A STUDY OF SCHEDULING ALGORITHMS IN CLOUD COMPUTING

Baldev Singh
Lyallpur Khalsa College, Jalandhar, Punjab (India)
bsd.lkc@gmail.com

Abstract
Cloud computing based services to the cloud users on Pay-as-Usage basis at anytime and anywhere. To ensure the fast and timely availability of the cloud resources to the cloud consumers is must for quality of service (QOS). Optimal resource utilization can be made possible if the appropriate scheduling is made for the cloud resources. Scheduling is the phenomenon that refers to the appropriate use of resources to the cloud based jobs. Scheduling leads to the efficient resource utilization and better performance in cloud environment. Effectual ordering and proper mix of jobs is not an easy job that is demanded for QOS and better throughput. Efficient use of scheduling algorithms helps in effective cloud resources' utilization and management. This paper is a study of several scheduling algorithms that assist in better resource provisioning and better load balancing.

Keywords: Cloud computing, Resource utilization, Scheduling Algorithms, QOS, Virtulization.

I. Introduction

Technologies and technology based paradigms are developed and evolutionized from time to time. Cloud computing is emerging as hot spot as per technological developments are concerned which is evolutionized from the existing technologies. This distributed paradigm is being used for providing variety of services to the cloud users. Cloud is defined more precisely by NIST. The NIST definition of cloud which is used widely defines cloud as [2, 15] "a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." Cloud computing is capable to provide IaaS, PaaS and SaaS services to the cloud users on Pay-as-Usage basis at anytime and anywhere. Various services or features of cloud computing encompasses on-demand service, broad Internet based access, resource pooling, elasticity, scalability, pay-as-usage basis service [1,12].

Internet is the host of cloud services. Cloud has the distributed architecture and provides computational facility as well through which various cloud computing features like availability, scalability, adaptability, agility and collaboration which are helpful for increasing the efficiency and performance of an enterprise or organization. Cloud has a vast pool of resources which are actual and virtual in nature. Virtualization is used for providing the secure services to the cloud users. As various issues, threats and vulnerabilities [3,6] exist in traditional information technology (IT) tools, there exists more or less similarities in cloud computing environment.

Figure-1 depicts the cloud computing reference architecture which is base of various cloud services. This generic reference model depicts high-level conceptual design which encompasses the requirement specification, structures, as well as various operations and activities of cloud computing. Cloud computing is mainly based on cloud computing reference model which consists of three main services [2] known as Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS).

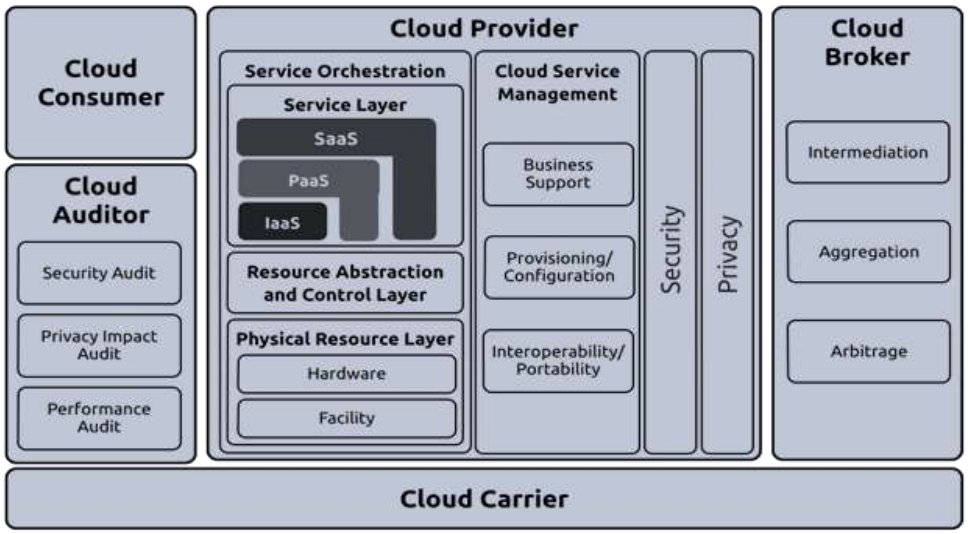

Figure 1: Cloud Computing Reference Architecture

Software as a Service (SaaS): Software as a Service is the capability provided to the consumer by the provider to use the providers' applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (e.g., web-based email).

Platform as a Service (PaaS): Platform as a Service is the capability provided by the provider to the consumer to deploy onto the cloud infrastructure his own applications without installing any platform or tools on their local machines. PaaS refers to providing platform layer resources, including operating system support and software development frameworks that can be used to build higher-level services.

