International Journal of Computer Sciences and Engineering Open Access

Survey Paper Volume-2, Issue-10 E-ISSN: 2347-2693

A Survey of Task Scheduling Methods in Cloud Computing

R.Nallakumar[1], Dr.N.Sengottaiyan[2] and S.Nithya[3*]

[1,2,3*]*Department of Computer Science and Engineering ,*
Anna University Regional Centre Coimbatore, Tamilnadu, India

www.ijcaonline.org

Received: 22 Oct 2014 Revised: 10 Oct 2014 Accepted: 24 Oct 2014 Published: 31 Oct 2014

***Abstract*—** Cloud computing is one of the latest computing technology where the data's applications and the IT services are provided with the help of Internet. The Grid computing, distributed computing and virtualization are the concepts that helped to build the cloud computing. Cloud computing provides various computing resources to the cloud users based on pay-per-usage basis. Resource scheduling and Job/Task scheduling are the most important considerations in the cloud and in which, the jobs submitted gets allocated by the resources efficiently. The goal of the cloud computing service provider, is to use the efficient resources to obtain the maximum profit and this results the task scheduling a core and one of the challenging issues in the cloud environment. The Task scheduling is an NP-hard optimization problem and where the meta-heuristic algorithms are proposed to provide an optimal solution. The scheduling strategy of a good task scheduler should get adopted based on the changing environments and the type of task scheduling used.

***Keywords*—**Cloud computing,Task Scheduling,Resource allocation,Quality-of-Service,Load Balancing

I. INTRODUCTION

Cloud computing means that the applications and resources are moved over the Internet where the devices like servers, printers, central processing unit helps to access the services where that do not have any limitations [1]. The combination of storage places is nothing but the data center, which is identical to a database [2]. In a cloud, several applications will be running and it is stored in data centers and the data's may be accessed by the user at anytime anywhere with the help of an Internet connection. Cloud computing has some of the features like efficiency, fault tolerance, reliability, security, etc., among different networks [3].

II. CLOUD SERVICES

Cloud computing provides the types of services [1] like

- Software-as-a-Service (SaaS)
- Infrastructure-as-a-Service (IaaS)
- Platform-as-a-Service (PaaS)

With SaaS, where to access the resources that are hosted in the cloud there is a need to get an authorization from the cloud consumers, Example: Microsoft Office, Web apps.

With PaaS, the cloud user's purchase the platform to access the services that helps to develop their own applications in the cloud, Example: Window's Azure, Google app engine.

And in IaaS the bandwidth, network connectivity and resource expenses are managed without concentrating about the infrastructure, Example: Amazon EC2, Rackspace.

Were the types of cloud services associated with the clients in a cloud environment are shown in Figure.1

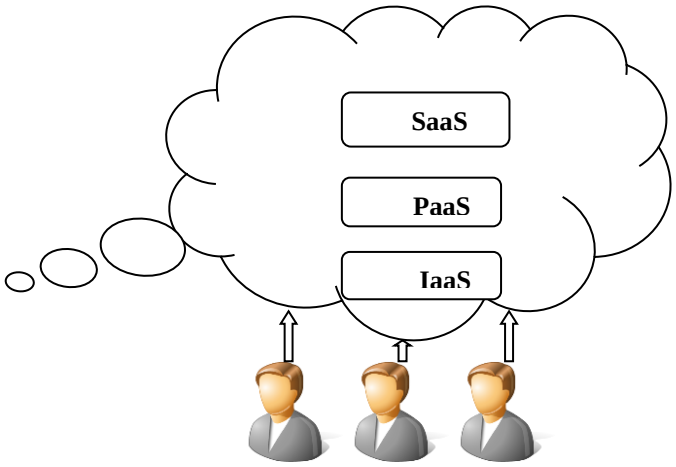

Figure.1 Types of Cloud Services

III. DEPLOYMENTS OF CLOUD

There are four different types of deployment models of cloud [4-11] and are listed below:

- Public cloud
- Private cloud
- Community cloud
- Hybrid cloud

Public Cloud

Corresponding Author: *S.Nithya*

The public clouds using service provider provides applications, resources and storage to the public based on the services like free or on a pay-per-use-basis. Example: Microsoft, Google.

