IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 6, June 2015.
www.ijiset.com
ISSN 2348 – 7968

A Survey of Scheduling Mechanisms in Cloud

Vinay Kumar[1], M. Sashi Kumar[2] Dr. M Naga Ratna[3]
[1, 2] Asst. Professor(s), Department of CSE, Vasavi College of Engineering, Telangana, India
[3]Asst. Professor, JNTUH College of Engineering
s.vinaykumar@staff.vce.ac.in

Abstract
The Cloud Computing environment has brought unprecedented computing capacity. Taking advantage of this complex infrastructure requires efficient middleware to support the execution of various distributed applications, comprising of a complex set of subtasks, for optimal performance. This presents the challenge of resource allocation and scheduling these subtasks in shared heterogeneous systems. In order to maximize resources utilization, many scheduling algorithms were introduced, analyzed and implemented. In this paper we study different scheduling algorithms widely used in the cloud.
Keywords: *Cloud Computing, Scheduling, Distributed System, Task Scheduling*

1. Introduction

NIST consolidated and defines cloud computing as: "a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., network, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction". Cloud computing has emerged as a new paradigm leveraging distributed computing in delivering infrastructure, platform, and software as services. These services are made available to the customers or consumers based on subscription in a pay-as-you-go model. The main characteristics of Cloud are

1. On-demand self-service
2. Broad network access
3. Resource pooling
4. Rapid elasticity
5. Measured service.

Cloud computing allows user by dynamically provisioning requisite number of compute resources at specified locations as and when required. Also, vendors allow applications to choose the storage locations to host their data at global locations.

In order to effectively and efficiently schedule the tasks and data of applications onto cloud computing environments, schedulers have different policies that vary according to different objectives like minimize total execution time, minimize total cost of execution, balance the load on resources used along with meeting the deadline constraints of the application.

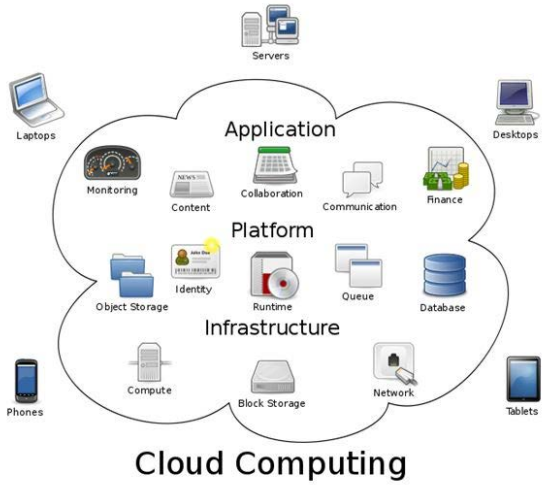

Fig.1: Cloud Computing

This technology is still in its early stages and the primary areas of focus include:

1. Resource Management
2. Task Scheduling
3. Power Management
4. Security
5. Standardization
6. Bandwidth & Latency
7. Legal Considerations

Task scheduling and provision of resources have been the concerns since the evolution of cloud computing.

In this paper, we focus on different scheduling algorithms in cloud that spread the load on processors and maximize their utilization while minimizing the total execution time of the task. The main purpose is to schedule tasks to the resources in accordance with time, which involves finding appropriate sequence in which jobs can be executed under various constraints.

ISSN 2348 – 7968

2. Purpose of Scheduling

Cloud computing has recently received a wide attention as an approach for delivering Info and Communication Technologies (ICT) services as a utility. In the mechanism of providing these services it is necessary to improve the utilization of data center resources which are operating in most dynamic workload environments. In a single data center generally hundreds and thousands of virtual servers run at any instance of time, hosting of many tasks at the same time on the cloud system keeps receiving the batches of task requests. During this context we have to notice, out many powered on servers only few target servers, which can fulfil a batch of incoming tasks. Therefore task scheduling is an issue that influences the performance of cloud service provider. Scheduling is a mechanism to determine which task to be executed on a resource and in which order. The primary concerns before solving the problems using the techniques of scheduling have to address these:

