International Journal of Latest Trends in Engineering and Technology
Special Issue SACAIM 2016, pp. 303-309
e-ISSN:2278-621X

A SURVEY ON RESOURCE SCHEDULING ALGORITHMS IN CLOUD ENVIRONMENT

Mithun D`Souza[1], Jassim Ibrahim[2], Ramnath Gaonkar[3] and Dr. S. Sathyanarayana[4]

Abstract- Cloud computing is a developing technology which provides on-demand user-requested services dynamically and efficiently. But nowadays the user demands to access more resources and it is difficult to manage and allocate cloud resources efficiently to satisfy user requirements. To allocate or manage resource efficiently, there is a need of effective resource allocation algorithms. In this paper, a detailed survey is done on various resource allocation algorithms i.e., Deadline Sensitive algorithm, Bat algorithm, Priority algorithm, Hungarian algorithm, Berger model, Modified Round-Robin algorithm and the comparative study is made on the basis of features and tools used.

Keywords—Cloud computing, Resource allocation, Resource allocation algorithms.

I. INTRODUCTION

Cloud Computing plays an important role in the modern computing system. These mainly ensures in the provision of computational services as well as resources like computing power, bandwidth, network, storage etc and also have on demand access to the same. NIST defines cloud computing as follows, "cloud computing refers to allowing ofuseful on-demand network access to a shared pool of resources (i.e. networks, servers, storage, applications and services) that can rapidly have provisioned and released with minimal management effort or service provider interaction" [1]. These shared resources, applications, infrastructures etc. are provided as a serviceto the end user over the network. The end users are not aware of the platform being used and from where they are gathering the service and where it is being stored [3].In cloud computing, there are mainly two types of cloud models i.e., Cloud Service model and Cloud Deployment model. Cloud Service Model includes SaaS (Software as a service),PaaS (platform as a service), and IaaS (infrastructure as a service).Thus cloud computing reduces the need to install the associated software and platform for a specific application andinfrastructure [2].**SaaS** is related to software or applications that the end user (customer) can access from the cloud. It provides a complete software application with a user interface[7]. **PaaS** is a platform with respect to deployment of application by the developers. Here the service provider

[1] *AIMIT, St Aloysius College Mangalore, Karnataka,India*
[2] *AIMIT, St Aloysius College Mangalore, Karnataka, India*
[3] *AIMIT, St Aloysius College Mangalore, Karnataka, India*
[4] *First Grade Women's College, Mysore, Karnataka, India*

providesdifferent environments to the user for development of applications and user can develop an application according to their requirements.**IaaS**, a bottom layer of cloud model provides resources with respect to computing power, storage and network according to user requirements7]. Cloud Deployment model includes private cloud, public cloud, hybrid cloud and community cloud.Private clouds can be accessed only by a specific entity i.e. a company or an organization [3]. It ensures maximum security and privacy.Public clouds can be used by the general public (end users) which is owned by large organization who offer services to the end users. These services and resources are set up by the cloud providers.People having internet connectivity are able to use public cloud anywhere and anytime. Cloud security is the main concerns the network is open to all the people [4]. Community clouds are used for the purpose of sharing cloud services in different organizations and companies[5]. An advantage is shared costs and convenient interplay between the groups. Hybrid cloud refers to the combination of one or more public, private and community cloud. They can move the data from one cloud to another [6].The service providers provide resources to the users without exposing the details like locations, hardware etc.

