International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 5, ISSN No. 2455-2143, Pages 30-35
Published Online March – April 2016 in IJEAST (http://www.ijeast.com)

TOWARDS THE VARIOUS CLOUD COMPUTING SCHEDULING ALGORITHMS: A REVIEW

Kulwinder Kaur
Swami Vivekanand Institute of Engineering & Technology,
Rajpura, India

***Abstract*- Cloud computing era refers to a dynamic, scalable and pay-per-use distributed computing model empowering designers to depart this world application amid job designation and storage distribution. The approach of cloud computing as another model of administration provisioning in circulated frameworks, urges analysts to examine its advantages. The essential advantage of moving to Clouds is application versatility. It is exceptionally advantageous for the applications which are sharing their assets on various hubs. The cloud computing mainly aims to give proficient access to remote and geographically distributed resources. Scheduling the assignment is truly a testing in cloud environment. Normally errands are scheduled by client prerequisites. New scheduling systems should be proposed to beat the issues proposed by system properties in the middle of client and assets. The New scheduling methodologies might utilize a portion of the traditional scheduling ideas to union them together with some system aware techniques to give solutions for better and more effective task scheduling. Job scheduling algorithms believed to be the most hypothetical problems in the cloud computing domain. Numerous deep investigations and efforts have been carried out in this regard. This paper intends to present the performance comparison analysis of varied pre-existing job programming algorithms considering various parameters.**

Keywords— ***Cloud Computing, Job Scheduling, Resource allocation, Efficiency, Performance, Cost***

I. INTRODUCTION

The most recent advancements in cloud computing are making our business applications significantly more versatile and community oriented, similar to famous purchaser applications like Facebook and Twitter. As purchasers, we now expect that the data we think about will be pushed to us continuously, and business applications in the cloud are travelling in that course also. Cloud computing is the early innovation which depends on pay-per-use basis. It is computing worldview wherever applications, data, information transmission and IT administrations area unit given over the web. Objective of the cloud administration suppliers to utilize plus effectively and attain the foremost extreme profit. Thus, this prompts errand scheduling as a canter and testing issue in cloud computing. Scheduling is the procedure of choosing how to submit assets between assortments of conceivable undertakings [1]. Cloud computing is the conveyance of figuring administrations over the Internet. Cloud administrations allow individuals and organizations to utilize programming and instrumentation that area unit overseen by outsiders at remote areas. Illustrations of cloud administrations incorporate online record stockpiling, long range interpersonal communication locales, webmail, and online business applications. The cloud computing model permits access to data and PC assets from anyplace that a system association is accessible. Cloud computing gives a mutual pool of assets, including information storage room, systems, PC preparing control, and concentrated corporate and client applications. Cloud computing could be a model for empowering useful, on-interest system access to a regular pool of configurable enrolled resources (e.g., networks, servers, repositing, applications, and administrations) which will be quickly provisioned and discharged with insignificant administration travail or service provider association. The qualities of cloud computing incorporate on demand self-administration, expansive system access, asset pooling, quick flexibility and measured administration. On demand self-administration implies that clients (typically associations) can ask for and deal with their own particular Computing assets. Expansive system access permits administrations to be offered over the Internet or private systems. Pooled assets imply that clients draw from a pool of figuring assets, more often in remote data centers. Administrations can be scaled bigger or smaller; and utilization of an administration is measured and clients are charged in like manner. The cloud computing administration models are software package as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) in an exceedingly software package as a Service model, a pre-made application, aboard any needed programming, operating framework, equipment, and system are given. In PaaS, a working framework, equipment, and system are given, and the client introduces or builds up its own particular programming and applications. The IaaS model gives only the equipment and system; the client introduces or creates its own particular working frameworks, programming and applications.