Private Cloud

A private cloud is used mainly in a single organization and is managed internally or by a third party which is either hosted internally or externally. Example: Amazon EC2

Community Cloud

In the community cloud where the infrastructure is shared between several organizations based on concerns like security, compliance, which is managed internally or by a third party that is hosted internally or externally. Example: Government, Banks.

Hybrid Cloud

The composition of two or more models like a private cloud, public cloud or community cloud is the hybrid cloud. The applications and programs are moved from one deployment model to another in an easy manner if multiple cloud systems are connected. Example: Electronic mail.

Also the cloud deployment model is shown in Figure.2

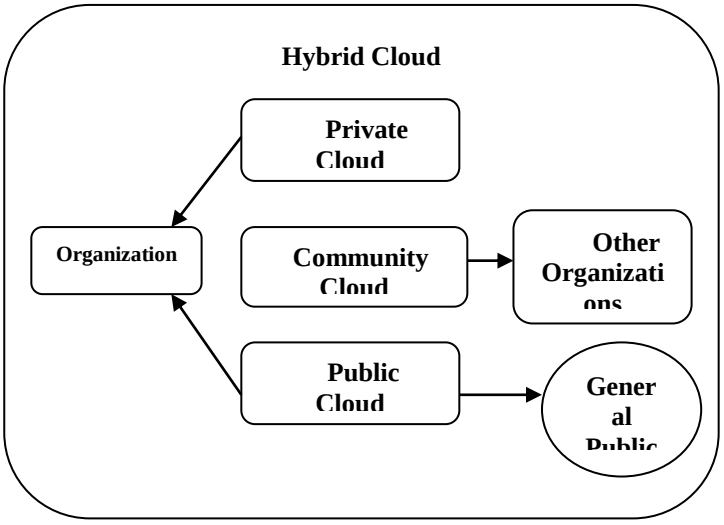

Figure.2 Cloud Deployment Model

IV. SCHEDULING IN THE CLOUD

Scheduling is the process of allocating resources based on budget constraints [12] to the needed jobs in the cloud environment and it helps to schedule the uploading and downloading files during transmission. The scheduling purpose is to increase the resource utilization. To avoid the circumstances of system hanging the application jobs and distribute the load network machines are to be scheduled properly if not several errors will occur [2].

The cloud computing services like storage, bandwidth and compute are available at lower costs [13] and the tasks are scheduled based on user requirements. Also, if the number of cloud users gets increased scheduling becomes quite difficult were a proper scheduling algorithm is to be used. At the initial stage the scheduling algorithms are implemented [14] [15] in grid computing and because of its reduced performance there occurs a way to move towards cloud computing.

In the cloud, the scalability of resources allows the users to meet the application requirements and this helps the workflow management system to meet the Quality-of-Service (QoS) [16]. Scheduling process [17] is classified into three stages in the cloud and they are:

Discovering and Filtering of Resources:
The resources presented in the network system are discovered by the data center broker and the status information related to it will be collected.

Selection of Resources:
Based on the parameters like task and resources the target resource is selected and is called deciding stage.

Submission of Tasks:
To the selected resource the task is submitted.

V. CLASSIFICATION OF SCHEDULING ALGORITHM

The cloud scheduling algorithms [12] are classified into groups as follows:

- Batch Mode Heuristics Algorithm (BMHA)
- On-line Mode Heuristics Algorithm (OMHA)
- Dependency Mode Heuristics Algorithm (DMHA)

In a Batch Mode Heuristic Algorithm when the jobs are arriving in the system, they get queued and collected into a set. Also, after a fixed period of time the scheduling starts. Example: First-Come-First-Serve, Round Robin, Shortest Job First scheduling, Minimum-Minimum, Maximum-Minimum.

The jobs are immediately scheduled when they arrive at the system in an On-line Mode Heuristic Algorithm. Example: Most-Fit Task Scheduling Algorithm.

Depending on the previous one, the jobs or resources are processed in Depending Mode Heuristic Algorithm. Example: Genetic Algorithm, Multi-Objective Genetic Algorithm.

VI. TASK SCHEDULING AND ALLOCATION OF RESOURCES

A Task is carried out like an action that takes the input as resources and produces the output efficiently in the computation nodes [1]. Task scheduling algorithms aim is to maximize the resource utilization in an efficient manner by minimizing the execution of tasks.