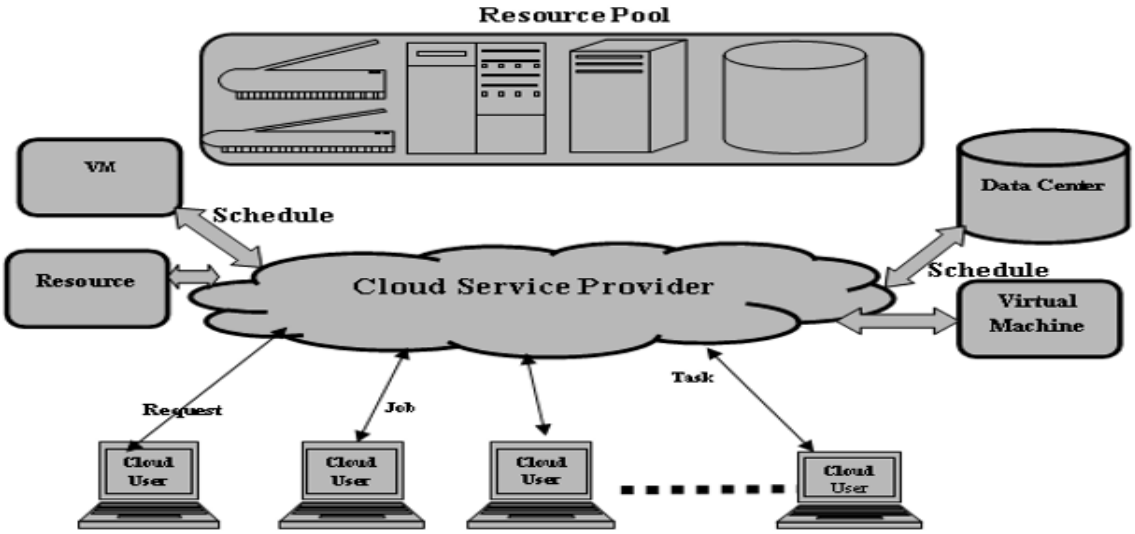

Figure: Basic Resource Management & Scheduling in Cloud Environment[12]

The above figure shows the basic Resource Management & Scheduling in Cloud Environment [12]. The resources available in the clouds are contained in the Resources Pool. VM is the Virtual Machine that schedules the resource which is being provided by the Service Provider. The users can request for resources from the cloud and the cloud service provider will provide the resources from the resource pool if it is available. If more resources are needed, then resource scheduling algorithms or methods must be used.

A. *CHARACTERISTICS OF CLOUD COMPUTING: -*

- Resource Pooling: Cloud Service Providers creates a pool of resources to share storage, bandwidth, networking capabilities etc., so that many users can dynamically acquire the

resources and release them when they no longer need for it. The user has no proper knowledge of the location of the resources which is being used [17].

- On-demand self-service: The provisioning of the resources to Cloud users is done without human interaction. This automated process reduces the work of the Cloud provider [15].

- Rapid Elasticity: The resources are provided to the cloud users at any time. If there is a need for more resources at a time, more resources are rapidly scheduled to provide for the rise in demand. Likewise, when there is less demand for resources, then the excess resources are quickly released[17]. This automated process decreases the acquired time for new computing capabilities when needed.

- Broad network access: The services provided by cloud service providers are available over the network and can be accessed through standard mechanisms that promote the use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops and workstations)[17].

- Measured service: Cloud systems automatically controls and makes better use of the resource by influencing a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth and active user accounts). The usage of the resource can be monitored, controlled and reported which provides transparency for the provider and consumer [17].

II. INTRODUCTION TO RESOURCE MANAGEMENT

In the context of cloud computing, resource management/allocation refers to the process of allocating computing power, storage, networking etc. to the end user according to their usage and requirements. For cloud users, the number of tasks must be completed on time with minimal cost [18].Hence there is a need for the cloud service providers to the schedule the resources to the users in an efficient manner. In some cases, a single resource can be allocated at the same time. So there is a need for checking of the reliability and availability and the load must be balanced among the resources of the same type[17].Resource Allocation Strategy (RAS) is the process of incorporatingthe number of activities for allocating and utilizing limited resources within the limit of the cloud environment so as to meet the requirements of the cloud application. In order to complete a user job, it requires the type and amount of resources needed by each application [18]. Normally when the service providers allocate resources to the users, they are done based on the user's requirements and requests. An important problem that must be addressed effectively in the cloud environment is how to manage Quality of Services (QoS) and maintain Service Level Agreement (SLA) for cloud users that share cloud resources [2]. The cloud users mainly focus on application performance, availability, the cost-effective scaling of available resources. The main aim of cloud computing is to provide resources with high availability, reliability and scalability in distributed environments [6]. This paper presents efficient resource allocation methods that will help the cloud provider to reduce wastage of resources and to achieve maximum profit. Efficient resource allocation in the cloud environment plays an important role in satisfying both the user's requirements as well as server's performance in an equal manner.

