Scheduling Algorithms: A Landscape on Pros and Cons

V. Rajeshram
Computer Science and Engineering
Sri Shakthi Institute of Engineering and Technology
Coimbatore, India

S. Rajasulochana
Computer Science and Engineering
Sri Shakthi Institute of Engineering and Technology
Coimbatore, India

Abstract—**Cloud computing, an internet based computing, is a new business paradigm that provides a dynamic and internet optimized infrastructure for deploying the applications. Cloud computing itself is a grid computing with additional property of elasticity and scalability. Since cloud has a heterogeneous pool of resources, scheduling plays a vital role in cloud computing. It is all about executing the current jobs in a given constraint. Scheduling, an NP-complete problem, is still under a serious research. Quality of Service is an inevitable issue needing to be deal with in task scheduling of cloud computing. Efficient scheduling mechanism should meet the QoS parameters and should enhance resource utilization. Traditional scheduling algorithms such as FCFS, SJF, EASY and backfilling possess fragmentation problems. This paper provides a landscape and discusses the pros and cons of various scheduling algorithms.**

Keywords – Quality of Service; Scheduling; Heuristics; Resource allocation

I. INTRODUCTION

Cloud computing is an emerging technology and also a trend which provides services at anytime anywhere through internet connection with on-demand and pay-as-you-go model. Cloud is a pool of resources where users can rent the resources as per needs and pay for what they use. It has less capital expenditure for setting up an organization. There is no need for huge volume of resources. It is enough to have small amount of physical resources from which users can set to use virtualized resources. Resource sharing is an important advantage in cloud computing that lead to job scheduling process. Job scheduling is one of the important challenge in cloud computing. There are multiple job scheduling algorithms are available in cloud computing environment like FCFS, SJF, Min-Max and Round Robin. Cloud computing is an enhanced concept of basic web services and grid computing. It contains three major services. They are IaaS (Infrastructure-as-a-Service), PaaS (Platform-as-a-Service) and SaaS (Software-as-a-Service). Everything is together combined and is known as Everything-as-a-Service (EaaS) which is needed by every organization. It helps to reduce the maintenance cost of all types of organization. The major service providers of cloud are Amazon(AWS), Google (Google App Engine) and Microsoft (Azure). They provide stream less service to their clients. Cloud computing deployment model can be classified as public cloud, private cloud, hybrid and community cloud. Public cloud denotes every users can access the service provided by the third party, that correctly defines all the users may share the space in same location. Private cloud is created and maintained by an organization for their own purpose and public users cannot access that.
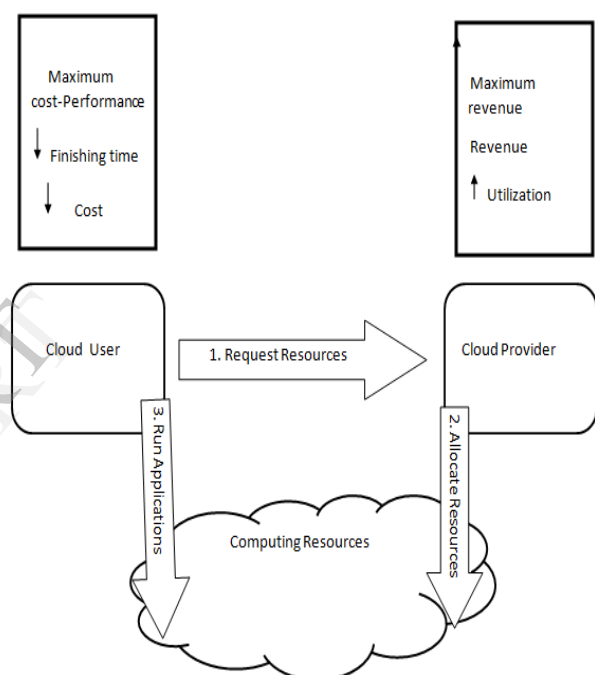

Figure 1 Cloud usage scenario

Hybrid cloud is the combination of public and private cloud. If set of organizations have similar requirements then they share the context known as community cloud. The main features of cloud computing are elasticity, economy, reliability and on-demand. This paper presents a bird's eye view on the various algorithms for scheduling cloud resources.

The paper is organized as follows: Section "Scheduling landscape" gives a brief introduction to scheduling. Section "Heuristic scheduling" discusses the ways to achieve scheduling efficiency based on past history. Section "Scheduling in Cloud" discusses the various scheduling algorithms in cloud. Section "Multi Queue job scheduling" narrates about scheduling using multiple queues. Finally the concluding remarks are presented in the "Conclusion and Future work" section.

