Volume 6, Issue 5, May 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Survey on Task Scheduling Optimization in Cloud Computing

Er. Mohinder Singh, Er. Ritu Marken
Department of Computer Science, Kurukshetra University,
Haryana, India

Abstract— Cloud computing is an emerging field, which enables one to achieve the aforesaid goal, leading towards increases business achievement. Cloud computing comes into center of attention immediately when you think about what IT continually needs: a means to increase capacity or add abilities on the fly without investing in new infrastructure, training current human resources, or licensing new software. The cloud should provide resources on demand to its clients with great availability, scalability and with low cost. Cloud computing environments provide dynamically manageable visualized computing resources. The management of these resources is taking big account in cloud. So the scheduling has of vital concern in cloud computing. In this paper, we give an elaborate idea about Genetic Algorithm and its several variants recommended for task scheduling in cloud environment and idea of GA based scheduler is proposed in which population is produced by enlarged Max Min by which makespan can be reduced and load of resources can be balanced.

Keywords— Cloud Computing, Make span, Task Scheduling, Genetic Algorithm (GA), Efficient GA, Load Balancing.

I. INTRODUCTION

Cloud computing is a trendy model in IT World, and provide lots of services instantly. It becomes popular in day to day life, and also becoming more challenging and developing field of computing. This is an emerging concept that has several computers interconnected through a real-time network like Internet, and it makes better use of multiple distributed resources that can be shared by the users as per requirements. It provides virtual resources that are dynamically scalable. It defines virtualized resources, software, platforms, applications, computations and storage to be scalable and provided to users directly on payment for only what they use[1].

Cloud is altering our life by providing users with new kind of services without paying attention to the details. It is high quality software having the ability to change the IT software Industry and making the software even more interesting. Hence it helps any organization in avoiding the capital costs of software and hardware.

Cloud ecosystem gives three main entities: Cloud Consumers, Cloud Service Providers, and Cloud Services. Cloud Consumers utilizes Cloud Services provided by Cloud Service Provider. These services may be presented by the service provider's own infrastructure or on the third party cloud infrastructure providers[2].

Cloud Computing Models

Cloud computing model is consist of three service models and four deployment models as shown below in Figure 1. Mainly three types of services are provided by the cloud. First is Software as a Service (SaaS), which brings the software to the users; so users don't need to install the software on their own machines and they can use the software directly from the cloud. Second is Platform as a Service (PaaS), which gives the platform to the clients so that they can prepare their applications on this platform. Third is Infrastructure as a Service (IaaS), which gives cloud users the infrastructure for various reasons,like the storage system and computation resources. Deployment Models are classified just as Public Clouds, Private Clouds, Community Clouds and Hybrid Clouds. This huge combination of services and sources, shared through clients on subscription basis needs a serious attention in terms of tasks scheduling, resource allocation and resource sharing. Cloud service providers are energy efficiency and bandwidth
management is another important concern. If seen by outsider's way, a cloud environment processes the tasks submitted by clients. Any concurrent access to resources needs to be addressed with objectives of improved resource utilization, reduced energy expenses

Cloud Computing Characteristics:

- On demand self services

Computer values like email, applications, network or server service can be presented without requiring human cooperation with each service provider. Cloud service providers providing on demand self services comprises Amazon Web Servicing (AWS), MS, Google, International Business Machines(IBM and Salesforce.com. New York Times and NASDAQ are examples of companies using AWS (National Institutes of Standards and Technoloy). Gartner defines this characteristic just as service based.

- Broad network connection

Cloud Capabilities are available over the network and accessed through standard mechanisms that further use by heterogeneous thin or thick client platforms such as mobile phones, laptops and Personal Data Assistant(PDAs).