Cloud allocates the resources based on pay-per-use basis and the demand to increase the flexibility [18]. The scheduling of a task in different environments under different circumstances may differ in cloud computing because of the algorithms used for the resource allocation.

In cloud, where the internal and external requirements of the resources are maintained and the requirements like bandwidth, storage, resource expenses, response time may differ for each task.

Load balancing, scalability, reliability, performance and dynamic re-allocation of resources to the computing nodes are all the major problems that occur in a task scheduling. Hence, an efficient scheduling algorithm is used for task scheduling in the cloud computing environment.

The overview of task scheduling in a cloud environment is shown in Figure. 3

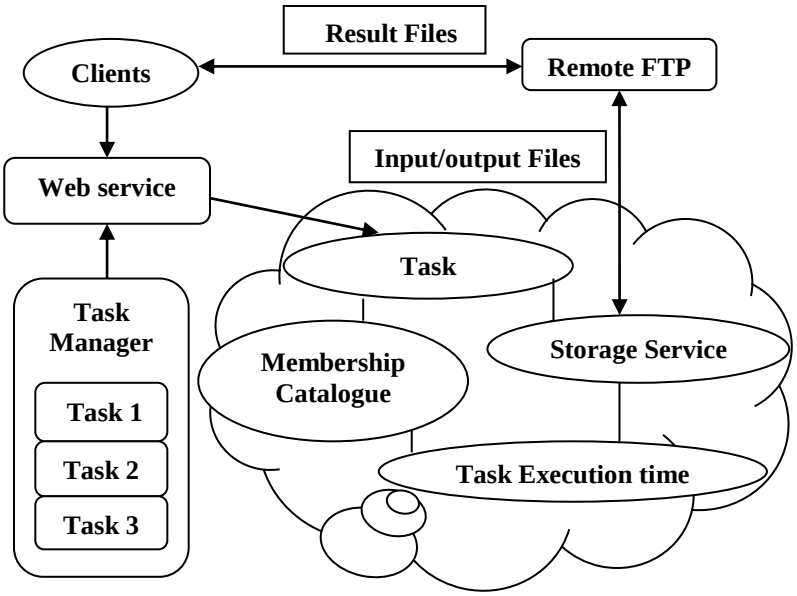

Figure. 3 Overview of Task Scheduling in a Cloud Environment

VII. TYPES OF TASK SCHEDULING

Cloud Service Scheduling:

Cloud Service scheduling is based on user level and system level [19]. The service provision problems raised between the providers and customers are done at user level scheduling.

The managing of resources with data center involves the system level scheduling. Several physical machines stay connected are said to be data center. The user sends millions of tasks and the tasks are assigned to the physical machines in the data center.

User-Level Scheduling:

For regulating the supply and demand of cloud resources the market based and action based schedulers are suitable. In case of market based schedulers where the resource is virtualized and delivered to the user as a service.

Service Level Agreement (SLA) includes service provisioning in the cloud, where the SLA indicates the contract to be signed among the providers and customers stating the agreement terms, including non-functional requirements specified by Quality-of-service (QoS), obligation and in penalties [20].

The SLA is used by novel cloud scheduling [21] having trust monitor to provide the scheduling faster with secure processing because of the over flooding user request. The novel approach is considered by two variants of heuristics [22] and they are stimulated Annealing and GI-FIFO scheduling.

Static and Dynamic Scheduling:

In the static scheduling where the required data are pre-fetched and the task execution at different stages are pipelined. And in dynamic scheduling where the task or job components are not known beforehand and whereas the task execution time may not be known.

The scheduling strategies in three tier cloud structure based on service request are

- Resource Providers
- Service Providers
- Consumers

A tremendous amount of energy is wasted by the large number of cloud computing servers and it emits a carbon dioxide, therefore the green task scheduling [23] [24] is needed to lower the energy usage and to reduce the pollution.

Heuristic Scheduling:

The NP hard problems are solved by enumeration, heuristic and approximation methods. If all possible solutions are enumerated and compared one by one, then the optimal solution is selected in an enumeration method. In case of large instances, where an enumeration is not feasible the heuristic handles the instances of finding the best solutions and reasonably fast.

