selected.

3. EXISTING SCHEDULING ALGORITHM

The Following scheduling algorithms are currently prevalent in clouds.

3.1 *Resource-Aware-Scheduling algorithm (RASA):* Saeed Parsa and Reza Entezari-Maleki [2] proposed a new task scheduling algorithm RASA. It is composed of two traditional scheduling algorithms; Max-min and Min-min. RASA uses the advantages of Max-min and Min-min algorithms and covers their disadvantages. Though the deadline of each task, arriving rate of the tasks, cost of the task execution on each of the resource, cost of the communication are not considered. The experimental results show that RASA is outperforms the existing scheduling algorithms in large scale distributed systems.

3.2 *RSDC (RELIABLE SCHEDULING DISTRIBUTED IN CLOUD COMPUTING):* Arash Ghorbannia Delavar,Mahdi Javanmard , Mehrdad Barzegar Shabestari and Marjan Khosravi Talebi[1] proposed a reliable scheduling algorithm in cloud computing environment. In this algorithm major job is divided to sub jobs. In order to balance the jobs the request and acknowledge time are calculated separately. The scheduling of each job is done by calculating the request and acknowledges time in the form of a shared job. So that efficiency of the system is increased.

3.3 *An Optimal Model for Priority based Service Scheduling Policy for Cloud Computing Environment:* Dr. M. Dakshayini, Dr. H. S. Guruprasad [3] proposed a new scheduling algorithm based on priority and admission control scheme. In this algorithm priority is assigned to each admitted queue. Admission of each queue is decided by calculating tolerable delay and service cost. Advantage of this algorithm is that this policy with the proposed cloud architecture has achieved very high (99%) service completion rate with guaranteed QoS. As this policy provides the highest precedence for highly paid user service-requests, overall servicing cost for the cloud also increases.

3.4 *A Priority based Job Scheduling Algorithm in Cloud Computing:* Shamsollah Ghanbari, Mohamed Othman proposed a new scheduling algorithm based on multi – criteria and multi - decision priority driven scheduling algorithm. This scheduling algorithm consist of three level of scheduling: object level, attribute level and alternate level. In this algorithm priority can be set by job resource ratio. Then priority vector can be compared with each queue. This algorithm has higher throughput and less finish time.

Extended Max-Min Scheduling Using Petri Net and Load Balancing: El-Sayed T. El-kenawy, Ali Ibraheem El-Desoky, Mohamed F. Al-rahamawy[5] has proposed a new algorithm based on impact of RASA algorithm. Improved Max-min algorithm is based on the expected execution time instead of complete time as a selection basis. Petri nets are used to model the concurrent behavior of distributed systems. Max-min demonstrates achieving schedules with comparable lower makespan rather than RASA and original Max-min.

3.5 *An Optimistic Differentiated Job Scheduling System for Cloud Computing:* Shalmali Ambike, Dipti Bhansali, Jaee Kshirsagar, Juhi Bansiwal[6] has proposed a differentiated scheduling algorithm with non-preemptive priority queuing model for activities performed by cloud user in the cloud computing environment. In this approach one web application is created to do some activity like one of the file uploading and downloading then there is need of efficient job scheduling algorithm. The Qos requirements of the cloud computing user and the maximum profits of the cloud computing service provider are achieved with this algorithm.

3.6 Improved Cost-Based Algorithm for Task Scheduling: Mrs.S.Selvarani, Dr.G.Sudha Sadhasivam [7] proposed an improved cost-based scheduling algorithm for making efficient mapping of tasks to available resources in cloud. The improvisation of traditional activity based costing is proposed by new task scheduling strategy for cloud environment where there may be no relation between the overhead application base and the way that different tasks cause overhead cost of resources in cloud. This scheduling algorithm divides all users Tasks depending on priority of each task into three different lists. This scheduling algorithm measures both resource cost and computation performance, it also Improves the computation/communication ratio.

3.7 *Performance and Cost evaluation of Gang Scheduling in a Cloud Computing System* with Job Migrations and Starvation Handling: Ioannis A. Moschakis and Helen D. Karatza has proposed a gang scheduling algorithm with job migration and starvation handling in which scheduling parallel jobs, already applied in the areas of Grid and Cluster computing. The number of Virtual Machines (VMs) available at any moment is dynamic and scales according to the demands of the jobs being serviced. The aforementioned model is studied through simulation in order to analyze the performance and overall cost of Gang Scheduling with migrations and starvation handling. Results highlight that this scheduling strategy can be effectively deployed on Clouds, and that cloud platforms can be viable for HPC or high performance enterprise applications.

4. COMPARISON

In this section we compare the existing algorithms on the basis of different parameters as shown in table 1.

Algorithms	Complexity	Allocation	Waiting Time	Type of system
FCFS Algorithm	Simplest Scheduling Algorithm	CPU is allocated in the order in which the processes arrive	More	Suitable for Batch system.
SJF Algorithm	Difficult to understand and code	CPU is allocated to the process with least CPU burst time	Lesser than FCFS	Suitable for Batch system
Priority Algorithm	Difficult to understand	Based on priority, So the higher priority job can run first.	Lesser	Suitable for both Batch and time sharing systems
Round Robin Algorithm	Performance heavily depends upon the size of time quantum	The preemption take place after a fixed interval of time	More than all	Suitable for time sharing system
Genetic Algorithm	Complexity depends on the task to be scheduled	This is a greedy algorithm and pick the best job to allocate the CPU	Waiting time is less	It deals with problem where the search space is large

5. CONCLUSION

Scheduling is one of the most important tasks in cloud computing environment. In this paper we have analyze various scheduling algorithm and tabulated various parameter. We have noticed that disk space management is critical issue in virtual environment. Existing scheduling algorithm gives high throughput and cost effective but they do not consider reliability and availability. So we need algorithm that improves availability and reliability in cloud computing environment.

REFERENCES

1. Arash Ghorbannia Delavar,Mahdi Javanmard , Mehrdad Barzegar Shabestari and Marjan Khosravi Talebi "RSDC (RELIABLE SCHEDULING DISTRIBUTED IN CLOUD COMPUTING)" in International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.2, No.3, June 2012

2. Saeed Parsa and Reza Entezari-Maleki," RASA: A New Task Scheduling Algorithm in Grid Environment" in World Applied Sciences Journal 7 (Special Issue of Computer & IT): 152-160, 2009.Berry M. W., Dumais S. T., O'Brien G. W. Using linear algebra for intelligent information retrieval, SIAM Review, 1995, 37, pp. 573-595.

3. Dr. M. Dakshayini, Dr. H. S. Guruprasad "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.

4. Shamsollah Ghanbari, Mohamed Othman "A Priority based Job Scheduling Algorithm in Cloud Computing"

5. El-Sayed T. El-kenawy, Ali Ibraheem El-Desoky Min Min Algorithms.

6. Shalmali Ambike, Dipti Bhansali, Jaee Kshirsagar, Juhi in Round Robin Algorithms.

7. Mrs.S.Selvarani1; Dr.G.Sudha Sadhasivam, improved cost based scheduling algorithms.

BIOGRAPHIE

Mr. Avinash Verma, He is Research Scholar and perusing Master of Technology from Computer Technology and Application from Gyan Ganga College of Technology, Jabalpur (Madhya Pradesh), His area of interest in Load Balancing in Cloud Computing Environment and doing his Research Work under the Guidance of **Prof. Saurabh Sharma**(Computer Science Department).