© 2016, IJARCSSE All Rights Reserved

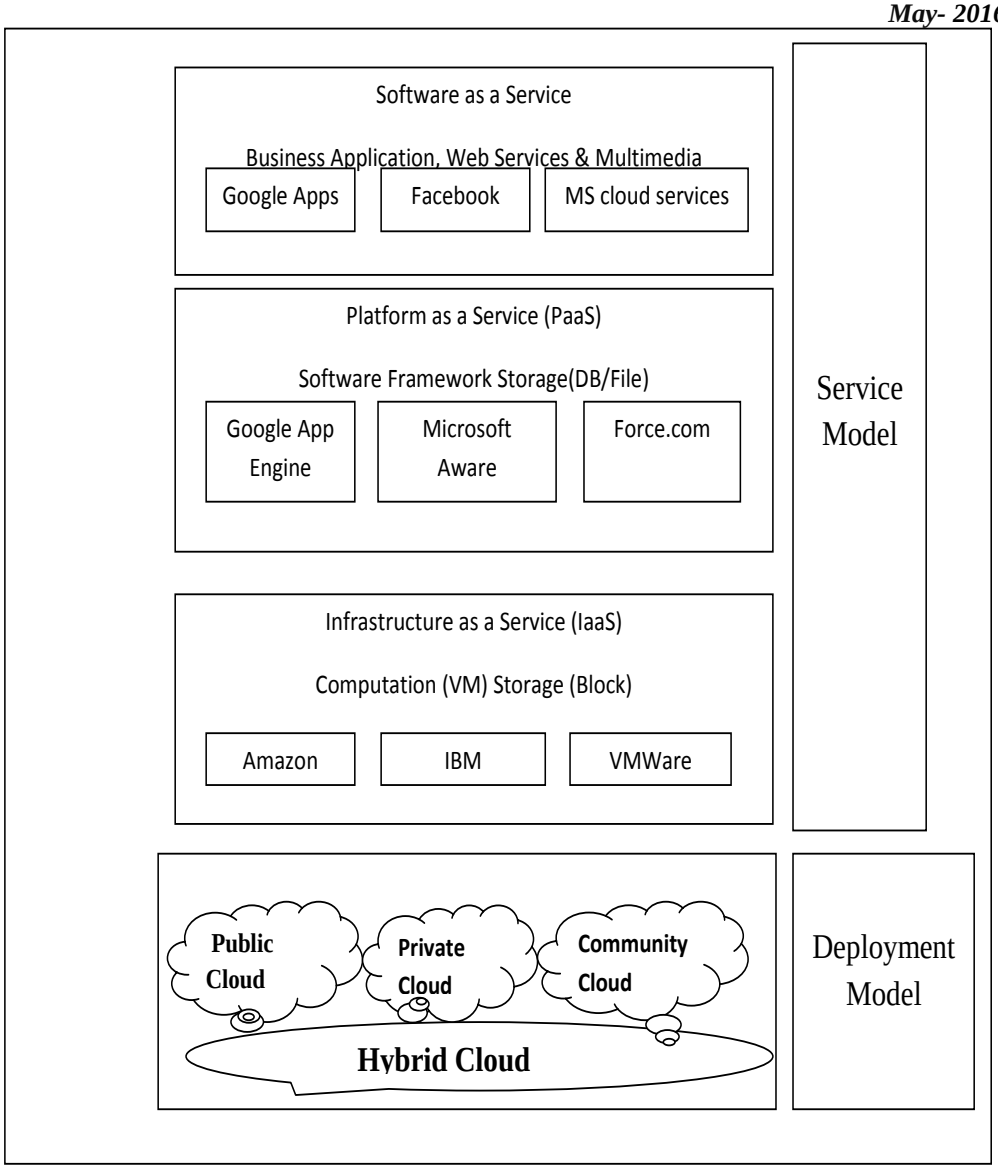

Figure 1 Cloud Platform

- Resource pooling

The provider's computing resources are pooled together to serve many consumers using multiple-tenant model, with distinct physical and virtual resources dynamically assigned and reassigned according to customer requirement. The resources include among others storage, processing, memory, network radio bandwidth, virtual machines and email services. The pooling together of the resource builds economies of scale.

- Rapid elasticity

Cloud services could be quickly and elastically provisioned, in some cases automatically, to quickly scale out and quickly released to instantly scale in. To the consumer, the capabilities available for provisioning often appear to be vast and can be purchased in any quantity at any time.

- Measured service

Their source usage can be measured, composed, and noted given that transparency for both the provider and consumer of the utilised service. Cloud computing services apply a metering ability which enables to control and optimise resource use. It is just related to air time, electricity or municipality water IT services are owed per usage metrics – pay per use.