Infrastructure as a Service (IaaS): Iaas is the capability provided to the consumer by the cloud service provider by way of provisioning processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications.

Resource provisioning in cloud is basically the distributed use of servers and storage using virtual machines' concept. The virtual machines are created using hardware resources at data centres of cloud providers and cloud resources are shared among cloud consumers by way of virtual machines. The cloud consumers pay-for-use of the cloud resources and they require the optimal and efficient access and use of cloud resources. The efficient and optimal use of cloud resource typically require the proper and efficient scheduling of cloud resources. The quality of service to be provided to the cloud consumers is the foremost goal of cloud computing which is possible if the cloud resources are managed by using efficient scheduling so that proper load-balancing of cloud resources can be achieved. Various parameters like bandwidth, time, cost, resource availability, priority as well as throughput of resources utilized are mainly base of the resource scheduling algorithms. Section II and II discusses cloud computing taxonomy and virtualization. Further sections IV and V describes the scheduling in cloud computing and related literature survey. Section VI consists of discussion and conclusion of the study presented in this paper.

II. Cloud Computing Taxonomy

Cloud computing taxonomy refers to various actors [2] along with their role, activities and functions that are part of cloud computing environment. Various views and descriptions, uses, as well as standards of cloud computing are depicts in this reference model. Table-1 shows various actors that has specific roles in cloud computing environment.

Table-1: Actors in Cloud Computing

Actor	Role
Cloud Provider	Cloud provider provides various services like IaaS, SaaS and PaaS to the cloud end users.
Cloud Consumer	Cloud consumer refers to the person or an organization who uses the cloud based services which are provided by the cloud service provider.
Cloud Broker	Cloud broker refers to an entity that provides three categories of services. These three categories are Intermediation, Aggregation and Arbitrage.
Cloud Auditor	A cloud auditor provides the role of auditing of cloud services like assessment of services provided by the cloud providers, security,

	privacy and performance aspects that are included in service level agreements.
Cloud Carrier	A cloud carrier provides the role of intermediary between the cloud consumers and cloud service providers. Cloud carrier acts as a bridge to provide connectivity and transport of cloud services. It provides the access of cloud computing devices like tablets, laptops, notebooks, navigation, computers, storage services etc.

III. Virtualization

Virtualization in cloud computing leads to huge paybacks to cloud stakeholders mainly cloud providers. It is treated as a technology or a process that creates a virtual environment by way of creating virtual servers, virtual infrastructures [2,6], virtual storage devices and/or computing resources and applications. Through virtualization, we can run desired program without interfering or changing any existing services provided by the server or platform. Multiple users can use a single physical instance of an application of a computer resource on shared basis, it is possible by creating logical resources of a physical resource through virtualization software. Virtualization is classified into four types:

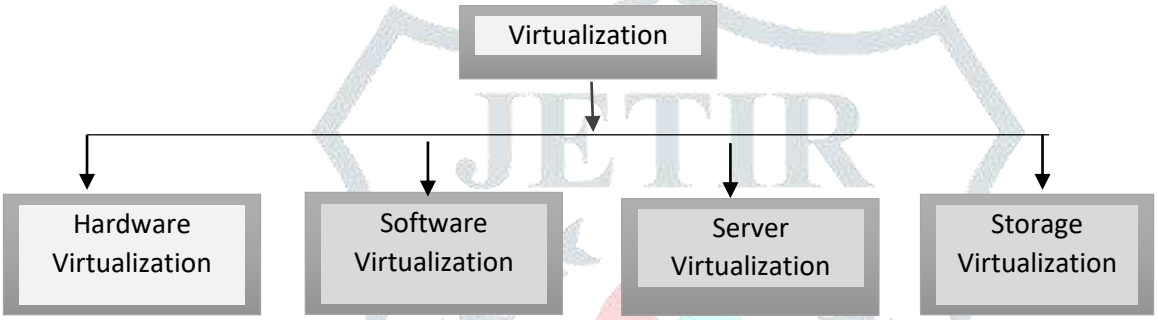

Figure-2: Virtualization

Hardware virtualization in cloud environment is taken place by installing virtual machine software on hardware (physical machine). Virtual machine manager (VMM) or hypervisor which is installed on the server machine is treated as server virtualization. Multiple servers are created on a single physical server and are used on demand basis as well as for purpose of load balancing. Virtualization software (VMM) also acts as a controller and monitor to control and monitor the various resources of the machine like processor, memory and other resources related to hardware.

Software virtualization further classified as Operating System Virtualization, Application Virtualization and Service Virtualization. Software Virtualization is used to create a virtual environments on the host machine to run multiple operating systems virtually, hosting multiple applications on the single instance of hardware as well as hosting particular services or processes for individual applications. For example (Operating System Virtualization) user can run Linux on the host machine and you are also able to run other operating system like Linux as native operating system on the same hardware.