II. SCHEDULING LANDSCAPE

Scheduling denotes the arrangement of process, threads, tasks or jobs based on some criteria and user requirements. In job scheduling the jobs allocated to resource. It is important because it manages the resource utilization and reduces the cost.
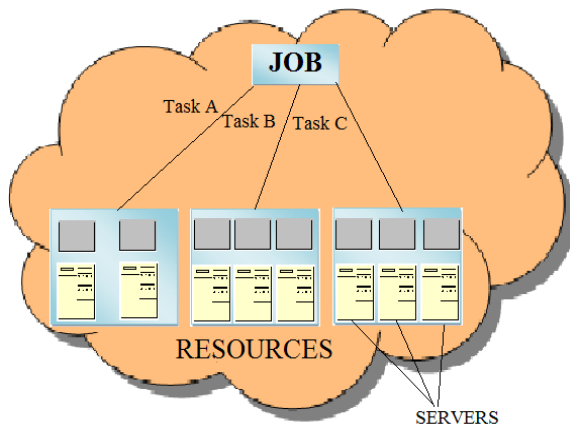

Figure 2 Scheduling tasks to available resources

There are several classifications in the job scheduling strategies. The two major types are preemptive and non-preemptive. Preemptive scheduling depends on the priority. The current executing process state is changed when the new job to be executed interrupts the current process. Examples for preemptive scheduling are round-robin and priority based scheduling. Non-preemptive scheduling denotes the state of the process cannot be changed i.e., the new process could not change the current executing process state. Examples for non-preemptive scheduling are FCFS and SJF. Priority scheduling is another aspect where the priority is assigned internally or externally. Internal priority is assigned by measurable qualities or quantities to compute process.

III. HEURISTICS SCHEDULING

Heuristics based job scheduling can be under both of Preemptive and non-preemptive. It denotes that the jobs are scheduled based on their previous history. It is again classified as batch mode heuristics and online mode heuristics. In the batch mode, the jobs are collected and arranged; finally it is executed in the system. The main examples for batch mode heuristics algorithm are First Come First Served scheduling algorithm (FCFS), Round Robin scheduling algorithm (RR), Min–Min algorithm and Max–Min algorithm. In the online mode, the order is not changed i.e.,the order in which the jobs are executed is the same. The example for online mode scheduling is Most fit task scheduling algorithm (MFTF).

IV. SCHEDULING IN CLOUD

Scheduling in cloud is done in three stages. They are

❖ Discover the resources and filter it: Datacenter Broker search the resources present in the cloud and gather the status information related to the resource.

❖ Select the suitable resource: Destination resource is selected based on desired parameters of task and resource. This is deciding stage.

❖ Submit the task to the resource: Task is submitted to resource selected depending on the parameter.

V. EXISTING SCHEDULING ALGORITHMS

There are several scheduling algorithms are available. Some of them are FCFS, SJF, RR and backfill algorithms.

A. FCFS

FCFS denotes First Come First Serve which allows the jobs to be executed in the same order in which they arrive. Here the smallest jobs wait in the queue for a long time till the large job finishes and so the resources also underutilized.

B. SJF

SJF denotes the smallest job first which also leads starvation for large jobs and also the resources are underutilized. Here the jobs with shortest execution time get preference and is executed first. It is also known as Shortest Job Next (SJN) algorithm.[5]

C. RR

In the round robin, jobs are scheduled by time slice and rounded queue manner. Each job in the queue executed for the predefined time slice. Once all the jobs are executed for the time slice the queue is started again from the first and for the same time slice the jobs are executed until all the jobs will be completed. Here the jobs are not finished within the scheduled time which also leads to performance degradation and increases cost.

D. Min-Max

In the Min-Max algorithm, the smallest jobs gets priority and longer jobs need to wait in queue until all the small jobs finish their execution.

E. Max-Min

In the Max-Min algorithm, the longer jobs are executed first and so the smaller jobs cannot get priority. They must wait until all the longer jobs finish their execution.

F. PSJN

Pre-emptable shortest job next scheduling algorithm which is known as PSJN is used in private cloud environment. PSJN combines the pre-emption technique of Round-robin algorithm with shortest process next (PSN). As a result, the cost benefits and also improves the response time and execution time. In case of Improved Cost Based algorithm the jobs are grouped according to processing capabilities of the available resources.

G. RASA

Resource-Aware-Scheduling algorithm (RASA). It combines the traditional scheduling algorithm that is Min-Max and Max-Min. It has the advantage of both algorithms and also covers the disadvantages. By considering the completion time of task and available

resources it applies Min-max or Max-Min algorithm randomly.