- Multi Tenacity

This is the last characteristic of cloud computing and, is advocated by the Cloud Security Alliance. In this, Consumers may utilize a public cloud provider's service offerings the same organization, like different business units rather than different organizational entities, however would still share framework0. There are many problems faces in cloud computing [6],[7]. like:

- Ensuring appropriate access control
- Network level migration, so that it requires low cost and time to shift a job
- Offer appropriate security to the data in transit and to the data placed at rest.
- Data availability problems in cloud.
- Data lineage, data origin and unintended leak of sensitive information is possible.
- And the most common problem in Cloud computing is the problem of Load Balancing.

© 2016, IJARCSSE All Rights Reserved

Problem issues in Clouds

Cloud computing is recently a growing area in the information technology region. However, the technology is still not totally developed. There are still some areas that are needed to be concentrate on:

- Resource Management

Users are utilizing less of their own existing resources, while increasing usage of cloud resources. With the evolution of new technologies such as mobile devices, these devices are commnly under-utilized, and can provide same functionality to a cloud provided they are properly configured and managed.

- Task Scheduling

Task scheduling and of resources are main problem areas in both Grid as well as in cloud computing. Cloud computing is an evolving concept in IT domain. The scheduling of the cloud services to the consumers by service providers reflects the cost benefit of these computing paradigms.

- Security

Security issues faced by cloud providers and their consumers.

- Fault Tolerance

The increasing demand of Cloud computing as an attractive alternative to information processing systems has increased the value of its correct and continuous operation even in the presence of faulty components.

- Load Balancing

Cloud Computing is an emerging computing paradigm. The main aim is to share data, calculations, and service done with a scalable network of nodes. As Cloud computing stores the data and resources in the open environment. So, the huge amount of data storage increases rapidly.

II. TASK SCHEDULING ON RESOURCES

Task scheduling and arrangement of resources are main problem areas in both Grid as well as in cloud computing. Cloud computing is an emerging technology in IT domain. The scheduling of the cloud services to the consumers by service providers effect the cost benefit of these computing paradigms. In a cloud environment, traditional scheduling methods are inconceivable owing to its properties - dynamical, distributed, and sharable. The main aim of resource allocation to tasks is for unified services to accomodate their performance targets. Several jobs demand different resources while running simultaneously. It is important for active working of cloud to balance these jobs on appropriate resources for optimal performance, and numerous task parameters need to be considered for proper scheduling. The available resources should be used effectively without affecting the service guidelines. Scheduling in the cloud environment system is an NP-complete problem. As the number of consumers increases, the tasks that need to be scheduled increase in proportion. Therefore, there are so many algorithms are provided by various researchers for task scheduling. In 2008, A heuristic method to schedule bag-of-tasks(tasks with short execution time and no dependencies) in a cloud is presented in [12] so that the number of virtual machines to execute all the tasks within allocated cost, is minimum and the same time speedup. In 2009, Marios D. Dikaiakos and George Pallis executed the concept of organization of Distributed Internet Computing as Public Utility and addressed the several serious problems and unexploited opportunities concerning the deployment, efficient operations and value of cloud computing framework [13]. In 2009, Dr. Sudha and Dr. Jayarani proposed the efficient Two-level scheduler(user centric meta-scheduler for collection of resources and system centric VM schedular for dispatching jobs) in cloud computing technology is based on QoS[14]. In 2010, Yujia Ge and Guiyi Wei proposed a new scheduler which makes the scheduling decision by classify the entire group of tasks in a job queue. A genetic algorithm is designed as the optimization method for a new scheduler who gives better makespan and better balanced load across all nodes than FIFO(First In First Out) and delay scheduling[15]. In 2010, An optimal scheduling scheme based on linear programming, to out source deadline restraint workloads in a hybrid cloud scenario is scheduled in [16]. In 2011, Sandeep Tayal proposed an algorithm based on Fuzzy-GA optimization which classify the entire group of tasks in a job queue on basis of prediction of execution time of tasks authorize to convinced processors and prepare the scheduling decision [17]. In 2011, Laiping Zhao, Yizhi Ren & Kouichi Sakurai recomended a DRR (Deadline, Reliability, Resource-aware) scheduling algorithm, which schedules the tasks such that all the jobs can be concluded before the deadline, ensuring the Realiability and minimization of resources [18]. In 2011,S.Sindhu & Saswati Mukherjee planned two algorithms for cloud computing environment and compared it with default policy of cloudsim toolkit , at the same time considering computational complexity of jobs. This paper presents a framework for our investigation [19].