A. Resource Allocation criteria in Cloud:

Resource allocation Strategy(RAS) is a process where it assigns the available resources to the cloud applications which need them through web. Recourse allocation might starve services if the allocation is not done accurately. RAS deals with combining cloud provider activities for better utilization and allocating lacking resources within limited cloud environment to meet the needs of the cloud application.
An Optimal RAS must avoid the criteria's which are stated below:

a. Resource contention- this is the state when multiple applications try to access the same resource simultaneously.
b. Scarcity of resources -this is the situation where resources fall short.
c. Resource fragmentation -fault arises when the resources are isolated.
d. Over-provisioning of resources arises when the application gets surplus resourcesthan the demanded one.
e. Under-provisioning of resources occurs when the application is assigned with few numbers of resources than the demand[16].

III. LITERATURE SURVEY:

Resource allocation strategy in cloud computing tends to focus on allocating the resources efficiently according to the user requirements.

A. Deadline Sensitive based lease scheduling algorithm:

SasmitaParida[9] proposed Deadline Sensitive resource allocation algorithm where new leases (i.e. the number of nodes i.e. like CPU utilization and Main Memory, start time, duration and end time which is called deadline) are declared and scheduling is done whenever a new lease is adapted. This algorithm is implemented in Haizea. Normally, the user requests may arrive and leave at any time. A user can request for more than one resource at a time. Every resource must have a start time and given end time (deadline). This type of condition must be given to Haizea in the form of deadline sensitive lease. When a request with deadline sensitive lease comes to Haizea, it first tries to find out a single slot which can provide the required resources for required duration of time. If Haizea can find such a time slot, it accepts the newly arrived deadline sensitive lease. If it cannot find such a single slot, then Haizea reschedules already accommodated deadline sensitive and apply best effort leases. It finds out which leases can be rescheduled and then how they can be rescheduled. Lease that are having start time or end time greater than or equal to the start time of the newly arrived lease, are measured for being rescheduled. Once these leases are set, Haizea adds the newly arrived deadline sensitive lease to the list [10].Haizearetains a Lease object to stores lease's information after receiving the lease, and then the scheduler schedules the lease. There are several scheduling options in Haizeathat specify how it selects the resources and schedules leases.There is no assurance that a submitted best-effort lease will get the resources to complete the processing within a certain deadline. If the system is busy with lots of advance reservation and immediate leases, then best effort leases will not have enough resources to run on [4].

B. Bat algorithm:

Liji Jacob [10]proposed a Bat algorithm which is simulated by Bat behaviors.Bats are the mammals with wings. The three guidelines for the development of bat algorithm are- (1) Bats use echolocation to sense the distance and difference between food or prey and barriers even in the darkness. (2) Bats fly randomly to search prey with velocity, fixed frequency and loudness;

(3) the loudness varies from the large loudness to the minimum loudness.The best location of a bat to its prey indicates the quality of the solution. For evaluating the performance of this algorithm, two patterns are considered i.e. makespan and cost. The makespan indicates total execution time and is measured in seconds. The cost indicates cost per unit resources and is measured in dollars.