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 5, ISSN No. 2455-2143, Pages 30-35
Published Online March – April 2016 in IJEAST (http://www.ijeast.com)

Cloud computing models are moving. In the cloud/customer design, the customer is a rich application running on an Internet-associated gadget, and the server is an arrangement of use administrations facilitated in an undeniably flexibly adaptable cloud computing stage. The cloud is the control point and framework or record and applications can traverse various customer gadgets. The customer environment might be a local application or program based; the expanding force of the program is accessible to numerous customer gadgets, versatile and desktop alike.[2]
Vigorous abilities in numerous cell phones, the expanded interest on systems, the expense of systems and the need to oversee data transmission use makes motivations, in some cases, to minimize the cloud application registering and capacity footprint, and to abuse the insight and capacity of the customer gadget. However, the undeniably complex requests of versatile clients will drive applications to request expanding measures of server-side figuring and capacity limit.

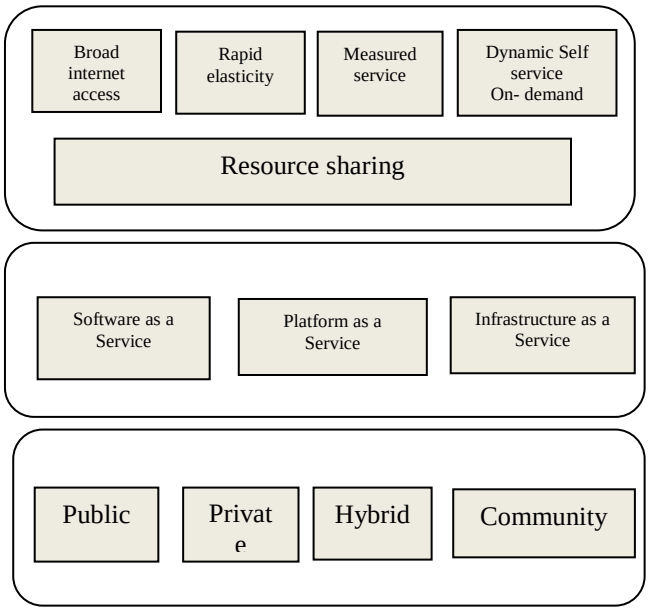

Fig.1 Overview of cloud computing

The rest of the paper is organized as follows: Section II proposes scheduling categories in cloud computing Section III explains different Scheduling Types. Section IV talks about the related work and Comparison in this field and Section V brings the conclusion and future extent of the paper.

II. SCHEDULING IN CLOUD COMPUTING

As a huge number of clients share cloud resources and dispatch their tasks to the cloud, it has become a challenge to schedule these tasks. Hence, job scheduling is a hot topic in distributed and cloud computing. Fig. 2 shows the scheduling stages in a virtualized environment of cloud computing.

The scheduling algorithms belonging to distributed systems mainly aim to partition the load and assign sub loads to processors to attain their most utilization whereas minimising the entire load execution time. It is a driving force in providing flexible and reliable systems. As cloud computing is highly dynamic and to cope up with the fluctuating demands of customers it becomes severely necessary to address the resource allocation problems. However, it's a giant challenge in itself to attain associate degree economic planning algorithmic program style and implement in cloud computing setting.

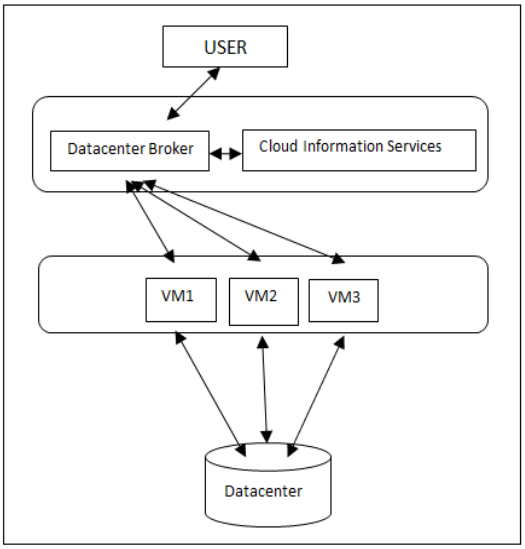

Fig.2 Scheduling stages in cloud computing

Cloud scheduling is categorised into- cloud service scheduling and user level scheduling. Cloud service scheduling is ordered at client level and framework level. At client level scheduling manages issues raised by administration procurement between suppliers furthermore, clients. The framework level scheduling handles asset administration inside datacenter. Datacenter comprises of numerous physical machines. Millions of errands from clients are received; task of these undertakings to physical machine is done at datacenter. This task or programing altogether has effects on the execution sequence of datacenter. User level programing incorporates Market-based and closeout primarily based schedulers that area unit appropriate for steering the availability and request of cloud assets. Market based asset distribution is successful in cloud figuring environment where assets are virtualized furthermore, conveyed to client as an administration.

