A Comparative Study of Various Scheduling Algorithms in Cloud Computing

Anil Kumar Mishra[1], Jyortirmayee Sarangi[2]
[1] *Computer Science Engineering, Gandhi Engineering College(GEC, Bhubaneswar)*
[2] *Electronics & Communication Engineering, Gandhi Engineering College(GEC, Bhubaneswar)*

Abstract Cloud Computing is known for providing services to variety of users by with the aid of very large scalable and virtualized resources over the internet. Due to the recent innovative trends in this field, a number of scheduling algorithms have been developed in cloud computing which intend to decrease the cost of the services provided by the service provider in cloud computing environment. Most of the modern day researchers, attempt to construct job scheduling algorithms to increase the availability and performance of cloud services as the users have to pay for the available resources/services based on time. Considering all the above factors, scheduling plays a crucial role to maximize the utilization of resources in cloud computing environment. Through this paper, we are doing a comparative study of various scheduling algorithms and the related issues in cloud computing.

Keywords *Cloud Computing, Scheduling, Algorithm*

1. Introduction

Cloud computing is known as a provider of dynamic services using very large scalable and virtualized resources over the Internet. Various definitions and interpretations of "clouds" and / or "cloud computing" exist. With particular respect to the various usage scopes the term is employed to, we will try to give a representative (as opposed to complete) set of definitions as recommendation towards future usage in the cloud computing related research space. We try to capture an abstract term in a way that best represents the technological aspects and issues related to it. In its broadest form, we can define a 'cloud' is an elastic execution environment of resources involving multiple stakeholders and providing a metered service at multiple granularities for a specified level of quality of service. To be more specific, a cloud is a platform or infrastructure that enables execution of code (services, applications etc.), in a managed and elastic fashion, whereas "managed" means that reliability according to pre defined quality parameters is automatically ensured and "elastic" implies that the resources are put to use according to actual current requirements observing overarching requirement definitions – implicitly, elasticity includes both up- and downward scalability of resources and data, but also load-balancing of data throughput.

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 algorithm has higher performance than static algorithm but has 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.

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 is suitable example of On-line mode heuristic scheduling 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 preempted and given to the next process waiting in a queue. The preempted process is then placed at the back of the ready list.

c. Min–Min Algorithm:

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

d. Max–Min Algorithm:

This algorithm chooses large tasks to be executed first, which in turn delays small tasks 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.

Scheduling Process

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

- Resource discovering and filtering: Data center 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.
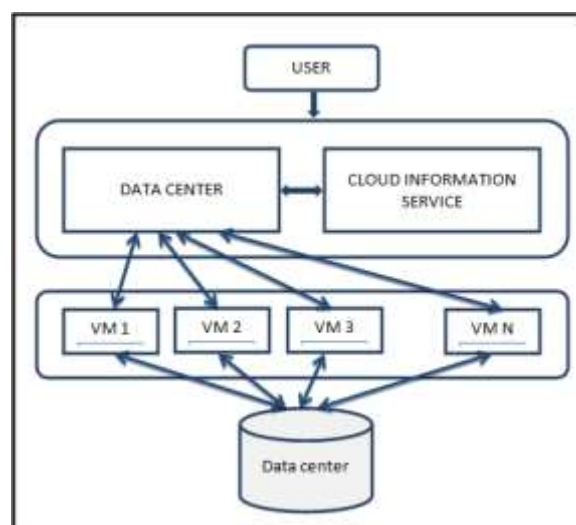- Task submission: Task is submitted to resource selected.

Figure 1. Scheduling Process

3. Existing Scheduling Algorithms

The following scheduling algorithms are currently prevalent in clouds.

Resource-Aware-Scheduling Algorithm (RASA)

Saeed Parsa and Reza Entezari-Maleki [1] 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.

An Optimal Model for Priority based Service Scheduling Policy for Cloud Computing Environment

Dr. M. Dakshayini, Dr. H. S. Guruprasad [2] 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.