To optimize the solution, the approximate solutions are found in case of approximation algorithms. The makespan of the given task is minimized while balancing the entire system load according to the load balancing task schedule [25].

These tasks are scheduled for some predefined time in batch mode. The actual execution time of a large number of tasks is known by the batch-heuristics. Also the heuristics Max-Min [26], Min-Min [27] are used by the batch mode scheduling.

Real Time Scheduling:

The real time scheduling involves some objectives and they are used to increase the throughput instead of meeting deadlines, also the average response time will be minimized.

To maximize the total utility [28] the real time tasks should be scheduled non-preemptively. The profit Time Utility Function (TUF) and penalty Time Utility Function (TUF) are two different utility functions associated at the same time for every task.

Workflow Scheduling:

In a Workflow scheduling where the applications are structured in the form of Directed Acyclic Graphs (DAG) [29]. Here the edges denote the task dependencies [30] and the nodes represent the constituent task.

A single workflow has some set of task where each task is communicating with another task in the workflow. In the workflow execution management the workflow scheduling is one of the major issues and in the cloud it is enabled to use various cloud services to facilitate the workflow execution.

VIII. CONCLUSION

Task Scheduling is one of the important scheduling mechanism in cloud computing. Task scheduling algorithm is designed in such a way to support the cloud infrastructure and a large number of applications. In a cloud environment the task scheduling aim is to minimize the execution time of tasks and to maximize the resource utilization. This paper surveyed about cloud computing, cloud services, and the deployment models of the cloud. Also the scheduling concepts, need of scheduling in cloud environment and the scheduling algorithm classification are explained. Task scheduling, resource allocation and the task scheduling types were also discussed. The major problems occurred in the task scheduling are load balancing, response time, bandwidth, memory storage and resource expenses which can be resolved by using efficient scheduling algorithms. In order to obtain the Quality-of-Service (QoS), an efficient scheduling algorithm is to be used to enhance the reliability and scalability in the cloud environment.

REFERENCES

[1]. M. Gokilavani, S. Selvi and C. Udhayakumar, "A Survey on Resource Allocation and Task Scheduling Algorithms in Cloud Environment", International Journal of Engineering and Innovative Technology (IJEIT), Volume 3, Issue 4, ISSN: 2277-3754, October 2013.

[2]. Sunny Kumar, Shivani Khurana, "Analysis of different Scheduling Algorithms under Cloud Computing",International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5, Issue 2, ISSN:0975-9646, 2592-2595,2014.

[3]. Monika Choudhary, Sateesh Kumar Peddoju, "A Dynamic Optimization Algorithm for Task Scheduling in Cloud Environment", International Journal of Engineering Research and Applications (IJERA), Vol. 2, Issue 3, May-Jun 2012, pp. 2564-2568.

[4]. "Cloud computing resources", http://www.cloud9s.net/cloudcomputingresources.html

[5]. http://www.thecloudcomputing.org/2011

[6]. http://salsahpc.indiana.edu/CloudCom2010/papers.html

[7]. http://www.itworld.com/internet/69141/5-coolcloud-computing-research-projects

[8]. http://en.wikipedia.org/wiki/Cloud_computing

[9]. http://www.thecloudcomputing.org/2012/history.html

[10]. http://www.whereisdoc.co

[11]. Torry Harris' "Cloud Computing – An Overview"

[12]. Maheswari. R and S. Selvi, "A Survey on Scheduling Algorithms in Cloud Computing", International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue 10, ISSN: 2278-0181, October – 2013.

[13]. Sujit Tilak, and Prof. Dipti Patil, "A Survey of Various Scheduling Algorithms in Cloud Environment", International Journal of Engineering Inventions (IJEI), Volume 1, Issue 2, ISSN: 2278-7461, September 2012, PP: 36-39

[14]. 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.

[15]. Nithiapidary Muthuvelu, Junyang Liu, Nay Lin Soe, Srikumar Venugopal, Anthony Sulistio and Rajkumar Buyya, "A Dynamic Job Grouping-Based Scheduling for Deploying Applications with Fine-Grained Tasks on Global Grids", in Australasian Workshop on Grid Computing and e-Research (AusGrid2005), Newcastle, Australia., Conferences in Research and Practice in Information Technology, Vol. 44.