Classification of scheduling algorithms

a. Static versus Dynamic scheduling algorithm. Static scheduling involves less runtime overhead as it involves prior fetching of the required data and pipelining of different task execution stages whereas in dynamic scheduling, system does

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 5, ISSN No. 2455-2143, Pages 30-35
Published Online March – April 2016 in IJEAST (http://www.ijeast.com)

not have any clue regarding upcoming job /task. Thus execution time of the task remains unknown and the allocation of tasks is done at the moment the activity executes [3, 4].

b. **Centralized and Decentralized scheduling**- As the name itself depicts, in centralized scheduling global decisions are made by centralized distributed scheduler. It has many benefits as it is easy to implement; more efficient and provides better control over resources. But since a scheduler for large scale grid environment lacks scalability, fault tolerance and efficient performance, so it limits the performance. Whereas decentralized scheduling is friendly towards real cloud scenarios due to absence of any centralised control entity, where each system manages and maintains state of jobs queue locally and never suffers from bottleneck.

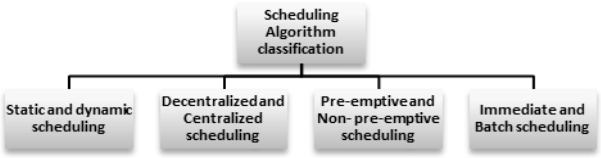

Fig.3 scheduling algorithms classification

c. Pre-emptive and Non- pre-emptive scheduling- The former scheduling category allows job interruption, midway execution and a job may migrate to alternative resource making its own assigned resource available to other jobs while in latter one, resources cannot be withdrawn until the execution of the running job finishes.

d. **Immediate and Batch scheduling**- In immediate scheduling, as the name itself signifies, jobs are scheduled as soon as they arrive without any delay of time interval on available resources while in batch scheduling, the scheduler assigns the arriving jobs into a batch to be dispatched over adjacent time intervals.

Scheduling in cloud is generally comprised of three stages:

- Resource discovery and filtering- resources in network and their related status information is collected.
- Resource selection- target resource is chosen based on various parameters.
- Job submission- job is submitted to the chosen resource.

III. TYPES OF SCHEDULING ALGORITHMS

Various advancements have been made towards different calculations for designating, scheduling and scaling the assets productively in the cloud. The essential target of scheduling calculation is: execution upgrade and enhancing the nature of administration alongside keeping up the productivity and decency among the employments and decrease the execution cost. Customary scheduling calculations are insufficient able to accomplish these destinations. So to overcome these limitations various enhanced algorithms are proposed .Cloud computing is a rising technique. So to upgrade the utilization of advantage in cloud, minimizing the planning cost, extend the execution of the server, minimizing the taking care of time and complete time it is extraordinarily imperative to logbook the assignments in the cloud. Diverse creators have considered the scheduling issue and have been seen as N-P hard. With the utilization of various approaches, a few structures were proposed by creators to handle this scheduling issue and among those timetables that have achieved best results are:

A. *First Come First Serve (FCFS)*

First Come First Serve alias First in First out suggests that the jobs are executed as per the order of job arriving time. The FCFS algorithm may further breed the convoy effect which usually takes place when there is a job with a huge amount of workload in the job queue. In this scenario, all the jobs which are queued behind have to wait a long time for the long job to finish [3].

B. *Shortest Job First Scheduling Algorithm*

Shortest Job First (SJF) or Shortest Job Next (SJN) picks up the task with the least execution time. Highest priority is assigned to the jobs with minimum execution time and placed first in queue while the lowest priority is assigned to the job with the maximum execution time. It can be either pre-emptive or non-preemptive. A pre-emptive SJF algorithmic rule can stop the method presently in execution, whereas a non-pre-emptive SJF algorithmic rule can let the running method to end its half [3].