Load Balancing

Load balancing is a computer network method for distributing workloads beyond multiple computing resources, for example computers, a computer cluster, network links, central processing units (CPUs) or disk drives. Load balancing plans to optimize resource use, maximize throughput, minimize response time, and escape overload of any one of the resources. By the use of multiple components with load balancing rather than a specific component may increase reliability over redundancy. Load balancing in the cloud differs from classical thinking on load-balancing planning and implementation by using commodity servers to perform the load balancing because it's difficult to conclude the number of requests that will be issued to a server. This gives new opportunities and economies-of-scale, also presenting its own exclusive set of challenges. Load balancing is one of the central problem in cloud computing [8]. It is a tool that distributes the dynamic local workload evenly across all the nodes in the perfect cloud to avoid a situation where few nodes are heavily loaded while others are idle or doing some work. It helps to attain a high users satisfaction and resource

© 2016, IJARCSSE All Rights Reserved

utilization ratio, consequently improving the global performance and resource utility of the system. It also makes sure that every computing resource is distributed efficiently and fairly [9]. It further inhibit bottlenecks of the system which may occur due to load imbalance. When one or more factors of any service stop working, load balancing facilitates in continuation of the service by implementing fair-over, i.e. in provisioning and de-provisioning of details of applications without fail. Fig 1 depicts the Load Balancing requirement in cloud when there are requests from various clients. The existing load balancing techniques in clouds, consider different parameters like performance, response time, scalability, throughput, resource utilization, fault tolerance, migration time and combined overhead. The emerging cloud computing model attempts to address the explosive growth of web-connected devices, and hold huge amounts of data[10] and client demands. Thereby, giving rise to the question even if our cloud model is capable to balance the ever-increasing load in an effective way or not.

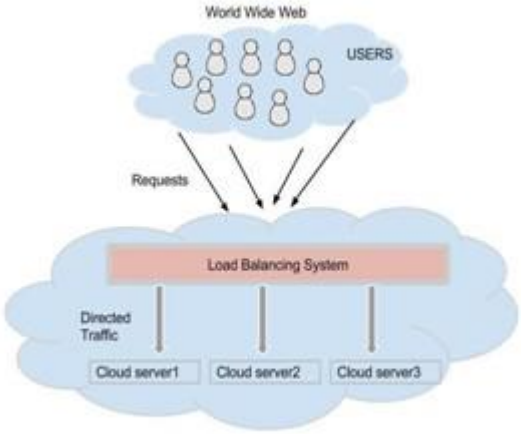

Fig 1. Load Balancing system in cloud computing

III. LITERATURE REVIEW

Shaminder Kaur.et.al(2012)

Cloud computing is recently a booming area and has been emerging as a commercial reality in the information technology field. Cloud computing shows supplement, consumption and delivery model for IT services that are on internet on pay as per usage basis. The scheduling of the cloud services to the consumers by service providers effects the cost benefit of this computing paradigm. In such a scenario, Tasks should be scheduled efficiently as the execution cost and time could be reduced. In this paper, we proposed a meta-heuristic based scheduling, which reduces execution time and execution cost as well. A revised genetic algorithm is developed by merging two existing scheduling algorithms for scheduling tasks taking into review their computational complexity and computing capacity of processing items. Experimental results show that, under the heavy loads, the proposed algorithm exhibits a good performance.

Saurabh Bilgaiyan.et.al(2014)

Cloud computing is a popular computing paradigm that performs processing of huge volumes of data using extremely accessible geographically distributed resources that can be accessed by users on the basis of Pay As per Use policy. In the trendy computing environment where the amount of data to be processed is increasing day by day, the costs engaged in the transmission and execution of such amount of data is mounting significantly. So there is a concern of appropriate scheduling of tasks which will help to manage the escalating costs of data intensive applications. This paper analyzes many evolutionary and swarm based task scheduling algorithms that address the above mentioned problem.