[16]. Meng Xu, Lizhen Cui, Haiyang Wang, Yanbing Bi, "A Multiple QoS Constrained Scheduling Strategy of Multiple Workflows for Cloud Computing",in 2009, IEEE International Symposium on Parallel and Distributed Processing.

[17]. Dr. Amit Agarwal and Saloni Jain, "Efficient Optimal Algorithm of Task Scheduling in Cloud Computing Environment", International Journal of Computer Trends and Technology (IJCTT), Volume 9 Issue7, March 2014.

[18]. Mrs. S. Selvarani, Dr. G. Sudha Sadhasivam, "Improved Cost - Based Algorithm For Task Scheduling In Cloud Computing", IEEE, 2010.
[19]. Fei Teng, "Resource allocation and scheduling models for cloud computing", Paris, 2011.
[20]. Emeakaroha, V.C., Brandic, I., Maurer, M. And Breskovic, I., "SLA-Aware Application Deployment and Resource Allocation in Clouds", IEEE, 2011.
[21]. Daniel, D., Lovesum, S.P.J., "A novel approach for scheduling service request in cloud with trust monitor", IEEE, 2011.
[22]. Boloor, K., Chirkova, R., Salo, T., Viniotis, Y., "Heuristic-Based Request Scheduling Subject to a Percentile Response Time SLA in a Distributed Cloud", IEEE, 2011.
[23]. Mehdi, N.A., Mamat, A. Amer, A., Abdul-Mehdi, Z.T., "Minimum Completion Time for Power-Aware
a. Scheduling in Cloud Computing", IEEE, 2012.
[24]. Luna Mingyi Zhang, Keqin Li, Yan-Qing Zhang, "Green Task Scheduling Algorithms with Speeds Optimization on Heterogeneous Cloud Servers", IEEE, 2011.
[25]. Xin Lu, Zilong GU, "A load-adaptive cloud resource scheduling model based on ant colony algorithm", IEEE, 2011.
[26]. Gao Ming and Hao Li, "An Improved Algorithm Based on Max-Min for Cloud Task Scheduling", Yunnam University, China, 2011.
[27]. Ching-Hsien Hsu, Tai-Lung Chen, "Adaptive Scheduling Based on Quality of Service in Heterogeneous Environments", IEEE, 2010.
[28]. Shuo Liu, Gang Quan, Shangping Ren, "On-Line Scheduling of Real-Time Services for Cloud Computing", IEEE, 2010.
[29]. J. Yu and R. Buyya, "Workflow Scheduling Algorithms for Grid Computing", Technical Report, GRIDS-TR-2007-10, Grid Computing and Distributed Systems Laboratory, The University of Melbourne, Australia, May 2007.
[30]. Anju Bala, Dr. Inderveer Chana, "A Survey of Various Workflow Scheduling Algorithms in Cloud Environment", 2nd National Conference on Information and Communication Technology (NCICT), 2011, Proceedings published in International Journal of Computer Applications® (IJCA).

AUTHORS PROFILE

Mr. R. Nallakumar received the Bachelor Degree in Computer Science and Engineering in 2009. He received the Master degree in Computer Science and Engineering in 2011. He also completed Master of Business Administration and currently he is pursuing his Ph.D. He is working as an Assistant professor at Anna University Regional Centre, Coimbatore, Tamilnadu, India. His area of interest is Cloud Computing.

Dr. N. Sengottaiyan received the Bachelor Degree in Electronics and Communication Engineering in 1986. He received the Master Degree in Computer Science and Engineering in 2004. He has received his Ph.D in 2011. He is currently pursuing P.D.F from California South University, California,U.S.A. He is working as a Principal at Indira Institute of Engineering and Technology Thiruvallur, Chennai, Tamilnadu, India. His area of interest are Cloud Computing and Wireless Sensor Networks.

Ms. S. Nithya received the Bachelor Degree in Computer Science and Engineering in 2012.She is currently pursuing her Master Degree in Software Engineering from Anna University Regional Centre, Coimbatore, TamilNadu, India. Her area of interest is Cloud Computing.