C. *Round Robin scheduling algorithm (RR)*

Round Robin algorithm is considered as one of the simplest, conventional and most used scheduling algorithms which works exceptionally better for timesharing systems. It distributes the load equally to all the resources. It works very similar way in cloud computing as it does in process scheduling. The working includes a circular queue and a fixed time unit called quantum. Each individual job's execution takes place only within this quantum. The whole process goes like this: first of all the first process from the queue is picked by CPU scheduler after which it sets a timer to interrupt after one quantum and then finally the process is dispatched. In case if the job does not achieve in one allotted quantum, it returns to the queue and wait for the following round. The massive benefit is that jobs are executed in a sequence turn wise and there is need to wait for the previous job to get complete. Therefore, there is no starvation issue. But the dark side is that if the workload is heavy and queue is fully loaded,

it takes a lot of time to accomplish all the jobs and moreover, a perfectly suitable time quantum is hard to decide [4].

D. *Priority based Job Scheduling Algorithm*
In order to reduce the make span time another Priority based Job Scheduling (PJSC) Algorithm for cloud computing has been proposed by Ghanbari.S, which is pre-emptive in nature where each process in the system is based on the priority and priority is allowed to run. The highest priority job can run first whereas lower priority job can be made to wait. Equal-Priority processes are scheduled in FCFS order. The drawback of this algorithm is starvation of a process.[5].

E. *Min-Min scheduling algorithm*
The working concept behind the Min-Min algorithm is to map each task to resources such that they can accomplish the task in the minimum possible time. It estimates the execution and completion time of each job on each available resource. There are two phase in the Min-Min algorithm. In the first phase it calculates the least execution time of all tasks. Further in second phase, the task with the least execution time among all the tasks is picked up. The algorithm further assigns the task to the resource producing the minimum completion time. The same procedure goes on repeating until all the tasks are scheduled [6].

F. *Max-Min scheduling algorithm*
The Max-min algorithm works very similar to the Min-min algorithm. The differentiating feature is as per the name because in this the word “min” is replaced by “max”; i.e the task having the utmost earliest completion time is allotted to the corresponding resource. Here, larger tasks are given priority over the smaller tasks [6].

G. *RASA Task Scheduling Algorithm*
RASA is a new scheduling algorithm composed of two traditional techniques- Max-min and min-min. Min-Min strategy is used to execute small tasks before large tasks and Max- Min strategy is applied to avoid the delays in large tasks execution. Both the ways are used for tasks and alternative exchange ends up in consecutive execution of a tiny low and an outsized task on totally different resources therefore ignoring the waiting time of the tiny tasks in Max-min rule and therefore the waiting time of the massive tasks in Min-min algorithm.[7]

H. *Schedule First, Manage Later: Network-Aware Load balancing*
A.Nahir proposed a novel plan that brings about no correspondence overhead between the clients and the servers upon job landing, in this manner expelling any scheduling overhead from the job's basic path. This Approach enhances the normal queuing overhead over customary plans by a component of 9 (or more) under all load conditions. Moreover, demonstrate that algorithm stays productive even at the point when the between server signal spread postponement is noteworthy (with respect to the jobs execution time). Heuristic answers for the execution corruption is provided that happens in such cases and appear, by re-enactments, proficiently alleviating the negative impact of proliferation postponements. For the effectiveness of proposed approach in a genuine situation, load balancing framework is executed in view of it, sending the framework on the Amazon Elastic Compute Cloud (EC2), and measuring its execution.[8].

I. *Pre-emptive Scheduling of On-line Real time services with task migration*
Santhosh.R presents in the paper an on-line planning issue of in progress errands utilizing "Infrastructure as a Service" model offered by cloud computing. The continued errands are set-aside pre-emptively with the goal of increasing the mixture utility and effectiveness to minimize the reaction time and to enhance the effectiveness of the assignments. The assignments are relocated to another virtual machine at whatever point any task misses its due date. This enhances the general framework execution and amplifies the aggregate utility. The proposed calculation can essentially beat the EDF and Non Pre-emptive scheduling calculation. [9].