Nikita Haryani and Dhanamma Jagli(2014)

The state-of-art of the technology focuses on data processing and sharing to deal with great amount of data and client's needs. Cloud computing is a promising technology, which allows one to achieve the preceding goal, leading towards enhanced business performance. Cloud computing comes into center of attention quickly when you think about what IT constantly needs: a means to increase capacity or add abilities on the fly without expanding in new infrastructure, training new human resources, or licensing new software. The cloud should give resources on demand to its users with high availability, scalability and with reduced cost. Cloud Computing System has widely been taken by the industry, though there are many existing issues which have not been so long wholly addressed. Load balancing is one of the primary challenges, which demands to distribute the dynamic workload across variety of nodes to assure that no single node is affected. This Paper gives an efficient dynamic load balancing algorithm to cloud workload management by which the load can be assigned not only in a balanced approach, but also it gives the load systematically and uniformly by checking positive parameters like number of requests the server is handling presently. It balances the goods on the overloaded node to under-loaded node so that response time from the server will decreases and performance of the system is increased.

Mala Kalra.et.al(2015)

Cloud computing has been a buzzword in the area of high performance distributed computing as it gives on-demand access to common pool of resources over Internet in a self-service, dynamically scalable and metered way. Cloud

© 2016, IJARCSSE All Rights Reserved

computing is still in its infancy, so to obtain its full benefits, much research is required across a broad array of concepts. One of the mandatory research issues which need to be focused for its valuable performance is scheduling. The main goal of scheduling is to shape tasks to appropriate resources that optimize more than one objectives. Scheduling in cloud computing relating to a category of problems known as NP-hard problem due to large amount of solution space and thus it takes a long time to find an best solution. There are no algorithms which may produce best solution within polynomial time to solve these problems. In cloud context, it is preferable to find suboptimal solution, but in concise period of time. Metaheuristic based techniques have been proved to attain near optimal solutions within reasonable time for said problems. In this paper, we provide an expanded survey and comparative analysis of many scheduling algorithms for cloud and grid environments based on three trendy metaheuristic techniques: Ant Colony Optimization(ACO), Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), and two unique techniques: League Championship Algorithm (LCA) and BAT algorithm.

Jyoti Thaman.et.al(2016)

Cloud computing is a development of parallel, distributed and grid computing which gives computing potential as a service to clients comparatively a product. Clients can operate software resources, valuable information and hardware devices just as a subscribed and monitored service over a network through cloud computing. Due to huge number of requests for access to resources and service level agreements between cloud service providers and consumers, few burning issues in cloud context like QoS, Power, Privacy and Security, VM Migration, Resource Allocation and Scheduling must attention of research community. Resource allocation among many users be assured as per service level agreements. Several techniques have been developed and tested by research community for generation of best schedules in cloud computing. A few encouraging approaches like Metaheuristics, Greedy, Heuristic technique and Genetic are tested for task scheduling in several parallel and distributed systems. This paper presents a review on scheduling proposals in cloud context.

IV. RELATED WORK

Scheduling of tasks is a critical problem in Cloud Computing, so a lot of researches have been done in this area. The following table summarizes distinct genetic task scheduling algorithms based on their scheduling goals, parameters and tools used along with future scope.

Paper	Authors	Findings	Scheduling Parameters	Tools	Future Scope
Efficient approach to GA for task scheduling in cloud computing environment	Shaminder Kaur, Amandeep Verma (2012)	Two-point crossover method is used.	Initial population is generated by using SCFP and LCFP techniques in private cloud environment, Genetic Algorithm, Makespan, Task-Scheduling, cost	CloudSim	Enhance algorithm by supporting runtime scheduling, priority of jobs for multiple users.
Independent task scheduling in cloud computing by improved Genetic Algorithm[7]	1Shekhar Singh, 2Mala Kalra (2014)	Fitness function is based on minimization of Make Span. Proportion selection method is used as selection operator	Make span, Resource Utilization	CloudSim	Execution cost can be taken as fitness criteria. Explore towards dependent and dynamic jobs.
Dynamic method of load balancing	Nikita Haryani[1], Dhanamma Jagli[2] (2014)	Estimation of load, comparison of load, stability of different systems.	Performance, response time, scalability, throughput, resource utilization, reducing overhead, low migration time and improving performance etc.	Hadoop MapReduce technology	
Task Scheduling optimization for the cloud computing system [9]	S.Tayal	Fuzzy GA optimization is used in which scheduling decision is made by evaluating the entire group of task in the job queue.	Fitness function is based on minimization of makespan	Cloud Sim	Improvement required on the accuracy of predicted completion time of job.