H. BACKFILL

Backfill is yet another important category in scheduling where the jobs can be moved if they will not delay the ahead job in the queue. It is known as Backfill strategy. Two types of backfilling are EASY (Extensible Argonne Scheduling System) backfilling and conservative backfilling.

a) EASY Backfilling and Conservative Backfilling

In case of EASY backfilling the jobs can be moved if they do not delay first waiting job. In the conservative backfilling the jobs can be moved if they do not delay any jobs in the queue.

b) Aggressive Backfilling

Aggressive backfilling also same as backfilling strategy but it ensures the first job of the queue must not be delayed.

c) CBA Backfilling

CBA denotes the Combinational Backfill Algorithm where a group of small jobs selected to backfill the resource gap and increases the resource utilization. Combination of more than one jobs to be backfilled is known as CBA. If there is no available jobs then that again lead to EASY backfilling. The above all creates fragment in the scheduling which leads to the heuristic based multi-queue job scheduling. [2]

d) SJF-BF

SJF-BF denotes Shortest job first Backfill. The job with minimum execution time move ahead in the queue and the backfill fills the space created by high priority jobs.[5]

Multi-Queue job scheduling

Multi Queue job scheduling is a new strategy that uses multiple queues to scheduling the jobs. Here the jobs are sorted based on the burst time and they are moved to the three different queues. The queues denote the jobs that are grouped based on the burst time as small, medium and long jobs. First small contains the 40% of the total job and the second medium queue also contains 40% of the total jobs and the last queue long contains the remaining 20% jobs. From these three different queues jobs are selected dynamically which eliminate the starvation that is "starve to death". [1]

AV. Karthick [1] proposed a multi queue job scheduling algorithm to reduce the cost of reservation and on-demand plan. Here jobs are put it into three different queues and selected randomly from the queue. That provides optimum solution the scheduling problem and efficient utilization of unused free spaces.

Lawson [2] proposed a backfilling scheduling algorithm based on aggressive backfilling strategy. Here multiple queues are used for scheduling. The waiting queues are used to separate short jobs from long jobs based on their estimated execution time.

A.D. Techiouba [3] proposed a job scheduling algorithm based on priority. Priority computed based on the set of heuristics. Heuristics represents the dead line and execution time of the job. Here the three different priorities combined namely administrative, user and scheduler. That provides improved resource utilization and meet the QoS requirement. Heavy workload is the drawback in this system.

Lawrence et al [4] proposed a resource scheduling methodology using potentially all pair-wise rankings of all possible alternatives (PAPRIKA). PAPRIKA method evaluates fairness based on user satisfaction. This method makes use of both task matrix and resource matrix. Priority of the resource is calculated with respect to a threshold value. Tasks are then mapped to the resource that gives higher user satisfaction, thus improving resource utility and minimizing the allocation time.

VI. CONCLUSION

Cloud computing is an emerging topic in distributed computing and is a door to the future of information technology. Scheduling is among the several issues that cloud computing is facing today. This paper presents a brief survey on the pros and cons of various scheduling algorithms. In future, resources can be effectively and efficiently scheduled by leveraging techniques like, scheduling and resource provisioning, request scheduling and task scheduling, scheduling and load balancing etc where efficiency and fairness will be the end goal.

VII. ACKNOWLEDGMENT

VIII. REFERENCES

1. AV.Karthick, R.Ganapathy Subramanian, Dr.E.Ramaraj, "An Efficient Multi Queue Job Scheduling for Cloud Computing," 2014 World Congress on Computing and Communication Technologies, ISSN 978-1-4799-2876-7/13, March 2013.
2. Lawson, Barry G., Smirni, Evgenia, "Multiple-queue Backfilling Scheduling with Priorities and Reservations for Parallel Systems" Department of Computer Science, College of William and Mary Williamsburg, VA 23187-8795, USA
3. A.D.Techiouba, G.Capannini, Ranieri Baraglia, D.Puppin, M.Pasquali, Laura Ricci (2002), "Backfilling strategies for scheduling streams of jobs on computational farms", European CoreGRID NoE, IST-2002-004265.
4. Hilda Lawrance and Salaja Silas, "Efficient Qos Based Resource Scheduling Using PAPRIKA Method for Cloud Computing," International Journal of Engineering Science and Technology (IJEST), Vol. 5, No.03, ISSN : 0975-5462, March 2013.
5. Anurag Mishra, Dharmender Singh Kushwaha and Shakti Mishra, "An Improved Backfilling Algorithm: SJF-BF," Int. J. on Recent Trends in Engineering & Technology, Vol. 05, No. 01, Mar 2011.