J. *Hyper-Heuristic Scheduling Algorithm*
C.W.Tsai proposed a novel Hyper-Heuristic Scheduling Algorithm to grasp JSP to diminish make span time and to find better scheduling answers for cloud computing structures. Two discovery administrators have been used by the proposed calculation to modify the escalation and extension in the chase of arrangements amid the meeting technique.[10].

K. *Particle Swarm Optimization Algorithm*
Pandey.S introduced a particle swarm optimization (PSO) based heuristic to timetable applications to cloud assets that considers both calculation expense and information transmission cost. It is utilized for work process application by changing its calculation and correspondence costs. It compare about the cost investment funds when utilizing PSO and existing 'Best Resource Selection' (BRS) calculation. The outcomes demonstrates that PSO can accomplish 3 times cost investment funds as contrasted with BRS, and great circulation of workload onto resources.[11].

L. *Ant Colony Optimization scheduling*
Several researchers have impressed from social behaviour of insects and ants have impressed variety of strategies of improvement technique called Ant colony optimization (ACO).This algorithm has an edge over the unique approach in which every ant assemble their own particular individual result set and it is later on incorporated with a complete arrangement.
The basic approach is to emulate the foraging behavior of ants. While searching for food, ants normally communicate using a particular kind of chemical known as pheromone. Starting

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 5, ISSN No. 2455-2143, Pages 30-35
Published Online March – April 2016 in IJEAST (http://www.ijeast.com)

with random search once the ants get succeeded in finding food source, they leave pheromone on that path and it becomes traceable by other ants. Ant colony optimization is irregular enhancement seek approach that will be utilized for designating the approaching occupations to the virtual machines.[12] With the continuation in process, ants tend to choose the shortest path with huge amount of pheromone accumulated over there [13].

M. *Bees Life algorithm for job scheduling in cloud computing*
S.Bitam display another Bee Swarm advancement calculation called Bees Life Algorithm (BLA) connected to profitably arrange calculation jobs among making ready assets onto the cloud datacenters. it's thought of as NP-Complete issue and it goes for spreading the workloads among the making ready assets in a perfect manner to diminish the mixture execution time of occupations and afterwards, to create strides the viability of the whole cloud computing administrations. [14] [15].
It is used for both combinatorial and computational problems and provides both kind of general and optimal solutions. It aims at spreading the workloads among resources in an optimal fashion to reduce total execution time, cut execution costs and improve the effectiveness of cloud services [16].

Table 1: Theoretical findings of Scheduling Techniques

Algorithms	**Scheduling principle**	**Allocation order**	**Findings**	**Type of system/ environment**
FCFS scheduling	Order of arrival	In the order of process arrival	More waiting time	Batch system
SJF scheduling	Arrival time and process time	To the processes with least execution time	Lesser waiting time than FCFS	Batch system
RR scheduling	Time quantum difficult to decide	Pre-emption after fixed quantum	More waiting time than all	Time- sharing system
Priority based scheduling	Assigned priority	To the job with highest priority	Lesser finish time	Batch and Time- sharing system
Min- Min Scheduling	Execution and completion time	Least exec time task assigned to resource producing min completion time	Increase resource utilization rate	Batch system
Max-Min Scheduling	Completion time	Max earliest time task assigned to resource	Small task delays for long time	Batch system
RASA	Make Span has to be reduced	Number of resources decide the strategy	Reduce make span	Batch system
Schedule First , Manage, later Network Aware Load Balancing	Dynamic	Enhances the normal queuing overhead over customary plans by a component of 9 (or more) under all load conditions.	Load , Execution Time	Cloud System
Pre-emptive Scheduling of Online Real time Services with Task Migration	Dynamic	It intends to diminish the reaction time and to enhance the effectiveness of the tasks. Enhances The general Framework Execution and boosts the aggregate utility.	Execution time ,Efficiency, Cost	Batch Mode
Hyper-Heuristic Scheduling algorithm	Dynamic	It reduces the total makespan time of jobs.	Makespan Time	Batch Mode
PSO based heuristic scheduling	Resource utilization time	Random initialization	Good distribution of workload onto resources	Cloud environment
ACO scheduling	Pheromone updating rule	Loads each VM and defines load balancing factor	Reduced time, improved results by local search	Cloud Environment
Honeybee Scheduling	Scout bees forage the food source	Each class performs Waggle dance and advertise about food source	Reduced make span and improved performance	Cloud Environment