© 2016, IJARCSSE All Rights Reserved

V. CONCLUSION

This paper presents survey on Task Scheduling optimization in cloud computing system. it provides Fuzzy GA optimization that is used in scheduling decision and Fitness function is based on minimization of makespan. In this paper, we also studied about load balancing, that how to balance the load when more than one systems are connected. Techniques used for load balancing is reducing overhead, reducing the migration time and improving performance etc., but the response to request ratio is rarely considered. In efficient research to GA in cloud computing environment, we have proposed a altered genetic algorithm for single user activity in which the fitness is developed to encourage the formation of solutions to attain the time reduction and related it with current heuristics. Experimental results show that, under the heavy loads, the recomended algorithm exhibits a good performance.

REFERENCES

[1] M. R. Miryani, and M. Naghibzadeh, "Hard real-time multiobjective scheduling in heterogeneous systems aquire genetic algorithms." In Computer Conference. CSICC 2009. 14th International CSI, pp. 437-445. IEEE, 2009.

[2] S.Tayal. "Task scheduling optimization in cloud computing systems." International Journal of Advanced Engineering Sciences And Technologies (IJAEST) 5, no.2(2011):111-115.

[3] Kumar, P., Verma, A. (2012). 'Independent Task Scheduling in Cloud Computing by revised Genetic Algorithm'. International Journal of Advanced Research in Computer Science and Software Engineering. 2, 5, 111-114

[4] Patel, R., and Patel, S., "Survey on Resource Allocation Strategies in Cloud Computing", International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 2, February- 2013, 1-5.

[5] Junwei, G. and Yongsheng, Y. (2013). Research of cloud computing task scheduling algorithm based on improved genetic algorithm. In dealings of second International Conference on Computer Science and Electronics Engineering, 2134-2137.

[6] Dikaiakos, M., katsaros, D., Mehra, P., Vakali, A.: Cloud Computing: Distributed Internet Computing for IT and Scientific Research‖. In IEEE Transactions on Internet Computing 13(5), pp. 10-13 (2009)

[7] Sindhu S., Mukherjee S.: Efficient Task Scheduling Algorithms for Cloud Computing context. In International Conference on High Performance Architecture and Grid Computing (HPAGC-2011), vol 169, pp 79-83 (2011)

[8] SHANTI SWAROOP MOHARANA, RAJADEEPAN D. RAMESH & DIGAMBER POWAR, "ANALYSIS OF LOAD BALANCERS IN CLOUD COMPUTING", International Journal of Computer Science and Engineering (IJCSE) ISSN 2278-9960 Vol. 2, Issue 2, May 2013, 101-108.

[9] Y.Fang, F.Wang, and J.Ge, "A Task Scheduling Algorithm Based on Load Balancing in Cloud Computing", Web Information Systems and Mining, Lecture record in Computer Science, Vol. 6318, 2010, pages 271-277

[10] http://www.ukessays.com/essays/information-technology/implementating-distributed-load-balancing-algorithms-information-technology-essay.php

[11] Verma R., Dhingra S., "Genetic Algorithm for Multiprocessor Task Scheduling", IJCSMS International Journal of Computer Science and Management research, Vol.1, Issue 02, pp. 181-185, 2011

[12] Mei, L., Chan, W.K., Tse, T.H., ―A Tale of Clouds Paradigm Comparisons and Some ideas on Research Issues, In APSCC 2008, pp. 464-469 (2008)

[13] http://code.google.com/appengine

[14] Zhao C., Zhang S., Liu Q., Xie J., Hu J., "Independent Tasks Scheduling Based on Genetic Algorithm in Cloud Computing", IEEE fifth International discussion on Wireless Communications, Networking and Mobile Computing WiCom '09, Beijing, pp.1-4, 2009

© 2016, IJARCSSE All Rights Reserved