Storage virtualization technique in cloud environment is used to create a virtual pool of storage (simulated) by using various physical storages that are interconnected and is managed by using a single command console. Example of storage virtualization is VMware's vStorage. Other types of virtualizations are data, network and desktop virtualization. Some of the examples of hypervisor are Oracle Virtual Box, VMware Fusion, Oracle OVM for SPARC, Hyper-V. Virtualization is the process used in cloud computing through which software hardware is provided to the cloud users and also load balancing is possible in cloud resources.

IV. Scheduling in Cloud Computing

Scheduling is the process of allocation of variety of jobs to the available cloud resources. The main objective of scheduling is to optimize the use of resources whether it is a computation activity, control activity or other relevant activity. Generally a job is a collections of various tasks that are performed by using various resources like, CPU, memory etc. Scheduling activity [5,8,9] has mainly two actors cloud consumers and providers for which it is required. Scheduling activity is must for the possible shortest time required for optimal utilization of cloud resources and quality of service (QOS) demanded by the consumers. Scheduling activity consists of resource discovery i.e. receiving and releasing of resources on requirement basis, rescheduling of tasks that are incomplete due to some failure, optimal utilization of the resource infrastructure, QOS by way of monitoring, secured delivery of resources,

real time workflow, managing the virtual machines. Scheduling can be the static scheduling or dynamic scheduling. In static scheduling, the mechanism is based on compile time arranging of the tasks for execution. In dynamic scheduling, run time execution of the scheduling mechanism is adopted. Cloud task scheduling effectively done at two levels. First level is the level when cloud users send the tasks known as cloudlets for execution. The second level of cloud task scheduling is the mapping of tasks with the appropriate resources so that maximum utilization of resources can be taken place with the minimum time span.

V. Literature Review

Scheduling algorithms are deployed to attain the highest level of resource utilization and to achieve maximum throughput. Resources are basically CPU, memory and bandwidth in which proper usage of virtual machines and cloudlets are required. The top performing algorithms support to maximize the resource utilization and minimizing the total execution time along with providing QOS and improving load balance. Various performance parameters of scheduling algorithms are as mentioned below:

- Resource Utilization
- Processing Cost
- Makespan
- Memory Allocation and Utilization
- Average Waiting Time
- Average Response Time
- Energy Consumption
- Quality of Service
- Time Complexity

Scheduling algorithms used in cloud environment have their pros and cons. Table -2 below shows various scheduling algorithms.

Scheduling Algorithm	**Features**
First Come First Serve (FCFS):	Small waiting time, simple and easy to understand, Non-preemptive
Round Robin Algorithm	Equal load distribution among VMs, starvation free, length of time slice decides switching, preemptive
Modified Round Robin	Improved Round robin algorithm, burst time basis time slice.
Shortest Job First	Ordering of tasks as shorter task in the first in the queue and longer task at the end, Reduced waiting time, Can lead to starvation for longer jobs. [15]
Min-Min Heuristic Algorithm	Minimum Completion Time is the Base, Smaller Makespan, Unbalanced load. [5,7]
QoS Guided Min-Min Heuristic Algorithm	Tasks Classification on the Basis of Bandwidth, High Bandwidth Tasks Scheduled First.
Max-Min Heuristic Algorithm	Maximum Completion Time for all Tasks is the base, Minimum Completion Time is the base for individual Task, Large Tasks execution first leads to Non-Effective Load Balancing, Reduced makespan, and Response Time is Minimized. [12]
Improved Max Min Algorithm	Minimize Waiting Time in case of Short Tasks, Concurrent execution of Short and Long Tasks [5,9,11]
Resource Awareness Scheduling Algorithm	Hybrid Algorithm (Max-Min and Min- Min Algorithms), Max-Min strategy for Even Resources, Min-Min strategy for Odd Resources [8].
Reliable Resource Awareness Scheduling Algorithm	Highly Reliable Resources used for Tasks Allocation that leads to lower resource failure [8]
Particle Swarm Optimization Algorithm	Meta- heuristics Algorithm, Tasks are treated as Particles, Mapping between Tasks and Resources by way of Particles, Simple and Effective, Little Computation Overhead, Slow Convergence.
Ant Colony Optimization Algorithm	Find the Shortest Path using Ant food searching mechanism, Inner Parallelism in algorithm, Extensible, Stagnation Phenomenon leads to same solution sort out by all.
New Flower Pollination based Task Scheduling Algorithm	Nature-Inspired Algorithm, Global pollination and Local Pollination based Solution [6,10]
Bacteria Foraging Based Task Scheduling Algorithm	Bacterial Foraging based Mechanism, Optimal Heuristic Task selection using Chemotaxis, Usage of Fitness Function Health [4]