- Fair resource allocation – Scheduling is done so that allocation of resources is done in fair manners.
- QOS – Resources and jobs are scheduled in such a way so that quality of services is achieved.
- Resource utilization - It is the degree to which the resources of the system are utilized. A good scheduling algorithm provides maximum resource utilization.
- Energy consumption – It is the degree to which the resources of system are consumed. A good scheduling algorithm saves energy consumption

Scheduling in cloud computing involves three steps
a) Discover a resource
b) Select a target resource (Decision stage).
c) Assigning a task to a target resource

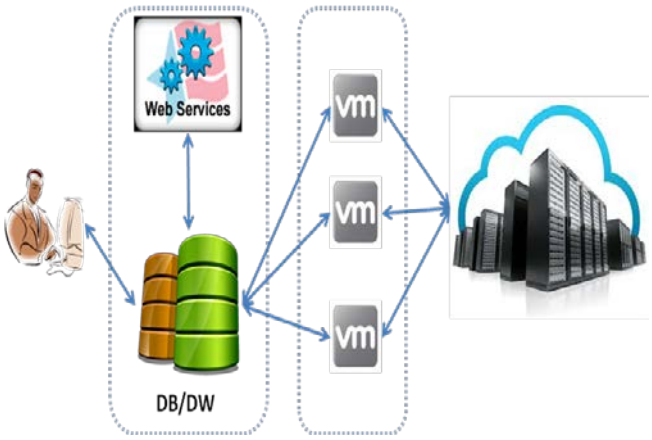

Fig.2 Cloud Server

The objective of scheduling is to utilize the resources along with managing the load between the resources to execute in minimum time.

3. Scheduling Mechanisms

This section describes various scheduling algorithms that are currently implemented in cloud. The motivation to all these traditional and heuristics scheduling algorithms is to ensure proper utilization of resources.

Shortest Job Scheduling [1]: this scheme is implemented on multiple clouds in both the under load and the over load conditions with request from the user with certain parameters defined such as arrival, process time, deadline and the input output requirement of the processes. Here the scheduling of the processes is performed with respect to the memory requirements.
Algorithm:

```
for i = 0 to i < main queue-size
        if taski+1 length < taski length then
                add taski+1 in front of taski in the queue
        end if
        if main queue-size = 0 then
                taski last in the main queue
        end if
end for
```

Preemptable Shortest Job Next Scheduling Algorithm [2]: This algorithm is run on a private cloud. This employs pre-emption technique of Round-Robin algorithm combined with Shortest Process Next to obtain cost benefits and improve the response time and execution time.

Optimized Activity Based Costing Algorithm [3]: Proposed an Activity based costing a mechanism which evaluates every task separately to give the cost. It is decided based on the space and time taken by every activity of every task and resources. Activity based costing is a way to measure both cost of the object and its performance. It helps to solve the problems like poor cost control and also starvation.

Improved Cost-Based Algorithm for Task Scheduling [4]: Proposed an improved cost-based scheduling algorithm by grouping the tasks according to the processing capabilities of available resources and make appropriate mapping tasks to resources in cloud. The strategy is for cloud environment where there may be no relation between the overhead application and the way different tasks cause overhead cost of resources. This scheduling algorithm measures both cost of the resource and computational

ISSN 2348 – 7968

performance, and it also improves the computation/communication ratio.

Min-Min Scheduling Algorithm [5]: In this paper an improved load balanced algorithm is introduced on a base of Min-min algorithm taking in to account priority given by the user in order to minimize the make span and maximize the utilization of resource.

Scheduling algorithm based on QoS [6]: Proposed an algorithm that takes in to account quality of service. It computes the tasks priority based on different attributes of tasks and after that apply sorting on tasks onto a service which can further complete the tasks.

Ant Colony Optimization [7]: Ant colony optimization is to simulate the foraging behaviour of ant colonies. Ants generally form a group and search for food during which they use a pheromone (chemical) to communicate with each other.