IV. CONCLUSION

Efficiency of cloud depends on the type scheduling algorithm used in environment. Scheduling still one of the ongoing research oriented domain in cloud computing scenario. In this paper, a number of existing algorithms for job scheduling are discussed, compared with each other and tabulated with their findings. It helps to understand the wide variety of scheduling options in order to select one for a given environment. Most of the algorithms schedule tasks supported single criteria (i.e. execution time). However in cloud setting it's needed to think about numerous criteria like execution time, cost, information measure of user etc. Multi-objective task scheduling algorithm provides better solutions to cloud environment. In order to improve improve the throughput of the datacenter and reduce the cost without violating the SLA (Service Level Agreement) for associate degree application in cloud SaaS environment.

.

V. REFERENCES

[1] Peter Mell and Timothy Grance. "The NIST Definition of CloudComputing (Draft)", 2011.

[2] L.Wang, G.Laszewski, M.Kunze, J.Tao, "Cloud Computing: A Perspective study ," J. New Generation Computing, pp.1-11, 2010.

[3] Kansal, N. Jain and I. Chana. "Cloud load balancing techniques: A step towards green computing", IJCSI International Journal of Computer Science Issues vol. 9, issue. 1, pp. 238-246, 2012.

[4] L. Kumar and A. Verma. "Workflow Scheduling Algorithms in Cloud Environment - A Survey", In Proceedings of RAECS UIET Panjab University Chandigarh, pp. 1-4, 2014, DOI: 978-1-4799-2291-8/14.

[5] Ghanbari.S,Othman.M," A Priority based Job Scheduling Algorithm in Cloud Computing", International Conference on Advances Science and Contemporary Engineering,2012.

[6] S. Devipriya and C. Ramesh. "Improved Max-min heuristic model for task scheduling in cloud.", International Conference on Green Computing, Communication and Conservation of Energy (ICGCE), pp. 883-888, 2013.

[7] Parsa.S , Maleki.R.E , " RASA: A new task scheduling algorithm in grid environment ",World Applied Sciences Journal 7 , pp.152-160,2009.

[8] A.Nahir, A.Orda, D.Raz," Schedule First , Manage Later: Network Aware Load Balancing ", IEEE ,2013.

[9] Santosh.R,Ravichandran.T," Pre-emptive Scheduling on Online- Real Time service with task migeration for cloud computing ", International Conference on Pattern Rcognition,IEEE,2013.

[10] C.W.Tsai, W.C.Huang ," A Hyper- Heuristic Scheduling Algorithm for Cloud," IEEE Transactions on Cloud Computing vol.2,pp.236-250,2014.

[11] Pandey.S.Wu.L, Guru.S,Buyya.R," a Particle Swarm Optimization – based Heuristic for scheduling workflow applications in cloud computing environments",24th IEEE international Conference on Advanced Information Networking and Applications,2010.

[12] M. Tawfeek, A. El-Sisi, A. E. Keshk and F.A. Torkey. "Cloud Task Scheduling Based on Ant Colony Optimization",computer engineering and systems , pp 64-69, 2013.

[13] L. Kun, G. Xu, G. Zhao, Y. Dong, and D. Wang. "Cloud task scheduling based on load balancing ant colony optimization.", In Sixth Annual China Grid Conference (ChinaGrid), pp. 3-9, IEEE 2011.

[14] S.Bitam," Bees Life algorithm for job scheduling in cloud computing ", International Conference on Computing and Information Technology ,2012.

[15] L.D. Dhinesh Babu and P. Venkata Krishna. "Honey bee behavior inspired load balancing of tasks in cloud computing environments", Applied Soft Computing, vol. 13, issue 5, Elsevier, pp. 2292-2303,2013.

[16] Garg, A. Krishna, C.R, An improved honey bee's life scheduling algorithm for a public cloud, IEEE 2014