Extended Max-Min Scheduling Using Petri Net and Load Balancing

El-Sayed T. El-kenawy, Ali Ibraheem El-Desoky, Mohamed F. Al-rahamawy [3] 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 behaviour of distributed systems. Max-min demonstrates achieving schedules with comparable lower makespan rather than RASA and original Max-min.

Reliable Scheduling Distributed in Cloud computing (RSDC)

Arash Ghorbannia Delavar, Mahdi Javanmard, Mehrdad Barzegar Shabestari and Marjan Khosravi Talebi [4] 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.

Improved Cost-Based Algorithm for Task Scheduling

Mrs. S.Selvarani, Dr. G. Sudha Sadhasivam [5] 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 user 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.

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.

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 [7]. 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.

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

Sl. No.	Scheduling Algorithm	Scheduling Method	Scheduling Parameter	Scheduling Factor	Findings	Environment
1	Resource Aware Scheduling Algorithm (RASA)	Batch mode	Makespan	Grouped task	1. It is used to Reduce makespan	Grid Environment
2	An optimal model for priority based Service scheduling Policy for cloud Computing environment	Batch mode	Quality of Service, service request time	An array of workflow instances	1. High Quality of Service 2. High Throughput	Cloud Environment
3	Extended Max-Min Scheduling using Petri net and Load Balancing	Batch mode	Load balancing, finish time	Grouped task	1. It Is used for Efficient load Balancing 2. Petri net is used to remove limitation of Max-Min algorithm	Cloud Environment
4	Reliable Scheduling Distributed in Cloud computing (RSDC)	Batch mode	Processing time	Grouped task	1. It is used to reduce processing time 2. It is efficient for load balancing	Cloud Environment
5	Improved Cost-based Algorithm for Task Scheduling	Batch mode	Cost and performance	Unscheduled task group	1. Measures both resource cost and computation performance 2. Improves the computation /communication ratio	Cloud Environment
6	An Optimistic Differentiated Job Scheduling System for Cloud Computing	Dependency mode	Quality of Service, maximum profit	Single job with multiple user	The QoS requirements of the cloud computing user and the maximum profits of the cloud computing	Cloud Environment

					service provider are achieved	
7	A Priority based Job Scheduling Algorithm in Cloud Computing	Dependency mode	Priority to each queue	An array of job queue	1. Less finish time	Cloud Environment
8	Performance and Cost Evaluation of Gang Scheduling in Cloud Computing System with Job Migrations and Starvation Handling	Batch mode	Performance and cost	Workflow with large number of jobs	1. The application of migrations and starvation handling had a significant effect on the model 2. It improves performance	Cloud Environment

5. Conclusions

Scheduling is considered as one of the most important tasks in cloud computing environment. In this paper, we have analyzed a number of scheduling algorithms and also tabulated the associated parameters. It has been noticed that disk space management is critical issue in virtual environment. Existing scheduling algorithms give high throughput and are cost effective but they do not consider reliability and availability. So there is a need for algorithms those improve availability and reliability in cloud computing environment.

REFERENCES

[1] 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.

[2] 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.

[3] El-Sayed T. El-kenawy, Ali Ibraheem El-Desoky, Mohamed F. Al-rahamawy "Extended Max-Min Scheduling Using Petri Net and Load Balancing" International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-4, September 2012.

4]   Arash Ghorbannia Delavar, Mahdi Javanmard, Mehrdad Barzegar Shabestari and Marjan Khosravi Talebi "RSDC (Reliable Scheduling Cloud environment Cloud environment Cloud environment Distributed In Cloud Computing)" in International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.2, No.3, June 2012.

[5] Mrs. S. Selvarani; Dr. G. Sudha Sadhasivam, improved cost-based algorithm for task scheduling in Cloud computing, IEEE 2010.

[6] Shalmali Ambike, Dipti Bhansali, Jaee Kshirsagar, Juhi Bansiwal "An Optimistic Differentiated Job Scheduling System for Cloud Computing" International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 2, Mar-Apr 2012, pp.1212-1214.

[7] Shamsollah Ghanbari, Mohamed Othman "A Priority based Job Scheduling Algorithm in Cloud Computing" International Conference on Advances Science and Contemporary Engineering 2012 (ICASCE 2012).