GASA (Grid Advance Reservation API) is a subsystem of Globus project. It provides a mechanism for resource reservation so that applications can receive a certain level of service from a resource.

Condor system [8] provides a match making mechanism to allocate resources with ClassAds (Classified Advertisements).

Legion System also supports resource reservation. It focuses on providing basic mechanisms for building application level scheduling algorithms rather than construct scheduling algorithm itself. A simple random selection policy is provided as the default scheduling mechanism.

The scheduling algorithm Sun Grid Engine [9] offers a scheduling policy for distributed jobs that allow priority to be assigned to users or groups. It also contains a model of deadline for job execution. However, in this system, deadlines are defined in terms of start time of the job.

A Compromised-Time-Cost Scheduling Algorithm [10]: Presents a novel Compromised-Time-Cost scheduling which considers the characteristics of cloud to accommodate instance-intensive cost-constrained workflows by compromising execution time and cost with user input enabled on the go. The simulation results demonstrated show that Compromised-Time-Cost algorithm achieves lower cost than others while meeting the user-designated deadline or reduce the mean execution time.

A Particle Swarm Optimization based Heuristic for Scheduling Workflow Applications [11]: Presents a particle swarm optimization that is based on heuristics to schedule applications taking in to account computation cost and data transmission cost to cloud resources. It can used for workflow applications by varying its computation and communication costs. The experimental result have shown that it can achieve cost savings and better distribution of workload onto resources.

Resource-Aware-Scheduling algorithm (RASA) [10]: a new task scheduling algorithm RASA is composed of two scheduling algorithms Max-Min and Min-Min. RASA uses the benefits of both Max-Min and Min-Min algorithms. RASA doesn't consider the deadline of each task, arrival rate of the tasks, cost of the task execution on each resource and cost of the communication. The experimental results show that RASA performs better than the existing scheduling algorithms in large scale distributed systems.

Innovative Transaction Intensive Cost-Constraint Scheduling Algorithm [12]: A scheduling algorithm that takes in to account cost and time. The simulation results demonstrated by the author show that this algorithm can achieve lower cost than others while satisfying the user designated deadlines.

Scalable Heterogeneous Earliest-Finish-Time Algorithm (SHEFT) [13]: SHEFT workflow scheduling algorithm schedules a workflow elastically in Cloud environment. The experimental results of SHEFT show that it outperforms several workflow scheduling algorithms in optimizing workflow execution time. It also enables resources to scale elastically at runtime.

Multiple QoS Constrained Scheduling Strategy for Multi-Workflows (MQMW) [14]: Takes in to account multiple workflows and multiple QoS parameters. This scheduling strategy has increased the access rate by minimizing the make span and cost of workflows on cloud platform.

4. Comparison

The following table compares various scheduling algorithms based on different parameters and environment.

Table -1: Comparison of Scheduling Algorithms

Scheduling Algorithm	Scheduling Parameters	Objective	Tool	Scheduling Factor	Environment

ISSN 2348 – 7968

Shortest Job scheduling	Arrival time, process time, deadline and I/O requirement	Effective resource allocation under defined parameters	MATLAB	Group task	Cloud environment
Priority Shortest Job Next	Cost and time	Effective and fast execution of task	Private cloud	Group task	Cloud environment
Optimized ABC Algorithm	Cost, profit and priority	Measure the cost and performance more accurately	Sim Grid	Array of task	Cloud Environment
Improved Cost Based algorithm	Cost and task grouping	Minimizing the cost and completion time	Cloudsim	Group task	
User-Priority Guided Min-Min scheduling Algorithm	Makespan	To promised the guarantee regarded the provided resources.	MATLAB	Independent Task	Cloud Environment
Ant Algorithm	Pheromone updating rule	Enhance the performance of basic ACO	Cloudsim	Independent task	Cloud environment

5. Conclusion

Scheduling is an important activity in cloud computing environment. This paper presented various scheduling algorithms in Cloud. The appropriate scheduling technique with respect to the mechanisms discussed seemed to be heuristic technique.