Agent-based scheduling	Dynamically Provisioning of Resources, Allocation of Real-Time Tasks, Manager Agent usage for scheduling the Tasks, Forward Announcement Information generation by Task Agents [14]

VI. Conclusion

Cloud computing emergence requires efficient scheduling and allocation of resources so that cloud consumers avail optimal computation solutions. Scheduling leads to the efficient resource utilization and better performance in cloud environment. Efficient use of scheduling algorithms helps in efficient cloud resources' utilization and management. This paper highlights various scheduling algorithms. These scheduling algorithms categorize as traditional scheduling algorithms used in operating systems and the modified forms of these traditional scheduling algorithms. Here is a study of several scheduling algorithms that assist in better resource provisioning and better load balancing. It is better to understand the efficiency and effectiveness of these scheduling algorithms so that enhanced modified or new scheduling algorithms can be developed for more efficient scheduling so that QOS can be ensured and throughput of the cloud computing can be enhanced.

References

[1] Selvarani, S., and G. Sudha Sadhasivam. "Improved Cost-Based Algorithm For Task Scheduling In Cloud Computing." Computational Intelligence And Computing Research (ICCIC), 2010 IEEE International Conference on. IEEE, 2010.
[2] O.M.Elzeki, M.Z.Reshad and M.A.Elsoud, "Improved Max-Min Algorithm in Cloud Computing", International Journal of Computer Applications, July, 2012.
[3] Saeed Parsa and Reza Entezari-Maleki, "RASA:A New Grid Task Scheduling Algorithm,", International Journal of Digital Content Technology and its Applications, 2009.
[4] Shridhar Domanal, Ram Mohana Reddy Guddeti and Rajkumar Buyya, "A A Hybrid Bio-Inspired Algorithm for Scheduling and Resource Management in Cloud Environment", IEEE TRANSACTIONS ON SERVICES COMPUTING, 2016.
[5] 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),2013.
[6] K Nishant et al., "Load balancing of nodes in cloud using ant colony", UKSim 14th int confcomput model simul, pp. 3-8, 2012.
[7] Kobra Etminani and Prof. M. Naghibzadeh, "A Min-Min Max-Min Selective Algorithm for Grid Task Scheduing", Third IEEE/IFIP International Conference on Internet, Uzbekistan, 2007.
[8] Sunilkumar Nakum, C. Ramakrishna and Amit Lathigara, "Reliable RASA Scheduling Algorithm for Grid Environment,",2014 IEEE International Conference on Computer Communication and Systems, 2014.
[9] Upendra Bhoi1and Purvi N. Ramanuj, "Enhanced Max-Min Task Scheduling Algorithm in Cloud Computing,", International Journal of Application or Innovation in Engineering Management, 2013.
[10] Santanu Dam, Gopa Mandal, Kousik Dasgupta and Paramartha Dutta, "An Ant Colony Based Load Balancing Strategy" in Advanced Computing Networking and Informatics-, Switzerland:Springer International Publishing, vol. 2, 2014
[11] Bhavisha Kanani, Bhumi Maniyar, "Review on Max-Min Task scheduling Algorithm for Cloud Computing", Journal Of Emerging Technologies And Innovative Research (JETIR), Volume 2, Issue 3, March 2015.
[12] Baldev Singh, "Resource Provisioning in Cloud Environment for Enhanced Quality of Service", International Journal of Latest Trends in Engineering and Technology (IJLTET), 2012
[13] HE XiaoShan, SUN XianHe and Gregor yon Laszewski,"QoS Guided Min-Min Heuristic for Grid Task Scheduling", Journal of Computer Science and Technology, 2003.
[14] Xiaomin Zhu, Chao Chen, Laurence T. Yang and Yang Xiang, "ANGEL: Agent-Based Scheduling for Real-Time Tasks in Virtualized Clouds", IEEE Transactions on Computers, 2015.
[15] Baldev Singh, "An Analysis of Cloud Computing Security Issues", International Journal of Scientific Research in Science and Technology, 2016.
[16] Xiao, Jing, and Zhiyuan Wang. "A Priority Based Scheduling Strategy for Virtual Machine Allocations in Cloud Computing Environment." Cloud and Service Computing (CSC), 2012 International Conference on. IEEE, 2012.
[17] TAREGHİAN, Shahab, and Zarintaj BORNAEE. "A new approach for scheduling jobs in cloud computing environment." Cumhuriyet Science Journal 36.3 (2015): 2499-2506.