C. Priorityalgorithm:

K C Gouda[11] proposed a Priority algorithm in cloud enironment. This algorithm mainly decides which user requests must be given more priority among several user requests based on the restrictions like access time, amount of resource, type of the task etc. Here many users are submitting the job request that is they are requesting for the same resources. In case of high computational environment which mainly deals with scientific models like rainfall simulation, weather pridiction, monsoon prediction etc., requires huge amount of computing resources. Many users may be requesting these computational resources for scientific models/predictions. This will be a problem for cloud service provider as to whom to allocate the available resources among the requested users. In this situation, the proposed prioirity algorithm comes in handy which helps the cloud service provider to decide prioirity among the users and allocate the resources efficiently based on their requirements.

D. Hungarianbasedresourceallocationalgorithm:

Disha Patel [12] proposed Hungarian method for resource allocation. One of the problem in transportation problem is an assignment problem where the objective is to assign a number of resources to an equal number of tasks so as to minimize the total cost or maximize the total profit of allocation. To solve this problem, Hungarian method is used. The main objective of this method is to give an optimal solution of the assignment problem. In this algorithm, the priority of the task is considered as the length of the task. The task with longer length should have high efficient resources so that the execution time of the resources is balanced and the total execution time is also reduced. A table is created based on the value of data for solving the assignment problem. In cloud computing, a table is created based on the task and resource to solve an assignment problem. With respect to table, resource is denoted in columns and tasks are denoted in rows.

E. Job scheduling algorithm based on Berger model in cloud environment:

BaominXua[13] proposed the Berger model which is based on expectation states. The idea behind this model is two-fairness constraints job scheduling that is shown in cloud computing. This algorithm deals with the effective implementation of user tasks with advance fairness. In future work, the author plans on dealing with building a fuzzy neural network of ‘QoS feature vector of task’ and ‘Parameter vector of resource’ based on the non-linear mapping relationship between QoS and resource.

F. Modified Round-Robin algorithm:

PandabaPradhanet al [7] proposed Modified RR. This algorithm starts at the time of the first request and changes after the end of the first request.When a new user request is added into the ready queue, the algorithm calculates the average of requests in the ready queue and even the new arrival request. This needs two registers: (i)SR: To store the sum of the remaining burst time in the ready queue and (ii)AR: To store average of the burst time by dividing the value found in the SR register by the count of requests found in the ready queue. After the execution, the request will be removed from the ready queue once it finishes its burst time.Otherwise the

request will be moved to the end of the ready queue. SR will be updated by subtracting the time consumed by this request. Thenew data will be updated to AR. This algorithm is implemented in MATLAB.

IV.COMPARATIVE ANALYSIS

Table 1: Comparison of different Scheduling Algorithms

Sr.no	Algorithm	Author	Tools used	Features
1	Deadline Sensitive based lease scheduling algorithm	SasmitaParida, SuvenduChandanNayak[9]	Haizea	Maximization of resource utilizationand acceptance of leases compared to the existing algorithm of Haizea.
2	Bat algorithm for Resource Scheduling.	Liji Jacob [10]	BAT Algorithm based Scheduling Tool (BAST)	High Accuracy value, High Efficiency Value and High Convergence rate than other alogorithms. Makespan and total cost of resourcesare used for evaluating the performances.
3	Priority based resource allocation model for cloud computing.	K C Gouda, Radhika T V, Akshatha M [11]	CloudSim	Decides priority among different user request based on the parameters.
4	Hungarian Method Based Resource Scheduling algorithm	Disha Patel, Ms.JasmineJha[12]	CloudSim	This method finds the optimal solution for a given assignment problem.
5	Job scheduling algorithm based on Berger model in cloud environment	BaominXua, Chunyan Zhao b, EnzhaoHua, Bin Hu.[13]	CloudSim	Effective implementation of user tasks with advance fairness.
6	Modified Round Robin Algorithm for Resource Allocation.	PandabaPradhanPrafulla Ku. Behera B N B Ray	MATLAB	Reduces the turnaround and waiting time and the quantum is considered as dynamic.

V. CONCLUSION

In this paper, we have made a detailed survey on various kinds of algorithms. Each algorithm has their own way of scheduling the resources and respective features. As cloud computing is an emerging technology, they must be more advanced to focus on resource management. This paper

will help the future researchers to understand how the resources are managed in cloud and hopefully come up with the new ideas.