References

[1] Poonam Devi, Trilok Gaba, "Implementation of Cloud Computing By Using Short Job Scheduling", International Journal of Advanced Research in Computer Science and Software Engineering, Vol.no.03, Issue 7, pp 178-183, July 2013.

[2] Nishant.S.Sanghani, R.J. Khimani, K.K. Sutaria, Pooja. P. Vasani, "" Journal of Information, Knowledge and Research in Computer Engineering, Vol no. 02, Issue- 02, pp 385-388, Oct 2013.

[3] Qicao, Zhi-Bo Wei , Wen- Mao Gong, "An Optimized Algorithm for task Scheduling Based on Activity Based Costing in Cloud Computing" Bioinformatics and Biomedical Engineering , pp 1-3, 11-13 June 2009 (IEEE).

[4] S.Selvarani, G. Sudha Sadhasivam, "Improved cost-Based Algorithm for Task Scheduling in Cloud Computing" Computational Intelligence and Computing Research (ICCIC), pp 1-5, 28-29 Dec 2010 (IEEE).

[5] Abramson D., Sosic R., Giddy J. and Hall B., "Nimrod: A Tool for Performing Parameterized Simulations using Distributed Workstations", The 4th IEEE Symposium on High Performance Distributed Computing, Virginia, August 1995.

[6] Henri Casanova and Jack Dongarra, "NetSolve: A Network Server for Solving Computational Science Problems", The International Journal of Supercomputer Applications and High Performance Computing, Volume 11, Number 3, pp 212-223, Fall 1997.

[7] Umarani Srikanth G. , V. Uma Maheswari , P. Shanthi , Arul Siromoney , "Tasks Scheduling using Ant Colony Optimization", Journal of Computer Science, Vol. 8, pp 1314-1320, Science Publications, 2012.

[8] Michael Litzkow, Miron Livny, and Matt Mutka, "Condor - A Hunter of Idle Workstations", Proceedings of the 8th International Conference of Distributed Computing Systems, pp 104-111, June, 1988.

[9] Gentzsch, W., "Sun Grid Engine: Towards Creating a Compute Power Grid", In: Proceedings of the 1st International Symposium on Cluster Computing and the Grid (CCGrid'01), Brisbane, Australia May 2001.

[10] 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, 37, pp. 573-595, 1995.

[11] Suraj Pandey, LinlinWu, Siddeswara Mayura Guru and Rajkumar Buyya, "A Particle Swarm Optimization-based Heuristic for Scheduling Workflow Applications in Cloud Computing Environments", 2010.

[12] Y. Yang, K. Liu, J. Chen, X. Liu, D. Yuan and H. Jin, "An Algorithm in SwinDeW-C for Scheduling Transaction-Intensive Cost-Constrained Cloud Workflows", Proc. of 4th IEEE International Conference on e-Science, 374-375, Indianapolis, USA, December 2008.

ISSN 2348 – 7968

[13] Cui Lin, Shiyong Lu, "Scheduling Scientific Workflows Elastically for Cloud Computing" in IEEE 4th International Conference on Cloud Computing, 2011.
[14] Meng Xu, Lizhen Cui, Haiyang Wang, Yanbing Bi, "A Multiple QoS Constrained Scheduling Strategy of Multiple Workflows for Cloud Computing", IEEE International Symposium on Parallel and Distributed Processing, 2009.

First Author Received M.Tech. (CSE) degree in 2011 from JNTU, Hyderabad. Currently working as Asst. Professor in CSE department, Vasavi College of Engineering, Hyderabad. His research interests include Cloud Computing, Data Science and Internet of Things.

Second Author Pursuing Ph.D. in the area of Cloud Security from Osmania University and have received Masters of Technology from JNTU in 2008. He is working as Assistant Professor in the department of CSE at Vasavi College of Engineering. His research interests include Cloud Computing, Information Security, Internet of Things, Sensor Networks and Storage Management.