REFERENCES

[1] P.Mell and T.Grance, "The NIST definition of cloud computing (draft)", NIST special publication, vol.800,pp.145.

[2] V.P. Anuradha, D.Sumathi, "A Survey on Resource Allocation Strategies in Cloud Computing", ICICES2014 - S.A.Engineering College, Chennai, Tamil Nadu, India.

[3] Deepesh Kumar, Dr.AjayShankerSingh, "A Survey on Resource Allocation Techniques in Cloud Computing", International Conference on Computing, Communication and Automation (ICCCA2015).

[4] Srushti Patel, KrunalSuthar. "A Survey of Resource Allocation Policies in Cloud Computing". International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 10, ISSN: 2277 128X.

[5] SangeetaKumari, Ravi Kant Kapoor, Shailendra Singh, "A Survey on Different Techniques for Information Resource Scheduling in Cloud Computing", 2015 International Conference on Computational Intelligence and Communication Networks.

[6] Aditya Narayan Singh and Shiva Prakash, "Challenges and Opportunities of Resource Allocation in Cloud Computing: A Survey"

[7] Mohd Hairy Mohamaddiah, Azizol Abdullah, ShamalaSubramaniam, and MasnidaHussin, "A Survey on Resource Allocation and Monitoring in Cloud Computing", International Journal of Machine Learning and Computing, Vol. 4, No. 1, February 2014.

[8] AmitNathani , Sanjay Chaudhary, GauravSomani, "Policy based resource allocation in IaaS cloud" , Future Generation Computer Systems 28, Elsevier, 2012 ,pp 94–103.

[9] SasmitaParida, SuvenduChandanNayak, "Study of Deadline Sensitive Resource Allocation Scheduling Policy in Cloud Computing", International Journal of Computer Science and Mobile Computing, ISSN 2320–088X IJCSMC, Vol. 3, Issue. 12, December 2014, pg.521 – 528

[10] Liji Jacob, "Bat Algorithm for Resource Scheduling in Cloud Computing", International Journal of Research in Applied Science and Engineering Technology (IJRASET), Volume 2, Issue 4.

[11] K C Gouda, Radhika T V, Akshatha M, "Priority based resource allocation model for cloud computing", International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 1.

[12] Disha Patel, Ms.JasmineJha, " Hungarian Method Based Resource Scheduling algorithm in Cloud Computing", IJARIIE-ISSN(O)-2395-4396 Vol-2 Issue-3 2016

[13] BaominXu a, Chunyan Zhao b, EnzhaoHua, Bin Hu c,d, "Job scheduling algorithm based on Berger model in cloud environment", Science Direct, Advances in Engineering Software 42 (2011) 419–425.

[14] PandabaPradhan, Prafulla Ku. Behera, B N B Ray "Modified Round Robin Algorithm for Resource Allocation in Cloud Computing", International Conference on Computational Modeling and Security (CMS 2016), 2016

[15] Kuyoro, S. O., Omotunde, A. A., Ajaegbu, C., &Ibikunle, F. A. (2012), "Towards Building a Secure Cloud Computing Environment", International Journal of Advanced Research in Computer Science, 3(4), 166-171.

[16] V.Vinothina, Dr.R.Sridaran, Dr.PadmavathiGanapathi, "A Survey on Resource Allocation Strategies in Cloud Computing", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No.6, 2012

[17] Omotunde, Ayokunle. A, Okolie, S.O, Idowu, S. A, Ajayi, O.B, ".Resource Allocation in Cloud Computing – An Exposé", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 9, September 2015, ISSN: 2277 128X.

[18] VinayakAwasare, SudarshanDeshmukh, "Survey and Comparative Study on Resource Allocation Strategies in Cloud Computing Environment", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 2, Ver. I (Mar-Apr. 2014), PP 94-101.